BY THE NUMBERS

The Newsletter of the Statistical Analysis Committee of the Society for American Baseball Research

Volume 1, Number 2 **August, 1989**

THE NEWS FROM SABR XIX

The 19th Annual SABR Convention in Albany, New York, was, well, interesting. I'm going to leave a report on the business meeting (and the internal politics) for the SABR Bulletin and concentrate on what we were able to do for the Statistical Analysis Committee.

Committee Meeting. The Committee met for about an hour, to discuss projects which the committee can undertake as a committee and to discuss the general direction of the committee.

Pete Palmer suggested developing an index of research articles using statistical techniques which can be made available to researchers. Andy Finn added the suggestion that the article index be organized around a series of "key words," indicating the primary topics discussed in the articles. Clem Conly agreed to try to develop a list of key words, which we plan to publish in the Newsletter as soon as it's available.

What this will require from you is that you look through publications which you have and identify which of the key word categories apply to the publication or to some part of it. Among the publications we hope to index are the Bill James Baseball Abstracts, the Baseball Analyst (the mimeographed publication edited by Bill James), the Bill James Baseball Abstract Newsletter (anyone out there still have copies of it?), the Elias Analysts, the Baseball Research Journal (published by SABR), and the Sabrmetric Review, as well as anything else we can find.

Once we publish the key words list, we hope you will submit your classifications of pieces in the following format:

Author, "Title," Publication, Publisher, Date, Pages.

Pieces which are a part of a larger work (for example, essays or team comments in the Abstracts or in the Elias Analysts) can also be classified in this way. Simply identify the key words which apply to the essay and identify it in the same format as above.

We hope to be able to distribute this bibliography in two formats. One would be a hard-copy form. If we are lucky, we might be able to have this become a regular SABR publication. The second, which could be updated annually, is on floppy disks as a database, in which the key words would be included. This would allow a user to search on a (set of) key word(s) and print out a special bibliography.

We hope you will participate in this project. For people doing research, this could be one of the most important things we can do as a committee.

It was also suggested that the Committee maintain an article file, so that research work which appeared in obscure (or currently out-of-print) sources can be made available to researchers. I think this is feasible, with the proviso that people requesting copies of articles pay for photocopying and postage. I will begin maintaining such an article file. If you have something or want to submit something for inclusion in the "library," let me know.

Finally, we discussed the possibility of developing an index of data files which could be available to researchers. Such an index could include the source, contents, and cost of the data file. For small data files, the Committee could even serve as a clearinghouse or distribution center. Again, people requesting data

files would be expected to pay for materials and postage.

Committee Membership. I distributed about 200 copies of the first issue of the newsletter, with the result that about 30 additional people have signed up with the Committee; we now have almost 60 people on the mailing list as Committee members. We also obtained additional material for the newsletter, so things are looking up for a regular publication schedule.

We're rolling now, but remember, we need your assistance and support if we are going to continue to make a contribution to SABR and to baseball research. Write for us, comment on what other people have written, make suggestions about what we can or should be doing. And, participate in the research article indexing project.

Don Coffin
Indiana University Northwest
3400 Broadway
Gary, IN 46408
(219) 980-6646

(Please note: This is a corrected phone number; I managed to type it wrong in the first newsletter. Oops.)

****

HOME PARK CORRECTION FACTORS
BY PETE PALMER

While doing the work for Total Baseball, I ran across a problem with Craig Wright's formula for adjusting park factors. The idea was that if the park you were testing was better for hitters than average, then all other road parks with which it was being compared were slightly worse than average. For example, if the runs scored per game at Fenway Park by both teams were 14% higher than the runs scored per game on the road in Red Sox games, the Fenway Park rating would not be 1.14. This is because the other 13 parks were only 0.99. Thus the true Fenway figure was only 1.13, as (1.13)/(0.99) = 1.14. Since the league average should be 1.00, that would mean that (1.13 + 13*0.99)/14 would equal 1.00.

The correction formula for park factor (PF) was

PF(corrected) = (PF)*[(NT-PF)/NT-1)},

where NT is the number of teams in the league. So, in the Red Sox case,

1.14*[(14 - 1.14)/(14 - 1)] = 1.13.

Now this seemed to work perfectly fine for run factors, where the range for unadjusted park factors is nearly always between 0.8 and 1.2, but when I did home run factors, I found a problem. In particular, in 1884, when Chicago NL hit seven times as many homers at home as on the road, the corrected park factor came out to

7*[(8 - 7)/(8 - 1)] = 1.

So I looked at the formula more closely, and, for eight-team leagues, the maximum value you could get for a corrected PF was around 2.3, when the initial (uncorrected) PF was 4. After that, the corrected PF went **DOWN** as the uncorrected PF went up.

Wright's original formula was derived by using the fact that the overall league rating should be 1.0. However, in deriving the equation for the corrected PF, an incorrect adjustment factor was used [(NT-PF)/(NT-1)]. The correct adjustment factor is [NT/(NT-1+PF)], where PF is the unadjusted park factor. Here are some examples, using an eight team league:

Unadjusted PF	Adjusted Park Factors Old	New
1.0	1.000	1.000
1.1	1.084	1.086
1.2	1.166	1.171
1.3	1.244	1.253
1.4	1.320	1.333
1.5	1.393	1.412
2.0	1.714	1.778
3.0	2.143	2.400
4.0	2.286	2.909
5.0	2.143	3.333
6.0	1.714	3.692
7.0	1.000	4.000

As you can see, it makes very little difference for run factors, which are pretty much limited to 1.2 or lower. However, there have been 18 parks with unadjusted home run park factors of 4 or more. Here there is quite a difference between the old and new adjusted park factors.

****

DON'T PUT THE CLUTCH IN YET

BY JORGEN RASMUSSEN

Cappy Cagnon's report in the first sue of By the Numbers on how superstars performed as pinch hitters is interesting, but I can't accept the argument that the figures prove that clutch hitters don't exist.

The question that immediately springs to mind is, "Why were guys like Cobb and Sisler sitting on the bench and available to pinch hit, rather than being in the starting lineup?" Perhaps they were being given a day's rest or it was the second game of a double header. In such instances, we would expect them to perform at their usual standards.

But suppose they were pinch hitting because it was early in their careers and they had not yet won a regular job. Or it may have been at the close of their careers when they were only occasional players. Perhaps they were just coming off an illness or injury and weren't ready for full-time duty. In all these cases, we would expect them to perform less well than their career figures.

It is not necessary to go through opy's entire list to develop this point. Cobb's 69 appearances as a pinch hitter, 48 were during his last four years. Of Hornsby's 86 appearances, 57 were during his last five years, in which he played a total of 113 games. (That means he appeared in only 56 games as a position player during these years.) Of Ted Williams's 111, 43 were during his last two years, as were 66 of Paul Waner's 164.

At the other end of a career, 37 of Babe Ruth's 67 pinch hit appearances were during his first four years, while he still only played as a pitcher. And all of Lou Gehrig's 16 pinch hit at-bats, obviously, occurred during his first three years, before he became a regular.

Beyond this, Cappy makes the questionable assumption that all pinch hit appearances are clutch situations. What about the pitcher who is getting batted around, but manages to complete an inning and is scheduled to bat during his team's half? Even if two hitters prior to him have been retired and no one is on base, he's going to get the hook. That hardly is a clutch situation with pressure on the hitter.

Sure, it's easy to compile such a list of figures from the Baseball Encyclopedia, but that doesn't justify using them inappropriately. When we can't get the situational information we need to make readily available figures relevant to an argument, then we simply have to remain silent, not claim that they prove something of which they are incapable.

Maybe Cappy is right in saying that clutch hitters don't exist. On this basis of this piece of evidence, however, the jury will have to remain out.

[Editor's note: In addition to the issues Jorgen raises, there are some sample size problems here--a small number of at-bats is insufficient to reach conclusions about ability--or lack of it. In addition, because the Baseball Encyclopedia provides information only on batting average--not on power, walks, runs-batted-in, etc.--the evidence base is pretty limited. Is there any way in which pinch hitting data can be used to address the issue of clutch performance? Think about it, look at it, and let me know what you find out.]

****

SHORTS I

Subscribe to the Baseball Analyst, to read what some of the best analysts are writing. It's (I think) $18 a year.

Baseball Analyst
P. O. Box 171
Winchester, KS 66097

Norman Macht, chairman of the Oral History Committee asks anyone interested in the oral history project to help out. When an Equitable Old-Timer's Game is held, ask the PR Director of the host club to allow you on the field to tape interviews with players. You could also try to schedule longer times for longer interviews away from the field. Send any completed interviews to Norm at his address in the membership book.

****

ELIAS SPLITS--OR WHY CORY SNYDER DOES NOT HATE TO FACE ROGER CLEMENS

BY ROB WOOD

Among the statistical minutia for which Elias is known, perhaps the most often discussed among baseball fans is their selection of pitchers that each batter loves of hates to face, and the hitters that each pitcher loves or hates to face. Paradoxically, these particular breakdowns are undoubtedly the least significant in a formal statistical sense of all their numbers. The reasons for this are a bit complex and go beyond the considerations of clutch performance to be discussed in the November Newsletter. The fact that Joe Hitter has gone 0 for 8 against Jim Pitcher may seem to be saying something about this particular match-up. However, it is more likely brought about by chance than first appears.

If one asks at the beginning of the year what are the chances that Joe will go 0 for 8 versus at least one pitcher during the year, the chances are indeed very high.[1] What is confusing the issue is that if we had instead asked at the beginning of the year what are the chances that Joe will go 0 for 8 versus Jim Pitcher, the chances are much smaller.

The distinction can be made more apparent by a hypothetical experiment. If 1000 people were each given a "fair" coin to flip eight times each, about four of them will get eight straight heads. Do we conclude that these four people are particularly adept at flipping heads? Obviously not. Now give each of 1000 people one coin (not necessarily a fair coin) and ask them to flip it eight times. If about four of the people get eight straight heads, we may be tempted to conclude that they were given two-headed coins. But this would be faulty reasoning. Instead, since the number of all-heads outcomes is what we predicted had all coins been fair, we cannot rule out the possibility that all coins were "fair." That is, we must test the entire distribution of coin flips, and judge whether or not it conforms to that which would result from random forces affecting the repeated flips of a fair coin.

The same reasoning should direct us to test the entire distribution of pitchers that Joe Hitter has faced during his career. We all know that pure chance will lead to bizzare splits in baseball statistics. The question becomes, given that we observe radical splits, do they exceed those which pure chance would produce? It is the analyst's chore to start with the "null hypothesis" that Joe Hitter has the same inherent ability to get a hit off of each and every pitcher in the league, and test whether or not we can reject this hypothesis.

Another issue that arises in this context of loves/hates to face is our standard of reference. Do we measure these relative to Joe Hitter's batting average across all pitchers, or how he does against particular pitchers compared to other hitters in the league? For example, suppose there were only two pitchers in the league: Nolan Ryan, off of whom opponents hit .200, and Bob Forsch, off of whom opponents hit .400. If Joe Hitter bats .300 off Ryan and .350 off Forsch, which should we say he "loves to face"? Most would agree [at least the Elias method of ranking would agree--Ed.] that it is proper in this instance to list Forsch as the pitcher than Joe loves to face.

Joe may even be listed as the batter Ryan "hates to face" and as the batter Forsch "loves to face." Both Forsch and Joe may "love" to face each other! The point I am raising is that there is the pitcher's side of the story that we usually ignore. Indeed, batting average is calculated in such a way that only the hitter matters. Although the opponents' batting average off of a particular pitcher appears in Elias, it is rarely discussed. Thus it has been implicitly assumed throughout baseball that batting average (and virtually every other offensive statistic) is nearly 100% batter-determined. In the February issue of the Newsletter, I will present a table of "batter vs. pitcher"-determined percentages.

---

1. The chances are virtually 100%. [Ed.]

Reverting the the "loves-to face/hates-to face" question, and using the above framework as a guide, I explicitly tested whether several players batter/pitcher match-ups were significantly different than could be expected by pure chance. First, I tested whether Eddie Murray (and three other hitters) has the same underlying batting average versus each and every pitcher he has faced in his career. By the nature of repeated random trials, we expect to observe figures like his 8-for-13 versus Vida Blue or his 1-for-14 versus Joe Cowley. But do they actually mean anything? By testing the underlying disrtibution (using a Chi-squared test), we cannot reject the hypothesis that Murray has the same underlying batting average versus each and every pitcher in the league. This result is repeated for all the hitters I tested.

We have just seen that a particular hitter is unaffected by the name of the pitcher (at least as far as formal statistical tests can tell). Let us now consider the reverse situation: Is a pitcher unaffected by the name of the batter? We all believe in a baseball sense that both answers are no, but let us stick to the formal statistical tests of the Elias break-downs.

I undertook the same test for Roger Clemens (and three other pitchers). The result is that we cannot reject the hypothesis that Clemens's opponents batting averages are identical for each and every hitter in the league. A second type of test is available as well. If one believes that hitters do have different inherent batting averages, one can test whether how each hitter does against a particular pitcher is significantly different from how he bats against the rest of the league's pitchers. I tested Clemens and the other three pitchers in this manner as well. The result is that we cannot reject the hypothesis that Clemens (et al.) is indistinguishable from the universe of other pitchers in the league--that is that Roger Clemens does not introduce a significant element into the outcome of an at-bat over and above the batter's intrinsic abilities.

Let us re-cap these results. I have not found one instance in which we can accept the hypothesis that a particular batter or pitcher introduces a significant element to an at-bat. Extreme splits are well within the bounds of pure chance. But this does not mean that all hitters or all pitchers are identical. Our baseball sense (as well as formal statistical tests) tells us that Wade Boggs is not identical to Rafael Belliard, and that Roger Clemens is not identical to George Frazier.

In addition, you may say that Murray and Clemens are known to be "consistent." Thus they are prime candidates to "pass" my test. That is exactly my point as well. Other players (both hitters and pitchers) who are known to exhibit large splits in some dimension would probably not pass the above tests. Consider a batter who murders right-handed pitching but is anemic versus left-handed pitching. We would assuredly reject the hypothesis that he hits identically against all pitchers. But if we divide the universe of pitchers into left-haders and right-handers, we would probably not be able to reject the hypothesis that he hits equally well against all right-handers and identically well against all left-handers.

I do find the Elias "loves/hates to face" methodology capable of leading to insights. It would be valuable information to know what subclass of pitchers a hitter loves/hates to face (and similarly for pitchers). Some would "love to face" right-handed non-power pitchers, or "hate to face" left-handed flyball-power pitchers. But we should take better advantage of our baseball sense, as well as the extensive body of formal statistics, to prevent the ludicrous "loves to face Jerry Reuss on grass fields during the day in August on odd-numbered days of the week" kind of analysis.

[Editor's Note: The problem lurking beneath the surface here is the problem of small sample sizes. For individual hitter-pitcher match-ups, even an entire career of data is unlikely to resolve the issue in a formal statistical sense. For example, Jim Palmer and Carl Yastrzemski overlapped for 15 years in the Major

Leagues. Following the beginnings of divisional play, they were in the same division. Both were durable, missing very few games. And Palmer was probably not held out of series in Fenway. And yet I calculate that Yaz probably did not bat more than about 300 times against Palmer in their careers. (We really, right now, don't know how many at-bats he had.)

[Suppose we know that Yaz hit .300 over his career against right-handed pitchers generally and that he hit over his career .350 (or .250) against Palmer. Statistically, these are not significantly different from each other. And, in a baseball sense, they probably aren't either. The difference between .300 and .350 (or .250) over 15 years and 300 at-bats is an average of ONE hit per year. That is unlikely to be anything but chance.]

****

PITCHER'S GAME SCORES: AN ALTERNATIVE MEASURE OF EFFECTIVENESS

BY MURRAY BROWNE

The Game Score statistic first appeared in The 1988 Bill James Baseball Abstract. Not only is it easy to compute (an important consideration for a borderline innumerate like me), but it is useful in illuminating the standard measures of potching performance--wins, losses, and ERA. It also allows us to assign a game-by-game numerical value to each outing by a starter. We compute the game score using the following rules:

1. Each starting pitcher starts with a score of 50 points.
2. Add 1 point for each hitter retired.
3. Add 2 points for each inning completed after the 4th.
4. Add 1 point for each strikeout.
5. Subtract 1 point for each base-on-balls.
6. Subtract 2 points for each hit.
7. Subtract 4 points for each earned run.
8. Subtract 2 points for each unearned run.

For example, the highly-publicized April 30, 1989, pitching duel betwen Nolan Ryan and Roger Clemens scores out as shown in the following table:

	Clemens	Ryan
	(50)	(50)
IP	9 (+37)	8 (+32)
H	6 (-12)	3 (- 6)
ER	2 (- 8)	1 (- 4)
UR*	0 (0)	0 (0)
K	6 (+ 6)	11 (+11)
BB	2 (- 2)	4 (- 4)
GS	71	79

*Unearned Runs

Both pitchers pitched well, but Ryan pitched better, and deserved a win.

Last year in the American League, Game Scores ranged from 96 to -2. Pitchers earned game scores of 90 or above in about 0.6% of the starts made in the AL. Tom Browing's perfect game scored out at 94; there hasn't been a game score in excess of 100 since Tom Seaver did it in May, 1974 (12 IP, 3 hits, 1 run, 16 Ks, 2 BB)--a score of 106.

On the other end of the spectrum are single-digit Game Scores--"batting tee performances". On August 27 and 28, New York's Charles Hudson and Tommy John tossed back-to-back single-digit Game Score games (-2 and 4). I don't know whether these types of outings reflect more on the pitchers or on the managers who leave them out there.

The idea of the Game Score is to zip through the box score to judge within seconds the performance of last night's pitchers. As a result, the analysis must be simple: If the pitcher's Game Score is 50 or above, the starter will probably get a win. A Game Score below 50 will probably lead to a loss. The Game Score of 50, which was used by Bill James in this fashion as a "dividing line," isn't an entirely arbitrary score. If you check the Game Score averages by league in the past three years, you will see that 50 is around the major league average:

	1987*	1988**	1989**
NL GS	49.40	52.83	52.84
AL GS	48.90	50.60	49.65

*Source: 1988 Baseball Abstract.
**Source: Calculations by the author. 1989 through 6/17.

If the GS of 50 is an adequate cut-off, we should expect to find that teams are much more likely to win when the GS is 50 or above and much less likely to win when the GS is less than 50. As the following table shows, this is what happens.

Game Score Range	W-L	Winning Percentage
90-100	14-1	.933
80-89	77-8	.906
70-79	197-40	.831
60-69	265-130	.671
50-59	287-212	.575
40-49	160-245	.395
30-39	85-213	.285
20-29	39-205	.160
< 20	6-72	.077

1988 AL Game Scores.

Although high seasonal average Game Scores for a pitcher are usually accompanied by low ERAs, this is not always the case. Some of the reasons are obvious: Whereas ERA is based on the number of earned runs allowed, Game Scores include the number of unearned runs. In addition, a pitcher might surrender a large number of hits or walks and "escape" with relatively few runs; Game Scores capture that, and, additionally, suggest that you can't be lucky forever. The following table lists the top 10 seasonal average Game Scores among AL starters in 1988. Of these 10, eight were also in the top 10 in ERA. (See top of next column.)

Oddly enough, last year's AL ERA leader, Minnesota's Allan Anderson, did not make the top 10 in Game Scores. (The other pitcher in the top 10 in ERA who did not make the top 10 in Game Score was Charlie Liebrandt.) If we compare Anderson to Higuera, we can begin to understand some of the differences in what Game Scores and ERA tell us.

Pitcher	Average Game Score	ERA
Higuera	63.70	2.45
Clemens	63.63	2.93
Robinson	60.96	2.98
Gubicza	59.31	2.70
Viola	58.90	2.64*
Langston	58.34	3.34*
Steib	57.97	3.04
Swindell	57.42	3.20*
B. Witt	56.96	3.92*
Stewart	56.95	3.23

*Not in top 10 in ERA.

The difference between Higuera and Anderson in ERA was virtually non-existent, but Higuera (63.70 in 31 starts) led the league in average Game Score, while Anderson (55.43 in 30 starts) did not finish in the top 10. What were the differences? Higuera, in 25 more innings, allowed 30 fewer hits; he struck out over 100 more batters. Higuera had more innings per start (7.33 compared to 6.73).

Finally, Anderson gave up 15 unearned runs; Higuera only gave up four. You have to wonder whether Milwaukee (a team which committed a nearly league-leading error total) just didn't make any errors when Higuera pitched) or whether Minnesota (the team commiting the fewest errors) suddenly became butter-fingered when Anderson pitched. Or perhaps Higuera just "pitched through" his team's fielding miscues.

Other pitchers present equally interesting contrasts. Oakland's Dave Stewart and Texas's Bobby Witt have virtually identical Game Scores (56.95 and 56.96), although Stewart's ERA is 0.69 lower (3.23 compared to 3.92). I did a double take on this initially, since Witt did not seem to belong at the same level as a two-time 20-game winner. And Witt did have an awful beginning (38.3 average game score in his first six starts). But after his return from Oklahoma City, he pitched very well, with an average 63.94 Game Score in his last 16 starts. By the end of the season, Witt averaged, per nine innings, more strikeouts and fewer hits allowed than did Stewart.

Game Scores also allow us to look at "Cheap Wins" and "Tough Losses." Borrowing again from the 1988 Baseball Abstract, we can identify cheap wins--wins with a Game Score of less than 50--and tough losses--losses with a Game Score of 50 or greater. In the AL in 1988 there were 133 cheap wins and 252 tough losses.

Again, we can find some interesting results when we look at individual pitchers. Charlie Hough (15-16) and Jose Guzman (11-13), both with Texas, both had losing records with ERAs of 3.32 and 3.70 respectively. I usually find it hard to tell whether a pitcher's record is worse (or better) than he deserves just by looking at ERA. But by looking at Hough's average Game Score (56.79, compared with a league average of 50.60) and Guzman's Game Score of 54.17, I find it easier to believe that both pitchers were victims of poor and untimely run support. Hough led the AL with nine tough (Game Score) losses; he had no cheap wins. Based on Game Scores, Hough might have "earned" a record of 24-7. Guzman, with five tough losses and one cheap win, has an "adjusted" record of 15-9.

Incidentally, Texas starters had the best average Game Score on the season (52.55). Oakland, the league ERA leader, finished fifth in average Game Score at 52.09. No wonder Texas traded for hitting in the off-season.

Boston pitchers Roger Clemens (18-12) and Bruce Hurst (18-6) also serve as examples of how Game Scores can be used to understand better the nature of a pitcher's season.

Last year, Clemens tied for the highest single Game Score of the year (96); he had 12 starts with Game Scores in excess of 80. Clemens had six tough losses and no cheap wins, so if you adjust his record as we did for Charlie Hough, it looks more like a 24-6 season, which might have been good enough for a third Cy Young Award, even with Frank Viola and Teddy Higuera around.

Hurst, on the other hand, seems to have been lucky in 1988. Boston averaged more than 6 runs per game in his starts; he had seven (!) cheap wins and only four tough losses. That's 5.3% of the cheap wins in the entire league. (One of his cheap wins came on a game score of 22--11 hits and 7 runs--but Boston scored 11).

Sometimes Game Scores will uncover a decent pitcher with a poor ERA coupled with a losing record. Such is the case with Chuck Finley (9-15 and 4.17) who led the California staff with a seasonal average Game Score of 50.48. Although this doesn't say too much good about the Angels' rotation, Finley was doing something right, and he has done even better in 1989. He's among the leaders in ERA and has an average (1989) Game Score of 60.64 in his first 11 starts.

[Editor's Note: This could be the beginning of an interesting dialogue. Do Game Scores help us understand pitcher performance better? Are they subject to park effects? Or is it an analytical tool which doesn't add much to our knowledge?]

****

WINNING STREAKS, LOSING STREAKS, AND PREDICTING FUTURE TEAM PERFORMANCE

BY KEITH KARCHER

[Editor's Note: In the July, 1989, issue of Sport magazine, Danny Sheridan discussed various "systems" for winning bets on baseball games. Since any betting system is simply a method of predicting when a team is more likely to win, it is plausible to examine any betting system simply as a prediction system. Sheridan wrote that his favorite system is one in which a bettor bets on a team to win only when it has won three or more games in a row; the bettor bets on the team to lose when it has lost three or more games in a row.]

There is a statistical method which, I believe, allows us to test for the streakiness of a baseball team. It is the "runs test" which is used in regression analysis to determine if there are too many or too few "runs" in the signs of the residuals. If positive (or negative) residuals tend to occur together, there will be relatively few, but relatively long, streaks of positive (or negative) residuals. On the other hand, if positive residuals tend to

be followed by negative residuals, there will be relatively many, short streaks of positive (or negative) residuals. The same reasoning can be applied to wins and losses. For example, we can look at the first 21 games of the Los Angeles Dodgers 1988 season, grouped by streaks:

L-WWWWW-L-W-L-WW-L-WWW-L-W-L-W-LL

The Dodgers began the year with a one-game losing "streak, then won five in a row, lost one, and so on. Altogether, they won 13 times, lost eight, and had a total of 13 streaks. Now the question is this: Is 13 streaks "too many" (indicating the Dodgers tend not to have long winning or losing streaks), "too few" (indicating the Dodgers do tend to have long winning or losing streaks), or not different from the expected number of streaks, given 13 wins in 21 games? With only 21 games, we could generate every permutation of 13 wins and eight losses, then count up the number of streaks in each permutation. From these totals we could determine exactly the probability of having 13 or more streaks.

As you might imagine, when the totals get up to 161 games and 94 wins, compiling all the permutations can get a bit cumbersome. In these situations we must rely on the normal approximation to the exact distribution. To do this we need to be able to calculate the mean number of streaks and the variance in the expected number of streaks. For this problem, the required formulas are knwon; they are given in an appendix at the end of this article, along with the procedure for making a statistical test.

The following table gives, for each National League team in 1988, the team's record, the number of winning streaks, the number of losing streaks, the total number of streaks, and the expected number of streaks. The data for the number of streaks are from The Sporting News Baseball Guide.

Only one team had a number of streaks which is outside the normal range of variation, and that one team is, by a remarkable coincidence, the Cincinnati Reds. The Reds had 97 winning or losing streaks, whereas we would expect a team with their record to have only about 81 streaks.

That indicates that the Reds were much less likely to have long winning or losing streaks than any other team in the NL.

			Number of Streaks			
Team	W	L	W	L	Tot	Exp
LA	94	67	39	40	79	79
CIN	87	74	49	48	97	81
SD	83	78	41	41	82	81
SF	83	79	41	40	81	82
HOU	82	80	39	39	78	82
ATL	54	106	37	37	74	73
NY	100	60	35	34	69	76
PIT	85	75	39	39	78	81
MON	81	81	39	39	78	82
CHI	77	85	45	44	89	82
STL	76	86	40	41	81	82
PHI	65	96	37	38	75	79

Put together, the evidence so far indicates that baseball teams are not any more "streaky" than we would expect. Wins and losses do not seem to gather together in any manner out of the ordinary. Yet Sheridan seems to indicate that there is something special about "3 in a row" winning or losing streaks. So back to 1988 we go and this time we'll just look at the NL West. The following table shows, for each team in the NL West in 1988, their records following three-game winning streaks and three-game losing streaks.

	Following Three-Game Winning Streaks		Following Three-Game Losing Streaks	
Team	W	L	W	L
LA	19	13	9	0
CIN	8	7	4	7
SD	12	8	9	9
SF	0	10	9	11
HOU	10	11	9	6
ATL	0	5	16	30
Total	59	53	56	63

Only two teams (LA and San Diego) had better winning percentages after winning three games than they did overall. However, neither of these improved winning percentages were statistically significant. Three teams had lower winning percentages after a three-game losing streak

than they did overall (Cincinnati, San Diego, and San Francisco). In these three cases as well, the difference in the team's winning percentage after three losses and overall is not statistically significant. Thus, it would appear that teams are no more or less likely to win games following a three-game winning streak, and no more likely to lose a game following a three-game losing streak, than they are in any other situation.

If we used Sheridan's "streak" system for betting baseball games, we would not do very well. For teams coming off winning streaks, we would win 59 bets and lose 53; for teams coming off three-game losing streaks, we would win 63 bets and lose 56. Our overall record would be 122 bets won and 109 bets lost. In order to break even, none of our losing bets could be in games in which we had to wager more than $5.50 in order to win $5.00--in other words, all our losing bets would have to be in "pick-em" games. Based on all the analysis here, I believe that anyone using Sheridan's "streak" method for betting baseball games is likely to lose.

APPENDIX: The formulas for the mean (μ) and variance (σ^2) of the distribution of the number of win streaks are:

1. Mean (μ) Number of Streaks:
$$\mu = (2WL)/(W+L) - 1$$
2. Variance (σ^2) in the Number of Streaks:
$$\sigma^2 = [2WL(2WL-W-L]/[(W+L)(W+L-1)]$$

where W is the number of wins and L is the number of losses.

The test statistic is the random normal deviate Z, given by
$$Z = (S - \mu \pm 0.5)/\sigma$$
where S is the number of winning or losing streaks a team actually has and μ and σ are calculated as above (σ is the square root of the variance). We add or subtract 0.5 in order to correct for continuity because we are approximating a discrete distribution by a continuous one. To test for too few streaks, add 0.5; to test for too few streaks, subtract 0.5. The value of Z obtained here is compared with critical values of Z in a table showing Z-values for the normal distribution to determine the probability of a greater or lesser value.

****

SHORTS II

Data Request: For a new research projet, I need a file of annual attendance, by team, for the 1950-1988 period. I could also use information on ticket prices, by team, for the same period. If anyone has, or knows of, such a data set, could you please let me know? Don Coffin, Indiana University Northwest, 3400 Broadway, Gary, IN 46408, 219-980-6646.

Team Newsletters. Team newsletters are frequently of very little interest unless one is a fan of the team involved. Many of you might be interested in the Philadelphia Baseball File, 1510 Harrison St., Philadelphia, PA 19124 ($20 for four issues). I have seen Vol. 1, Nos. 3&4 and they are quite good. The work they are doing on defensive statistics (using Project Scoresheet data) will, I think, make a real contribution to our understanding of defensive play.

If anyone knows of other newsletters which may be of interest to our membership, please send a copy of one (or comments about it) to me.

****

IMPROVED OFFENSIVE PERFORMANCE
BY DON COFFIN

Living in Chicago, I get to read a lot about the Cubs and the White Sox. So when something happens (or fails to happen) to one of these two teams, I really get to read a lot about it. For almost the entire season, there has been weekly comment on the improvement in the White Sox offense, particularly as measured by team batting average. When the Chicago Tribune wrote its mid-year review of the Sox, a large part of it was devoted to the improvement in the White Sox team batting average and how happy GM Larry Hines has been to see it.

So I decided to look a little more closely at the improvement. It is real improvement, by the way. The table below presents four measures of team offense--runs per game (RPG, which is probably the most important of the four measures), team batting average (BA), team isolated power (slugging average minus BA, ISOP), and

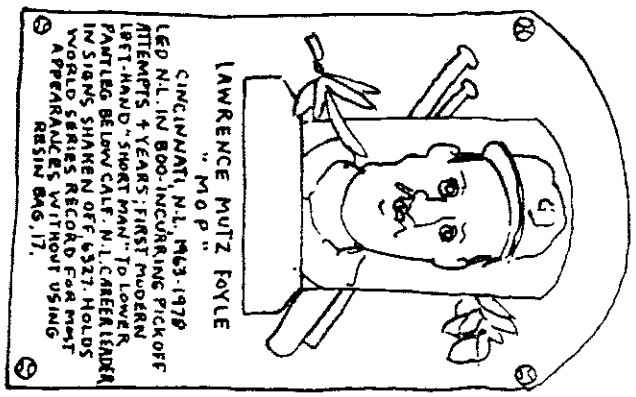
team walks per plate appearance (WPA). The White Sox rank in each category is shown in parentheses; the league figures are for the other 13 AL teams (excludes the White Sox). (Data as of 7/16/89.)

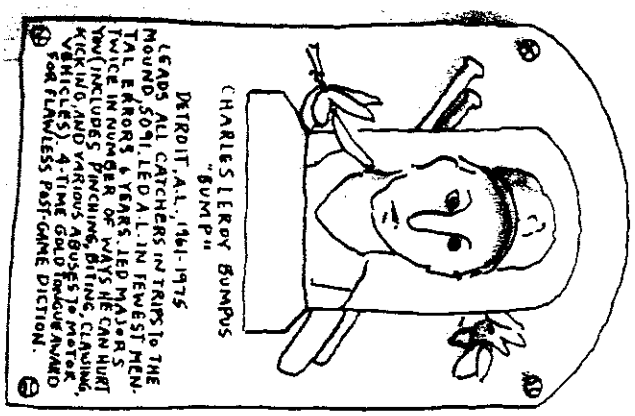
Category		1988	1989
RPG:	Sox	3.92 (13)	4.21 (10)
	League	4.39	4.41
BA:	Sox	.244 (13)	.270 (4)
	League	.260	.263
ISOP:	Sox	.127 (9)	.115 (12)
	League	.132	.124
WPA:	Sox	.074 (12)	.076 (13)
	League	.085	.087

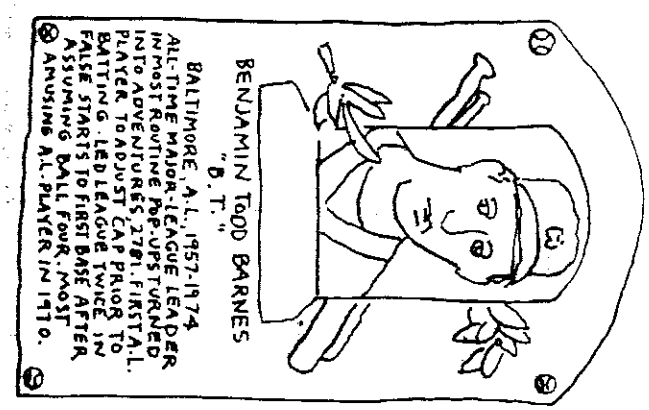
The Sox improvement in BA is remarkable--from 1.5 standard deviations below the league average to a standard deviation above the league average. The improvement in RPG is also quite noticeable--from 11% below the league average to 4.5% below the league average. Yet so far in 1989, the Sox are still only 10th in RPG. Why hasn't the improvement in BA had more of an effect on their RPG?

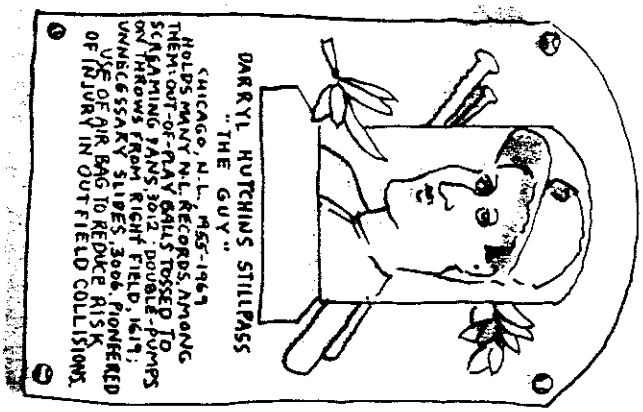
The answer is quite clear--while power is down a bit across the league (ISOP is down by 6% for the other 13 teams), the decline in Sox power is even more marked--almost a 9.5% decline in ISOP. And, while the Sox WPA is up slightly, it is down relative to the league. It appears that the increase in BA for the White Sox has been at the expense of power and that there has been no improvement in the team's ability to draw walks. So the modest increase in RPG becomes quite understandable.

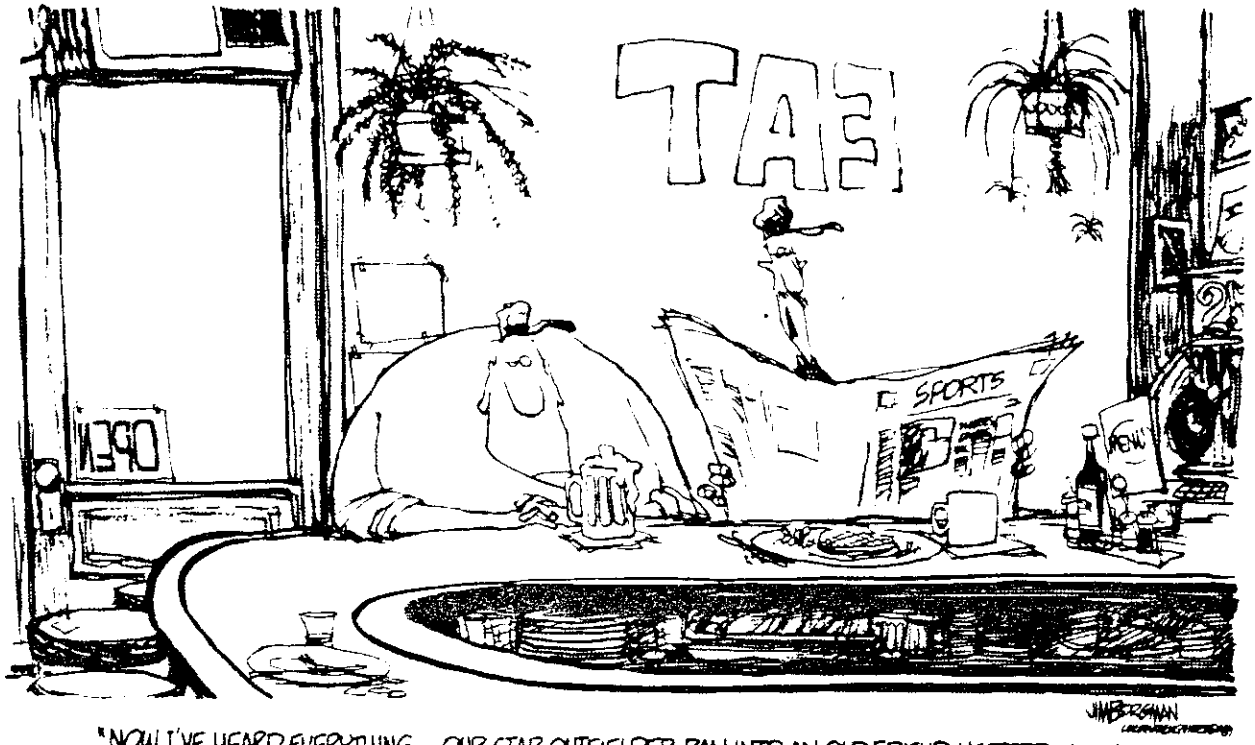
NEW INDUCTEES











"NOW I'VE HEARD EVERYTHING.... OUR STAR OUTFIELDER RAN INTO AN OLD FRIEND YESTERDAY, AND WAS PLACED ON THE 21-DAY DISABLED LIST."

Donald A. Coffin
Indiana University Northwest
3400 Broadway
Gary, IN 46408