By the Numbers

Volume 13, Number 1 The Newsletter of the SABR Statistical Analysis Committee February, 2003

News

Phil Birnbaum

As I write this, the 2002 "Baseball Research Journal" is in the mail. And this year, for what I think is the first time, BRJ will include a study previously published in *By the Numbers* – Tom Hanrahan's *"Does Experience Help in the Postseason?"* from last issue. (There would likely have been others, but many of them had already appeared in the "Best of" book.)

BRJ has taken some criticism in the past for allowing inappropriate statistical material to appear. This year, Jim Charlton, SABR's Publications Director, has rectified the situation, sending submissions out to statistically-oriented members for review and comment. As well, he's eager to publish articles that have already appeared in BTN.

Previously, many authors submitted to BRJ instead of here because of the larger, more general readership BRJ reaches. Now, they can submit to BTN, knowing that publication here automatically results in consideration for the upcoming BRJ.

I had the opportunity to review a few of the submissions Jim received for BRJ. What struck me was the number of authors who didn't seem to be aware of the state of research in their field. A couple had barely read the works of Bill James, and many were not even aware of "By the Numbers". Some weren't even SABR members.

In that light, Jim Charlton suggested we put togther a style sheet and/or submission guidelines for statistical research submitted to BRJ. My thought was that we'd concentrate on ranking stats, because many (perhaps even most) of the manuscripts Jim receives are on this topic. The guidelines would explain what's already out there, describe the results supporting their accuracy, insist on empirical results (that is, the new stat must measure something, not just arbitrarily combine numbers), and include critiques of stats that have not gained sabermetric acceptance.

The guidelines would be posted on the SABR website.

Any thoughts or suggestions on this topic would be welcome; you can reach me at the usual addresses.

In this issue

A couple of months ago, Rob Wood referred me to a website called *www.baseballprimer.com*. The site features some excellent sabermetric research from its writers, some of whom are SABR members (but many of whom are not). While much of the content is commentary on current players and teams, digging out the sabermetric articles is very worthwhile. See, for instance, a nice article on streakiness:

http://www.baseballprimer.com/articles/lichtman_2002-11-15_0.shtml

If anyone has other internet research to recommend, please let me know. There's lots of interesting work going on out there, and often we don't know it exists. So please, let us know where the good stuff is.

Phil Birnbaum, 18 Deerfield Dr. #608, Nepean, ON, Canada, K2G 4L1, birnbaum@sympatico.ca ♦

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 (but no death threats, please) are all welcome.

Articles should be submitted in electronic form, either by e-mail or on PC-readable floppy disk. 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. (But if you want to make my life a bit easier: please, use two spaces after the period in a sentence. Everything else is pretty easy to fix.)

If you can (and I understand it isn't always possible), try to format your article roughly the same way BTN does, and please include your byline at the end with your address (see the end of any article this issue).

Deadlines: January 24, April 24, July 24, and October 24, for issues of February, May, August, and November, respectively.

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

Send submissions to:

Phil Birnbaum
18 Deerfield Dr. #608, Nepean, Ontario, Canada, K2G 4L1
birnbaum@sympatico.ca

"Moneyball" – Sabermetrics Reaches the Major Leagues

Clem Comly

While it reveals no new sabermetric knowledge, "Moneyball," the author says, is an entertaining account of how sabermetrics finally comes to a baseball front office, helping the Oakland A's put together great teams cheap.

I liked it, but it barely deserves to be in our newsletter. This is a book about the Oakland front office making fools of the rest of baseball by putting together pennant contenders on the cheap. Billy Beane and the rest of the Oakland management gave author Michael Lewis access to the front office during the 2002 season. The book's key early scene (which runs over 30 pages) is the Oakland draft room before and during the 2002 June free agent draft of college and high school players. The book's key late scene is Billy Beane's office during late July 2002 as trades are being negotiated.

In between are biographical chapters of Beane, pitcher Chad Bradford, and Bill James, as well as significant biographical sketches of Voros McCracken and minor league catcher (and 2002 draftee) Jeremy Brown. Among the SABR names dropped by Lewis are Dick Cramer, Pete Palmer, and Craig Wright. The book covers people doing or using sabermetrics, but I will mine almost all of the sabermetric knowledge of the book in a few paragraphs below. I am in the same position as an AAAS reviewer of a book about Charles Darwin—the subject may be of interest to the members but they would much prefer a book about evolution.

Moneyball

By Michael Lewis

Norton, 288 pages, $24.95 (US)
ISBN 0393057658

The Oakland edge is the use of the following sabermetric principles familiar to our committee members: don't undervalue the ability to draw walks, don't undervalue extra base hits, don't be fooled by saves, don't overvalue batting average, don't overvalue speed, don't overvalue defense, don't undervalue the compensation for losing free agents, and don't waste outs on sacrificing or being caught stealing. Also, they use Bill James' conclusions from his study of the free agent draft: draft college players over high school players -- college players require less projection of their talents, because their statistics are meaningful and allow comparison.

The book does a nice job of explaining these principles, and, often as not, gives credit to the originator (usually Bill James). There must be almost 100 pages about Bill James that gives Lewis the chance to quote James' writings extensively (I mean extensively). James' prose brightens the book significantly. The book has no index, an omission that made writing this review a real chore—I can only assume the publisher didn't want to explain Bill James having an entire page in the index.

What was new or different that SAC would care about? Oakland places more emphasis on OBA than slugging. Beane's assistant Paul DePodesta is quoted as saying a point of OBA is worth three times a point of slugging average.

Under the category of "Momma, don't let your babies grow up to be sabermetricians", Lewis reports that a company called AVM Systems came up with a system to record where batted balls are hit and how hard with more accuracy than STATS. AVS used this information to figure out the value of the average ball hit to spot X with speed Y and could thus award that value to each hitter whether that ball was actually caught or went for a single or double, etc. AVM Systems could then derive the value of a player as the sum of all of these expected values. Lewis reports "in 1998, Paul [DePodesta] persuaded Billy [Beane] to hire AVM Systems…Billy and Paul used the system for a couple of years and then, to save money, copied what AVM did. Once Paul finished replicating the parallel world [of expected values]…The precision of the AVM System, copied by Paul, enabled him [Paul] to think about every event that occurred on the baseball field in a new and more satisfying way."

Can Oakland keep this up? While Lewis does not address this question, based on the facts in *Moneyball*, no. Toronto hired J. P. Ricciardi away from Oakland to be GM. Boston offered big bucks to get Beane to join them. The teams that are less rigid or more desperate will be able to mimic Oakland and reduce if not eliminate the bargains Oakland has been getting. So, unless Beane and DePodesta have several secrets they did not reveal to Lewis, Oakland is at or near its high-water mark.

So, while adding little to my sabermetric knowledge, *Moneyball* was nonetheless entertaining. The June 2002 free agent draft scene where Billy Beane mocks the Oakland scouts and uses DePodesta's evaluations instead is literally fantastic.

Clem Comly, 308 Colonial Drive, Wallingford, PA, 19086-6004; ccomly@erols.com ♦

Academic Research: Pitch Expectations and Home-Field Advantage

Charlie Pavitt

The author catalogs two more recent studies from academic journals: one on player pitch expectations and cognitive processes; the other on a possible cause of home-field advantage, with a rare finding of significance.

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 2002 edition of 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.

Rob Gray, Markov at the Bat: A Model of Cognitive Processing in Baseball Hitters, Psychological Science, Volume 13 Number 6, November 2002, pages 542-547

This one's a bit different from the type of article I usually review. Gray studied the timing of a swing in relation to pitches of different speeds given different counts. He assumed that (1) the ideal time to make contact is when a ball is nine-tenths of a meter in front of the plate, (2) the ideal trajectory for a swing makes contact with the pitch at its lowest point, and (3) batters expect "fast" pitches when the count is 2-0, 3-0, or 3-1, "slow" pitches (i.e., breaking balls and change-ups) when the count is 0-2 and 1-2, and have no expectation of either in other counts. He then ran a simulation using six players of differing levels (two each from College Division 1A, College Division 3, and recreational leagues) and a pitching machine in which pitch speed (fast versus slow) and location (strike versus ball) was varied; the former depended on the count but the latter was randomized. Gray then measured the "temporal error" in the batters' swings (the difference in time between when pitches were nine-tenths of a meter away and when the batters' swings were at their lowest point in their trajectories). All batters had smaller temporal errors when the pitch matched their hypothesized expectations than when the pitch did not, higher level players had smaller temporal errors than lower level players, and the entire exercise could be modeled as a relatively simple Markov process.

Richard Pollard, Evidence of a Reduced Home Advantage When a Team Moves to a New Stadium, Journal of Sports Sciences, Volume 20 Number 12, December 2002, pages 969-973

Back in the May 2001 *By The Numbers,* I reported on a study testing whether the home field advantage was at least partly due to the wear and tear of travel on visiting teams. The study showed no supportive evidence, and I concluded that "we still have no good idea what causes home team advantage." The present study examines another possible factor, the home team's relative familiarity with their ballpark, by seeing if it decreases when a team moves to a new park in the same metropolitan area. Based on a sample of 7 moves from 1988 through 2001, there was indeed evidence of a slight but insignificant decrease, from a winning percentage of .547 in the last year in the old park to .537 in the first year in the new. Analogous data from 17 NBA moves showed a slightly smaller decrease, but 13 NHL moves showed a much larger decrement. Pooling the data provided a large enough sample size for statistically significant findings ($p = .011$). While not quite earth-shattering, it is nice to find a study that finds some evidence for some explanation of home-field advantage.

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

The Predictive Value of Half-Season Statistics

Dan Levitt

A popular forecasting technique is to give special weight to a player's second-half statistics in predicting his next year's performance – the idea being that his ability is constantly changing, and the second half is closer to his current, true ability. Here, the author checks whether the data support that hypothesis.

A number of years ago as Rotisserie baseball evolved into a widespread phenomenon, several sources offering game advice began touting their baseball player forecasting techniques. One of the suggestions to help project a player's statistics for the coming year was to examine and give extra weight to a player's statistics over the second half of the season. This makes some intuitive sense: the second half is closer in time to the coming year and any changes in ability over the year should show up in a player's second half statistics.

Back in the early 1990s I made a quick study of this assertion by comparing player data for batters from 1991 to 1992, which I have summarized below. Today, with the data available from Retrosheet and several published sources as well, one could make a much more detailed analysis. Nevertheless, I think just looking at the one year can give a good quick perspective on the predictive value of second-half statistics.

In all, I found 109 players who had at least 200 at bats in both the first and second halves of the 1991 season, and at least 250 at bats in 1992. The statistic I used to compare the data between seasons was the stolen base version of runs created per game. Table 1 shows the correlation between the first half, second half, and full season 1991 runs created per game with the full year 1992 runs created per game.

Second half statistics appear a better predictor of the next year's performance than those from the first half. But the key issue is: do they add any additional predictive information beyond the full year statistics? In other words, does supplementing the 1991 full season statistics with the second half results help create a better forecast for the following year?

In Table 2, the first row reprints the correlation of the 1992 full season with just the 1991 full season data from Table 1. The final two rows present the correlation of the 1992 data with the combination of the 1991 full season data and half year statistics. [1]

Table 1: Correlation with Following Year RC/game

	Correlation With Full Year 1992
First Half 1991	0.56
Second Half 1991	0.64
Full Year 1991	0.68

Table 2: Correlation of Full and Half Year With Following Year RC/game

	Correlation With Full Year 1992
Full Year 1991	0.677
FY & First Half 1991	0.681
FY & Second Half 1991	0.682

[1] As most readers of this publication know, the correlation coefficient measures the strength and direction of the association between two variables—such as RC/g over the second half of 1991 and the RC/g during the full season 1992. If when one varies, the other varies in a related manner, the data is described as correlated. Specifically, the correlation value can run from -1 to 1; 1 signifying an exact association (the two variables move in lockstep), -1 an exact negative correlation, and 0 no correlation. A multiple correlation coefficient can be used to measure the strength of a relationship between a dependent variable and a set of independent variables—such as the full season statistics in combination with the second half.

As can be seen from Table 2, inclusion of the second half data with the full season adds very little to the correlation coefficient. In other words, this analysis strongly implies that the additional predictive information provided by second half statistics over and above that available from the full season alone is basically zero.

In conclusion, although this study is based on a fairly small (one-year) sample, it appears that in projecting player statistics from one year to the next, knowing how the player performed in each half of the season does not offer any significant assistance.

Dan Levitt, 4401 Morningside Road, Minneapolis, MN, 55416, danrl@ibm.net ♦

Applications of Win Probabilities

Phil Birnbaum

In another application of Retrosheet play-by-play data, the author computes the average probability of winning in every game situation, and applies the results to questions of run timing, such as intentional walks and relief pitching.

Conventional wisdom is that a hitter should be judged by how many runs he creates for his team.

That's a bit of an oversimplification. A player's job is not really to create runs (or, in the case of a pitcher, to prevent runs). The player's job is to create *wins*. Runs, as Bill James once wrote, are the currency with which runs are purchased.

But not all runs are equal in purchasing power. A solo home run in the ninth inning of an 12-1 game is worthless in terms of wins. On the other hand, with the score tied in the bottom of the ninth, that same home run buys half a win. The average, as Pete Palmer discovered, is that overall, a run is worth a tenth of a win. Each extra win costs ten runs.

Bill James, and others, have noticed that some players seem to produce more runs in close games, and others appear to produce lots of runs when they don't really mean much. James created the "Victory-Important RBI" statistic to illustrate the difference. Though this statistic will give us the historical fact of whose runs happened to be more or less clutch than others, studies have shown that clutch hitting is not a skill, but rather appears to be the result of luck. For batters, then, it is probably fair to say that runs are generally a reasonable and accurate surrogate for wins.

Nonetheless, there are exceptions -- situations when runs is *not* an appropriate substitute for wins. Studies have shown a stolen base to be worth a certain number of runs, on average. But in the bottom of the ninth with the score tied and Rod Carew at bat, a stolen base is much more valuable than in the top of the third. To study the value of the steal, it's not enough to simply compute how many runs it theoretically creates. A player may not be able to choose when he hits a home run, but he and his manager *are* able to choose when he attempts to steal. Since we expect that more steals will be attempted when the stakes are higher, a runs-only analysis may undervalue the stolen-base strategy.

Another unsolved sabermetric problem is determining the value of the relief pitcher. A closer might have an ERA of 2.50 in 60 innings pitched. A good middle reliever might have the same stats – but is nowhere near as important to his team. The stopper pitches when runs are most valuable, and his work must be evaluated in that context. Who has more value to a team: Mariano Rivera or Andy Pettitte? We still don't know the answer to that question.

But now, thanks to Retrosheet, we have the evidence to find out.

Win Probs

Using Retrosheet's play-by-play data of both leagues from 1974 to 1990, I determined a team's probability of winning based on

(a) whether it was the home team or visiting team;
(b) how many runs it was ahead or behind;
(c) what inning it was;
(d) the number of outs; and
(e) which bases were occupied, if any.

For instance, consider the situation where it's the bottom of the ninth, the score is tied, and the home team has a runner on third with one out. That happened 299 times in our sample, and the home squad won 249 of those games. Its probablity of winning, then, is 249/299, or 83.2%:

Batting Team	Inning	Score	Bases Occupied	Outs	Probability
Home	9th	Tied	Third	One	.832

I repeated this calculation for every possible combination of team, inning, score, and base/out situation.[1] There were 5,616 situations – 2 teams, times 13 scores, times 9 innings, times 3 numbers of outs, times 8 possible bases-occupied situations.

It's not possible to print a 5,616 row table here[2]. But, by way of example, Table 1 shows part of that dataset: bottom of the 9th with the score tied.

The table entries are the chance of the home team eventually winning the game. So with a man on first and nobody out in a tie game in the bottom of the ninth, the home team has a .713, or 71.3%, chance of eventually winning the game.

I'm going to call this number the "Win Prob."

The numbers in the chart should be close to what you'd expect. At the start of the inning, for instance, the home team has a win prob of .649 (64.9% chance). We know that must be more than .500, since the visiting team has already batted, and this half-inning is therefore a free chance to win. So .649 seems reasonable.

And, to take one more example, with two out and nobody on, the .549 sounds about right. There's still a chance of scoring a run before the last out, and, even if that doesn't happen, there is still a greater than 50% chance of the home team winning in extra innings.

Table 1: Probability of home team winning for all base/out situations, score tied in the bottom of the 9th

	0 out	1 out	2 out
Nobody on	.649	.597	.549
Man on 1st	.713	.663	.581
2nd	.792	.703	.606
1st and 2nd	.821	.727	.631
3rd	.873	.832	.634
1st and 3rd	.868	.820	.671
2nd and 3rd	.880	.826	.642
Loaded	.899	.831	.650

More importantly, the *order* of the probabilities is approximately in line with what we'd expect. The chance of winning with a runner on 2nd is consistenly better than if the runner is only on first; the runner on 3rd improves chances over the runner on 2nd, and so on. There are exceptions, though – with two outs, first-and-third appears better than second-and-third, which can't be right. That's probably a sample size issue – even with 12 years of games, there's still a bit of randomness in the data.

In the above chart, the smallest sample is for second-and-third with nobody out: it's based on only 84 games. The probability estimate of .880 has a standard error of .035, which means an approximate 95% confidence interval is approximately .810 to .950. The largest sample, is for the nobody-on-nobody-out cell – with 9,363 games, the standard error is .005, so the confidence interval is about .639 to .659.

Intentional Walks

The intentional walk is perhaps the most obvious strategy in which timing is important. For the free pass, run-based methods really can't help us evaluate the strategy. But win probs can help.

In our bottom of the 9th situation, with a runner on second base, should we walk the next batter? Looking at the chart:

- With no outs, the walk increases the home team's chances by 29 points (.029 probability).
- With one out, the walk increases the home team's chances by 24 points.
- With two outs, the walk increases the home team's chances by 25 points.

So far, it looks like the walk is a bad strategy. Of course, we are looking for tiny differences in a sample size that might be too small for the task. But even if we knew these results are accurate, there are a couple of reasons that the free pass might still be the right strategy.

First, our analysis assumes average players coming up before and after the walk. But in real life, if the batter is Barry Bonds or Alex Rodriguez, the advantage of pitching to a weaker hitter could easily be more than the 25-or-so points the home team gains by the baserunner.

Second, and more subtly, we need to consider that the probabilities in the charts already take into account that the manager is going to walk the next player in many of these cases. The apparent benefit of the IBB to the home team might be the result of the fact that when it's not advantageous to walk the next batter, the manager won't do it. Indeed, this result could be entirely the result of managerial astuteness.

[1] To keep things simple, if a team was more than 6 runs ahead or behind, I counted it as 6. Also, extra inning data was combined with ninth inning data, since the strategic implications are exactly the same.
[2] The full table of data is available from the author, or temporarily at http://www.philbirnbaum.com/winprobs.txt .

To see why, let's assume a simple case: suppose with a runner on first and nobody out, the probability of winning is 70% half the time, and 80% the other half, depending on who the next batter is. With either batter walked, the chance becomes 75%.

In that case, our analysis would show the IBB is a wash: the chance of winning is 75% before the walk, and 75% after. But the team gets a new manager, a savvier one, who realizes that it's better to pitch to the worse hitter but walk the better one. After that strategy, the probability with a runner on first has been reduced from 75% to 72.5%. The probability with runners on 1st and 2nd is still 75%. So the IBB looks like a bad move – but that's only because the manager has already used the intentional walk in the situations where it was a good move!

And so, our analysis here might be faulty, and we can't conclude that the IBB is a bad strategy in this case. But we *can* conclude that, in general, in the cases where the manager chose *not* to use the IBB, that was the correct strategy more often than not.

But having said that, it turns out that the runner on third situation gives different results:

- With no outs, the walk decreases the home team's chances by 5 points.
- With one out, the walk decreases the home team's chances by 12 points.
- But with two outs, the walk increases the home team's chances by 37 points.

Go figure: here, the IBB looks like a good strategy with no outs or one out, but a large loss with two outs. This is probably a sample size issue.

Linear Probs

The chart we've been looking at is very similar to the run potential chart, first used by Pete Palmer in 1984 in *The Hidden Game of Baseball*. Here's Pete's chart, denominated in runs (from the paperback edition, page 153):

	0 out	1 out	2 out
Nobody on	.454	.249	.095
Man on 1st	.783	.478	.209
2nd	1.068	.699	.348
1st and 2nd	1.277	.897	.382
3rd	1.380	.888	.457
1st and 3rd	1.639	1.088	.494
2nd and 3rd	1.946	1.371	.661
Loaded	2.254	1.546	.798

With no outs and nobody on, .454 runs would be expected to score. If the next batter singles, putting a runner on first, the expectation is now .783 runs. That particular single, then, was worth .329 runs. If we were to compute the run increase for every single hit by every batter in every circumstance, and average them all, we would find that the average single is worth approximately .46 runs – which is the term for singles in Palmer's Linear Weights formula.

We can do the same thing with Win Probs. Again using Table 1, a leadoff single in the bottom of the ninth with the score tied increases the probability of winning from .649 to .713. That single is therefore worth .064 wins. And if we average the value of all singles over our entire database, we get the equivalent to the Linear Weight: the Linear Prob. And it turns out that the Linear Prob for a single is .043 wins.

We can do the same for doubles, triples, steals, and so on. Here they all are in chart format, compared to Linear Weights:

	Linear Weight	Linear Prob
Single	.46	.043
Double	.80	.069
Triple	1.02	.094
Home Run	1.4	.138
Walk (unintentional)	.33	.029
Steal	.3	.022
Caught Stealing	.6	.046
Strikeout	- .25	-.0258
Other batting out	- .25	-.0265

There's an obvious relationship here – the Linear Prob seems to be very much proportional to the linear weight. Let's divide the first column by the second:

	Linear Weight	Linear Prob	LW divided by LP
Single	.46	.043	10.7
Double	.80	.069	11.6
Triple	1.02	.094	10.9
Home Run	1.4	.138	10.1
Unintentional Walk	.33	.029	11.4
Steal	.3	.022	13.6
Caught Stealing	.6	.046	13.0
Strikeout	- .25	-.0258	9.7
Other batting out	- .25	-.0265	9.4

The third column is the Linear Weight divided by the Linear Prob: runs divided by wins. The results confirm the theory that it takes ten runs to make a win, as evidenced by the fact that the Linear Weight is generally about 10 times the Linear Prob. The exceptions are SB and CS. I think I read somewhere that Pete Palmer bumped up the values of these two events, citing the fact that they tend to take place at more important moments of the game. This analysis suggests that perhaps they were bumped about 30% too much.

Relief Pitching

Relief pitching is another situation where runs don't tell enough of the story. Runs scored in a close game are much more valuable, in terms of wins, than runs scored in blowouts. And since the best relief pitchers are used only in close games, the number of runs they prevent underestimates their value, since each run is so much more valuable than the same run prevented by a middle reliever.

For instance, with the score tied in the bottom of the ninth, runs are obviously much more important – instead of 10 runs creating a win, it's going to be a lot smaller. But how much smaller?

To find out, I ran the Linear Probs analysis again, but this time, I included only play-by-play data from the bottom of the ninth, and only if the inning started with the score tied. (I included the entire inning, whether or not the score stayed tied.)

As we expected, events are worth a lot more in this critical situation:

	All situations Linear Prob (A)	Only "Bottom of the 9th score tied" Linear Prob (B)	Relative Importance of Situation (A/B)
Single	.043	.109	2.5
Double	.069	.152	2.2
Triple	.094	.232	2.5
Home Run	.138	.366	2.2
Unintentional Walk	.029	.063	2.2
Steal	.022	.049	2.2
Caught Stealing	-.046	-.103	2.2
Strikeout	-.0258	-.060	2.3
Other batting out	-.0265	-.064	2.4

It would be reasonable, here, to say that events are a bit more than twice as important as usual.

But stoppers aren't normally used in tie-game situations. Let's try the classic save situation, where the pitching team is one run up in the bottom of the 9th:

	All situations Linear Prob (A)	Only "Bottom of the 9th, up a run" Linear Prob (B)	Relative Importance of Situation (A/B)
Single	.043	.162	3.8
Double	.069	.272	3.9
Triple	.094	.386	4.1
Home Run	.138	.556	4.0
Unintentional Walk	.029	.105	3.6
Steal	.022	.064	2.9
Caught Stealing	-.046	-.125	2.7
Strikeout	-.0258	-.051	2.0
Other batting out	-.0265	-.049	1.8

Here, the values of hits and walks are about four times as valuable as average. The values of outs, though, are only about twice average. It may seem like something's wrong, but if you think about it, it starts to make sense.

Down a run in the bottom of the ninth, the home team has only a .190 chance of winning the game. If it makes three outs in a row, it goes to zero, a decrease of 190 points. If it scores two runs, it goes to 1.000, an increase of 810 points. Since the team has further to rise than to fall, it follows that the positive events should be relatively more important than the negative events.

Can we collapse these two facts into a single statistic to measure "importance"? One thing we could do is average out the individual RIs, weighted by frequency of occurrence. I'm not crazy about the idea, though – it seems like we're losing valuable information, that hits and walks are very damaging to the visiting team, to a far greater extent than outs help it. But I can't really think of a decent alternative, so I'll do it anyway. By this measure, this bottom-9th-pitching-team-up-a-run situation has importance 2.53.

Here are the top 13 situation importances.

Inning	Pitching Team Situation	Relative Importance
Top 9th	Up 1 run	2.66
Bottom 8th	Up 1 run	2.63
Bottom 9th	Up 1 run	2.53
Bottom 9th	Tied	2.37
Top 8th	Up 1 run	2.36
Top 9th	Tied	2.18
Top 8th	Tied	2.03
Bottom 7th	Up 1 run	2.03
Bottom 7th	Up 2 runs	1.93
Bottom 6th	Up 1 run	1.93
Bottom 8th	Up 2 runs	1.86
Top 7th	Up 1 run	1.75
Bottom 8th	Tied	1.75

Other notable situations:

Inning	Pitching Team Situation	Relative Importance
Top 9th	Down 1 run	0.53
Bottom 1st	Up 2 runs	1.52
Bottom 1st	Tied	1.11
Top 1st	Tied	0.94
Top 9th	Up 2 runs	1.67
Top 9th	Up 3 runs	0.87
Bottom 5th	Up 5 runs	0.83
Bottom 5th	Down 5 runs	0.16

It's pretty clear from the data that everything is much less important when the batting team is ahead than when it's behind. Which makes sense: if you're up four runs, you're probably going to win no matter what you do at bat, whether you score zero runs or ten. But if you're down four runs, you can drastically change your chances of winning by scoring a few runs, and those situations are much more important for the opposition pitchers.

Using your stopper to protect a two-run lead has been criticized in some quarters, but this analysis shows that events are 50% to 100% more important than average in this situation, so using a better-than-average pitcher is certainly called for.

But three run leads don't have a very high importance. And so even though a save is awarded for protecting a three-run lead, it makes sense, then, to save your best pitchers for tie games instead.

In *Win Shares*, Bill James hypothesized that relievers' innings in save situations, because of their importance, should be weighted at a bit less than twice the value of other innings. It seems James was correct. The top situations, where a stopper is used most often, seem to average a bit over 2. Throw in those less-important situations where managers just want to give the reliever some work, and "a bit less than twice" seems right on.

Also vindicated is the sabermetric wisdom that it's better to use your ace reliever in ties than to protect a two-run lead. The tie situation is about a 2.3 in importance, while the two-run lead is only about 1.6.

Notes

Similar research on reliever importance can be found at http://www.baseballprimer.com/articles/tangotigre_2002-12-03_0.shtml . The author, "Tangotiger", provides situations with high importance (which he calls "leverage," probably a better description than "importance") by base-out situation, not just by beginning of inning as I do.

Phil Birnbaum, 18 Deerfield Dr. #608, Nepean, ON, K2G 4L1, birnbaum@sympatico.ca ♦

Platooning Voros McCracken

Clem Comly

In a now-famous study, Voros McCracken theorized that once a batter hits the ball and it stays in the park, whether it's a hit or an out depends only on the defense and not on who the pitcher was. If that theory is correct, the author argues, the platoon advantage should not be evident on those non-HR balls in play. Here, he checks if that is indeed the case.

Voros McCracken hypothesized that whether a ball put in play becomes a hit or an out is a function of the defense, not the pitcher. While pondering this, it dawned on me that if it is the defense that is the key, there should be no platoon advantage unless there is a weakness defensively that LHBs could take advantage of relative to RHBs or vice versa. If there is a platoon advantage on balls hit in play for both LHPs and RHPs on the same team, it would indicate the pitcher is affecting the balls put in play.

Anyone who recalls John Kruk facing Randy Johnson in the All Star game would think that if Kruk had managed to hit the ball, it would have been a weak grounder or pop up. That would be the conventional-wisdom platoon theory, while McCracken would not expect a difference (unless the fielders who get more balls hit at them by LHB than RHBs are weaker or stronger than the fielders that RHB hit at). So I decided to look at two power pitchers, one LH and one RH, playing in front of the same defense for part or all of the last 3 years. I broke OPS (OBP + slugging percentage) into a pitching component (HR, W, K) OPS_P and a defensive component (singles, 2B, 3B, non-strikeout outs) OPS_D:

$$OPS\_P = \frac{HR+W}{K+HR+W} + \frac{4(HR)}{HR+K}$$

$$OPS + D = \frac{H-HR}{AB-K-HR} + \frac{H-HR+2B+2(3B)}{AB-K-HR}$$

For each pitcher, I calculated these two stats for all LHBs faced, and separately for RHBs faced (switch hitters count as batting from opposing hand). I did not have HP or SF or IPHR broken down by opposing batter handedness, but I suspect this won't impact the numbers greatly.

The 2 pitchers, using only their statistics pitching for Arizona are Randy Johnson and Curt Schilling. PA_P is (K+HR+W); PA_D is (AB–K–HR). These are the aggregate results for 2000-2002:

Pitcher	OPS_P vs LHB	OPS_P vs RHB	OPS_D vs LHB	OPS_D vs RHB	PA_P vs LHB	PA_P vs RHB	PA_D vs LHB	PA_D vs RHB
Johnson (L)	.430	.460	.682	.726	148	1191	220	1394
Schilling (R)	.617	.570	.618	.665	405	437	858	793

OPS_P over the 3 years follows the traditional platoon pattern, but OPS_D favors RHB by roughly the same amount with or without the platoon advantage. This seems consistent with Voros' prediction.

I then did the next 2 most used Arizona pitchers, Brian Anderson and Miguel Batista. Anderson is for 2000-2002, Batista for 2001-2002:

Pitcher	OPS_P vs LHB	OPS_P vs RHB	OPS_D vs LHB	OPS_D vs RHB	PA_P vs LHB	PA_P vs RHB	PA_D vs LHB	PA_D vs RHB
Batista (R)	.923	.836	.629	.587	162	195	448	525
Anderson (L)	1.455	1.503	.662	.656	98	329	429	1208

The large difference between this pair of pitchers and the previous pair is found in OPS_P. The lean in favor of RHB in OPS_D before is now the other way a bit. For all four pitchers, the platoon theory correctly predicts the higher OPS_P between LHB and RHB. But for only two of the four does the platoon theory correctly predict the higher OPS_D between LHB and RHB. Two out of four, of course, is exactly what would be expected by chance.

I then calculated the defense's opposing batting average [BA_D = (H-HR)/(AB-HR-K)] for each pitcher for all three years by batter's handedness (the denominator is the same as PA_D above):

	BA_D vs.		H-HR vs.		PA_D vs.	
	LHB	RHB	LHB	RHB	LHB	RHB
Johnson (L)	.300	.320	66	446	220	1394
Schilling (R)	.273	.301	234	239	858	793
Batista (R)	.277	.259	124	136	448	525
Anderson (L)	.287	.287	123	347	429	1208

The platoon theory correctly predicts only 2.5 of the four higher BA_Ds.

Finally, I combined all four pitchers together. First, I grouped by batter handedness:

	BA_D vs.		H-HR vs.		PA_D vs.	
	LHB	RHB	LHB	RHB	LHB	RHB
All 4	.280	.298	547	1168	1955	3920

Then, I grouped by platoon advantage ("Same" is RHB vs RHP and LHB vs. LHP; "opp" is the opposite):

	BA_D vs.		H-HR vs.		PA_D vs.	
	same	opp	same	opp	same	opp
All 4	.287	.295	564	1151	1967	3908

There seems to be more discrimination between the batter handedness than the platoon, which again tends to confirm McCracken.

Clem Comly, 308 Colonial Drive, Wallingford, PA, 19086-6004; ccomly@erols.com ♦

The Bestest – Baseball Dynasties Revisited

Phil Melita

In their book "Baseball Dynasties," Rob Neyer and Eddie Epstein introduced a technique to evaluate which were the best team-seasons of all time. Here, the author argues that their system has a flaw, and suggests a correction to their formulas.

The remarkable success of the 2001 Seattle Mariners stirred thoughts of both the long-ago 1906 Cubs and the only-a-few-years-ago 1998 Yankees. It also revived the age-old question: how good, in the historical sense, is this modern team? How does the Mariner performance measure up against the 1902 Pirates, 1927 Yankees, 1955 Dodgers, 1975 Reds, and all the other "great" teams of baseball history? Is there a way even to compare teams separated by a century? And can we ever identify the best team of all time with anything approaching confidence?

In *Baseball Dynasties: The Greatest Teams of All Time*, Rob Neyer and Eddie Epstein attempt to do just that by examining the run-scoring and run-preventing ability of each team in every year since 1901, using standard deviations to measure the variability of runs in each league. They calculate standardized Runs Scored and Runs Allowed for each team by dividing the divergence from the mean by the standard deviation. Their Standard Deviation Score is the sum of these standardized values. This statistical analysis is wedded to insightful descriptive discussions of twenty great teams, although the conclusions lean heavily upon the SD Scores. The authors make an excellent case for preferring run data to win data, primarily because of the luck involved in winning and losing, and their work is a giant step in a promising new direction. Unfortunately for their results, however, the SD Score method fails under careful scrutiny: it is both skewed by park effects and influenced by the number of teams per league. Acting on the premise that knowledge is gained by building on the efforts of predecessors, it is my intention to reveal these shortcomings and propose an alternate method that more faithfully evaluates team performance using run data.

The 1933 American League clearly demonstrates the first of these failings. Table 1 shows the final standings, the standardized RS and RA, and the SD Scores.

Table 1: 1933 American League, Standardized

	W	L	RS	RA	RS-RA	R/G	StdRS	StdRA	SDS
WAS	99	53	850	665	185	10.0	0.902	1.473	2.375
NY	91	59	927	768	159	11.3	1.673	-0.124	1.549
PHI	79	72	875	853	22	11.4	1.152	-1.442	-0.290
CLE	75	76	654	669	-15	8.8	-1.062	1.411	0.349
DET	75	79	722	733	-11	9.4	-0.381	0.419	0.038
CHI	67	83	683	814	-131	10.0	-0.771	-0.837	-1.609
BOS	63	86	700	758	-58	9.8	-0.601	0.031	-0.570
STL	55	96	669	820	-151	9.9	-0.912	-0.930	-1.842
Totals	604	604	6080	6080	0				
Avg			760	760	0	10.1			
Std Dev			99.8	64.5	113.8				

Because the Yankees were a strong team in a hitter's park, their league-leading 927 runs scored drive the standard deviation of RS to a value of nearly 100. The bottom three teams all played in relatively neutral home parks, so their runs allowed totals were not as extreme, and the standard deviation of RA is below 65. The significant teams here are Philadelphia and Cleveland. Indians pitchers basked in cavernous Cleveland Stadium, while Athletics batters enjoyed cozy Shibe Park. An average of 8.8 runs was scored in each contest involving Cleveland, while the average for games featuring Philadelphia was 11.4 runs – 30 percent higher. Because the SDS system separately evaluates runs scored and runs allowed, the Indians earn a positive SD Score, thanks to their "low" total of 669 runs allowed and the small standard deviation for RA. As a result, Cleveland ranks ahead of Philadelphia in the SDS reckoning. This is a preposterous assertion, totally contrary to the actual run data. The Indians allowed more runs than they scored, while the Athletics scored more than their opponents: there should be no debating which is the better team.

Simply put, the goal of a baseball team is to outscore its opposition. It makes no difference if this is accomplished via an outstanding pitching staff and a middling offense (e.g. the 2002 Atlanta Braves) or a tremendous offense and adequate pitching (the 2002 New York Yankees). The net effect is the same: more runs scored than surrendered. Considering RS and RA independently obscures the contexts in which the runs were scored and allowed. The 2001 Texas Rangers boasted an explosive batting order, but their woeful pitching staff demonstrated the fallacy of looking at only half the equation.

A more consistent approach is to use the standardized RS-RA difference, which ranks teams in descending order of run differential, and places the 1933 Athletics above their Cleveland competition. The perhaps controversial implication is that a method using run differential as its criterion considers teams that win 2-1 equivalent to teams that win 14-13: the margin of victory becomes crucial, not the run totals. Such an assertion obviously challenges those who are under the impression that an additional run saved is not equivalent to an additional run earned. This myth is easily dispelled (please contact me for a brief illustrative example), and in fact the RS-RA differential is actually a slightly better predictor of success than the RS/RA ratio. In 243 league-seasons from 1876 through 2002, the ratio predicts the highest winning percentage 179 times, for a success rate of 74 percent. The correlation coefficient is .940, which is outstanding. The predictions of the differential actually match those of the ratio a remarkable 230 times (both right and wrong), and it is correct in identifying the best winning percentage a total of 182 times, for a success rate of 75 percent. The correlation coefficient is .942, also outstanding, and marginally better. Given the choice, why not select the indicator which is more reliable and even more accurate?

Even if one is not convinced of the equality of runs, it is indisputable that the RS-RA differential is a better tool than the composite of runs scored and runs allowed, as it cuts through the park effects. Teams that score and give up equal numbers of runs fall exactly at the zero midpoint: their run differential is zero, and as the average of the differentials in any league is always zero, the standardized score for a truly average team is also zero. If the Rockies both score and allow 1200 runs this season, they should be considered a perfectly average team, regardless of the National League RS and RA means.

An equally thorny issue is league size. As the number of teams per league has fluctuated, the maximum attainable SD Scores have varied accordingly, subject to a very real expansion effect. From the standpoint of runs scored and allowed, the ideal situation for a first-place team is realized when all the other teams in the league score the same number of runs as each other and also allow the same number of runs as each other, while the leader scores more and allows fewer. One team with a superior differential in a league of identical opponents is the epitome of the best team ever – all the other teams are of equal ability, and one team is equally better than each of them. No team is "almost as good" as the front-runner, and no team is a "pushover." This occurs when the run differential for the first-place team is the negative of the sum of all the other equivalent differentials. The leader thereby becomes the *primus supra pares* – first above equals.

What becomes clear from analyzing such a situation is the peculiar nature of the maximum SD Scores and standardized differential. For instance, in a league with five teams, the maximum standardized RS, RA, and differential are all 2 exactly, or the square root of 5 minus 1. In a league with ten teams, these values are 3 exactly, or the square root of 10 minus 1. The proof that the maximum of both SDS component values and the standardized differential is the square root of one less than the number of teams is available from me on request – the calculations are quite simple.

With $\sqrt{N-1}$ as the highest attainable value, the only relevant constraint is the number of teams. A glance at the top and bottom teams of the past hundred years ranked by standardized run differential supports this. Looking at seasons 1901 through 2002, eleven of the top twenty teams (and the first four) are from 1997-2002 while nine of the bottom twenty (and the last four) are from the same period. In raw form, the data indicate that the

Table 2: Top and bottom 30 one-year Z/root scores, 1901-2002 (* indicates World Series winner)

1	1944	N	STL*	.812
2	1915	N	PHI	.795
3	1917	N	NY	.783
4	1927	A	NY*	.773
5	1902	N	PIT	.768
6	1925	N	PIT*	.754
7	1929	A	PHI*	.749
8	1941	A	NY*	.749
9	1914	A	PHI	.741
10	1986	N	NY*	.723
11	1984	A	DET*	.713
12	1913	A	PHI*	.709
13	1906	N	CHI	.709
14	1939	A	NY*	.702
15	1911	A	PHI*	.689
16	1922	A	STL	.687
17	1968	A	DET*	.678
18	1937	A	NY*	.677
19	1918	N	CHI	.672
20	1966	A	BAL*	.668
21	1905	A	CHI	.667
22	1958	A	NY*	.666
23	1942	A	NY	.665
24	1998	A	NY*	.664
25	1949	N	BRO	.662
1	1916	A	PHI	-.966
2	1938	N	PHI	-.898
3	1919	A	PHI	-.886
4	1943	A	PHI	-.844
5	1925	A	BOS	-.832
6	1908	A	NY	-.830
7	1954	N	PIT	-.828
8	1932	N	CIN	-.822
9	1915	FL	BAL	-.818
10	1939	N	PHI	-.811
11	1952	N	PIT	-.807
12	1961	N	PHI	-.798
13	1921	N	PHI	-.781
14	1951	A	STL	-.780
15	1973	N	SD	-.780
16	1940	N	PHI	-.780
17	1903	A	WAS	-.763
18	1910	A	STL	-.749
19	1963	N	NY	-.748
20	1926	A	BOS	-.745
21	1904	A	WAS	-.745
22	1945	N	PHI	-.745
23	1941	N	PHI	-.742
24	1945	A	PHI	-.739
25	1923	N	PHI	-.738

1997-1999 Yankees were obviously the best dynasty of the twentieth century – their three seasons hold spots one, three, and thirteen out of two thousand teams. It is no coincidence that interleague play began in 1997: in essence, each league now consists of thirty teams, and the maximum standardized differential is 5.385, up from 3.606 as recently as 1996. The run totals from games pitting American League teams against National League foes must be included since 1997, effectively adding a slew of teams to each league, each with eighteen games played. This emphasizes the value of using the differential: when half the teams only play one-ninth as many games (and consequently score and surrender one-ninth as many runs), the averages and standard deviations for runs scored and allowed are rendered meaningless.

Dividing each standardized differential by $\sqrt{N-1}$ adjusts for this expansion effect and bounds the results between –1 and +1. I have named the final determination the Z/root score, as Z is the statistical symbol for standardized value, and there is the necessity of dividing by the square root of $N-1$. As a sampling of the results, the top 30 and bottom 30 teams since 1901 are listed at the end of this article.

What of our original question – where do the 2001 Seattle Mariners fit in all this? Although they tied the Major League record for games won, they are not very high on the list, due largely to the abysmal performance of the Tampa Bay Devil Rays and the strong second-place finish of the Oakland Athletics. Considering all seasons since 1901, the Mariners are comfortably in the top one hundred as the 67th team with a Z/root of .575: just below the 1961 Yankees and a bit above the 1907 Cubs. The 1906 Cubs are 13th, while the 1998 Yankees are 24th. The top one-year team is the 1944 Cardinals, who dominated a depleted National League. Discounting teams from the World War II era, the 1915 Phillies are first, while the top World Series-winning team is the vaunted 1927 Yankees. The 1881 White Stockings do hold the distinction under this analysis as the best team in any major league since 1876, but they played only an 84-game season, and twelve decades later most of the players surrounding Cap Anson and King Kelly have faded into the shadows of history. This recent Mariners team was indeed special, but we should not let its dazzling record blind us to the truly greatest teams of all time.

Notes

One possible shortcoming of the Z/root calculation is its disregard for the unbalanced schedule. The 2001 Mariners played more games against the strong A's squad than against each foe outside the AL West, yet their Z/root evaluation assumes that all opponents have equal representation. I am uncertain how to make a suitable adjustment, which may or may not prove significant in the final tally.

A related rating system may be used to compare won-lost records instead of run totals. Although wins are more subject to luck and therefore less preferable than runs, this alternative method does present a different perspective. The analogue to run differential is win differential (games over .500), which accounts for varying numbers of games played by each team – especially useful in this interleague era. Similarly, the maximum standardized value is $\sqrt{N-1}$, and the adjusted results fall between –1 and +1. The 1941 Yankees rank first, as they finished 48 games over .500 against very balanced opposition. Using this method, the "real winning percentage" of the 2001 Seattle Mariners is .617, good enough for 58th place all-time and 47th since 1901. By contrast, the 1998 Yankees with a .704 RWP are 16th since 1901, and the 1906 Cubs at .697 are 18th. Sometimes winning isn't all it's cracked up to be.

Phil Melita is a music industry consultant who believes the 1986 Mets have been underrated. He can be reached at 1610 Crestwood Drive, Columbia, SC, 29205, preterosso@bellsouth.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
Jim Box	im.box@duke.edu	Statistics
Keith Carlson	kcarlson2@mindspring.com	General
Rob Fabrizzio	rfabrizzio@bigfoot.com	Statistics
Larry Grasso	l.grasso@juno.com	Statistics
Tom Hanrahan	HanrahanTJ@navair.navy.mil	Statistics
John Heer	jheer@walterhav.com	Proofreading
Keith Karcher	kckarcher@compuserve.com	General
Chris Leach	chrisleach@yahoo.com	General
John Matthew IV	john.matthew@rogers.com	Apostrophes
Duke Rankin	RankinD@montevallo.edu	Statistics
John Stryker	johns@mcfeely.interaccess.com	General
Dick Unruh	runruhjr@dtgnet.com	Proofreading
Steve Wang	scwang@fas.harvard.edu	Statistics