By the Numbers

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

Volume 1, Number 3 December, 1989

COMMITTEE NEWS

An Apology. As you know, I had planned for this issue of By the Numbers to come out in November. I had forgotten that no one in college teaching ever gets anything done in November or December, so here, belatedly, is the December, 1989, issue of By the Numbers.

Publication Schedule. I have planned the 1990 publication schedule for By the Numbers as follows:

Vol. 2, No. 1: January.
Vol. 2, No. 2: March
Vol. 2, No. 3: June
Vol. 2, No. 4: September
Vol. 2, No. 5: December

SABR In Cleveland. By the March issue of the newsletter, I expect to know when the Statistical Analysis Committee will meet at the SABR convention in Cleveland. I hope many of you will be there. Also, I volunteered the Statistical Analysis Committee as a sponsor for a research session devoted to statistical analysis. I expect that we will be able to have five presentations in the two-hour session, leaving time for questions and discussion. If you would like to make a research presentation in this session, please send me a brief (one-page) outline of your research by March 31. I will announce the schedule for the session by April 1 in a special mailing to the committee.

March Newsletter. I continue to need material for future issues of the newsletter. If you are working on something which you think may be of interest to the other members of the committee, the deadline for inclusion in the March issue of the newsletter is February 28.

Donald A. Coffin
Division of Business and Economics
Indiana University Northwest
Gary, IN 46408
219/980-6646

****

MEASURING PITCHER PERFORMANCE

By Pete DeCoursey
Editor, The Philadelphia File

We ultimately measure baseball teams by one criterion: How many games have they won? So when we rate major league pitchers, it is very tempting to take their won-loss record as the best measure of their performance. It isn't; a pitcher's job is to keep people off the basepaths, and to keep baserunners from scoring. Measuring those aspects of performance will tell you how well he has performed. Won-Loss records depend not only on the pitcher, but on his teammates.

Let's use Shane Rawley, former Phillies staff ace as an example. Rawley was 17-11 in 1987 and 8-16 in 1988. Let's look at his actual performance in those years.

Performance Category	1987	1988
Games	36	32
IP/Start	6.4	6.2
Hits/9IP	9.8	10.0
BB/9IP	3.4	3.5
HR/9IP	0.9	1.3
ERA	4.39	4.18

Rawley reduced his ERA, while giving up about four more hits, two more walks, and six more home runs over the course of a full season. No matter where you place them, four more hits, two more walks, and six more home runs won't turn a 17-11 record into an 8-16 record, by themselves.

So if Rawley allowed hits and walks at about the same rate in 1987 and 1988, and allowed fewer earned runs per start, why did his record drop so sharply? Because the Phillies gave him fewer runs to work with! The Phillies gave Rawley 4.94 runs per game in 1987 and 3.84 runs per game in

1988 (source: Elias Baseball Analyst). Shane performed at about the same level in both years, he just had more runs to work with in 1987, so he won more games than ever before or since, without significantly changing how the other team hit him.

The difference between a starting pitcher's run support and his ERA is what I call that pitcher's Win Margin. For Rawley, that would be a Win Margin of +0.55 in 1987 (run support of 4.94 minus an ERA of 4.39) and a Win Margin of -0.34 in 1988 (run support of 3.84 minus ERA of 4.18). That means Rawley slipped from being half a run better than his opposition to being one-third of a run worse. That explains the change in his record.

Win Margin is the best and simplest indicator of a starting pitcher's effectiveness for two reasons. First, it measures how much he exceeded or was exceeded by his opponents. Second, it allows for all of the situational factors which can affect a pitcher's record. If your starter pitches in a ballpark in which a lot of home runs are hit, his opponents faced the same problem; if it's a hot night and he tires quickly, the thermometer puts his opponent under the same stress.

Win Margin compares how your ptticher fared against his opponents, working under the exact same conditions, on the same nights, in the same parks. It isn't just a measure of offensive support. It measures every context that affected a pitcher's performance by comparing him to his opponent in that set of contexts. Numbers measure more accurately when the comparisons are more carefully drawn. You can't draw a closer comparison than the man on the mound for the other team. That's the guy you have to beat.

While just under four runs per game did little for Rawley, look what a similar level of offensive support did for Orel Hersheiser. Hersheiser is a good example of how a terrific pitcher helps his team win. He received only 4.02 runs per game of support from his batters in 1988. But by holding his opponents to only 2.26 runs per game, his Win Margin was 1.76, and he made 4 runs of support per game sufficient for 23 victories. Nineteen NL pitchers received more run support than did Hersheiser in 1988. None won more games.

Let's examine the 13 pitchers who won 18 or more games in 1988. RS (Run Support) refers to how many runs the pitcher's team scored in the games he started; WM (Win Margin) is RS minus ERA. (Run support numbers are taken from the 1989 Elias Baseball Analyst.)

Pitcher	RS	ERA	WM	W-L
Viola	5.29	2.64	+2.65	24-7
Hersheiser	4.02	2.26	+1.76	23-7
Jackson	4.60	2.73	+1.87	23-8
Stewart	4.76	3.23	+1.53	21-12
Cone	4.64	2.22	+2.42	20-3
Gubicza	4.94	2.70	+2.24	20-8
Reuschel	4.44	3.12	+1.32	19-11
Browning	4.94	3.41	+1.53	18-5
Hurst	6.00	3.66	+2.34	18-6
Maddux	4.18	3.18	+1.00	18-8
Gooden	5.59	3.19	+2.40	18-9
Clemens	4.37	2.93	+1.44	18-12

The common denominator of a big win season is clear: If your Win Margin is in the +1.3 to +1.5 run range, you are going to win a bunch of games. It's also clear that a merely average season like Bruce Hurst's can be made to look much better if your teammates score tremendous numbers of runs when you're on the mound. Roger Clemens, Hurst's former teammate, had to win his 18 with more than one run per game less to work with. If you switched their run supports, then Hurst would be back to his traditional 13 wins or so and Clemens would have won close to his Cy Young total of 24. Since both pitchers pitched for the same team (in 1988) in mostly the same parks, their ERAs are directly comparable, and Clemens is 0.75 runs better.

That advantage is compounded by the durability of Roger Clemens. He averaged 7.54 innings per start, whereas Hurst averaged 6.58. That means the Red Sox had to use more than two relief innings per start for Hurst, but just more than one relief inning per start for Clemens. More innings of fewer runs per game is obviously the sign of the better pitcher.

So, in 1988, Roger Clemens pitched more games than Hurst, more innings per game, and allowed fewer runs per game. Despite the fact that Hurst had as many wins and a better winning percentage, Clemens was by far the better pitcher.

This happens fairly frequently. Let's pick two other pairs of teammantes: the Mets' Dwight Gooden and Bob Ojeda and the Cubs' Greg Maddux and Jamie Moyer and compare their records in 1988.

	Gooden	Ojeda	Maddux	Moyer
Games	34	29	34	30
IP	248	190	249	202
ERA	3.19	2.88	3.18	3.48
RS	5.59	2.97	4.18	3.27
WM	+2.40	+0.09	+1.00	-0.21
W-L	18-9	10-13	18-8	9-15

In both these cases, if you traded run support, it's very likely that you would switch records. Both Maddux and Gooden pitched more innings per start than did Moyer or Ojeda, which is one reason their ERAs are higher. But the big difference is in the Win Margin column: If you hold the opposition to a run or a run-and-a-half less than your teammates score for you, your're a Cy Young candidate.

If I could have only three numbers to rate starting pitchers, I would choose ERA, innings per start, and Win Margin. Win Margin measures whether he beat his opponent; IP/Start measures how long he pitches; and ERA measures what he gave up. Comparing ERA to Win Margin makes sure that the Ojedas and Moyers and Clemens are not cheated by the greater offensive support enjoyed by their teammates. Ultimately, however, a pitcher's job is to outduel his opponent, and that's what Win Margin measures.

One drawback to Win Margin is that, as a brand new statistic, it is not published in any newspapers--yet. Before 1989, no one published home-road or left-right stats either. So if you want to see Win Margin every week in the daily papers, let them know.

****

CAN MITCH WILLIAMS DO IT AGAIN?

By Don Coffin

One of the major surprises of the 1989 season was the performance of Mitch Williams for the Chicago Cubs. Going into the season, he had a lifetime record of 18-19, 30 saves, and a 3.70 ERA. He proceeded to turn in a 36 save, 2.64 ERA season. Can he do it again?

As Pete DeCoursey points out in the preceeding piece, one of a pitcher's main functions is to prevent runners from reaching base. One of Mitch Williams's problems throughout his career has been that he has not succeeded, very well, in that endeavor. The following table shows his numbers.

	1986	1987	1988	1989
IP	98.0	108.7	68.0	81.3
Hits	69	63	48	71
Walks	79	94	47	52
H/9IP	6.3	5.2	6.4	7.9
BB/9IP	7.3	7.9	6.2	5.8
*BR/9IP	13.6	13.1	12.6	13.7
HR/9IP	0.7	0.7	0.5	0.7
ERA	3.58	3.23	4.63	2.63
**R/9IP	3.58	3.89	5.03	2.98

*Baserunners per 9IP.
**Total runs allowed per 9 IP

His 1989 performance was, in some ways, his worst, not his best. He gave up more hits per nine innings, and hits tend to be more damaging than walks, than before. Following three years of reducing his baserunners per nine innings, he experienced a sharp increase. Yet he cut his ERA by 2.00 and his total runs per game by 2.05 (compared to 1988).

Williams's runs allowed and saves were startling, based on his record of allowing runners to reach base. I looked at all pitchers who pitched between 60 and 100 innings in 1989 (Williams pitched 81.3), and found the following relationship between baserunners per nine innings and total runs allowed per nine innings:

$R/9IP = 0.346*8R/9IP \quad R^2 = 0.59$
(3.43)

In running this regression, I forced the constant term to equal 0 (since we would expect that a pitcher who allows no baserunners should allow no runs). Based on this regression, we would have expected Williams to allow 4.70 runs per nine innings, rather than the 2.98 he did allow. Based on the 81.3 innings he pitched, this means he surrendered about 15 fewer runs than we would have expected.

Williams allowed more baserunners per nine innings than all except one other pitcher who pitched between 60 and 100 innings--Rick Rhoden, who allowed 13.87 baserunners per nine (with 2-6 record and a 4.28 ERA). I think it is fairly clear that the Cubs cannot count on a similar set of outcomes in 1990--36 saves and a 2.64 ERA--unless Williams manages to cut down on the number of baserunners he allows.

****

CLUTCH ABILITY: MYTH OR REALITY?

By Rob Wood

There are a couple of important distinctions that should be kept in the forefront of the discussion pertaining to "clutch." The first is the difference between clutch performance and clutch ability. We observe a player's performance in the clutch, and hope that this is a true and meaningful indicator of underlying ability. There is no doubt that clutch performance exists, but we need to investigate whether or not it is explainable solely by pure chance.

The second distinction is between pressure and importance of a situation. Pressure is felt by the batter, pitcher, fielders, announcers, fans, etc., at particular moments in particular games. Importance is what the analyst can say were crucial moments that determined the ourcome of a game--after the game is over. Pressure is almost always a late inning phenomenon, whereas turning a routine double play in the top of the first inning may prove to be the most important play of the game. Both are valid concepts in sabrmetrics. However, fans and analysts alike usually refer to "clutch" situations as "pressure" situations. For this reason the game-winning RBI statistic, which pertains more to importance than to pressure, went over like a lead balloon--and has been scrapped as an official statistic of major league baseball.

Whether or not clutch ability actually exists has been a topic of lively debate and serious statistical research for over 25 years. Dick Cramer and Pete Palmer analyzed the 1961-1971 baseball seasons in search of clutch ability. Using a variety of methods I will discuss below, they concluded that without question clutch ability does not exist.[1] But even with the reputations of these men, the issue is still with us--partly due to Elias publishing the "clutch performance" of each batter in their annual Analyst.

There are two ways to attack this issue. The first is to discuss what we mean by clutch ability and show that it is not manifested in the statistics. The second is to present the formal statistical significance tests of clutch splits. Cramer and Palmer dutifully employed the first method. Seeing that they were unsuccessful in eradicating clutch ability from our baseball lexicon, I will employ the second method. Hopefully the weight of the two methods will crush "clutch ability" once and for all.

It is not difficult to summarize the earlier Cramer-Palmer study, and virtually every other statistical analysis of the topic. If clutch ability did exist, we would expect it to show up in at least two places: first, reputable "clutch hitters" (for some reason clutch ability is assymetric in that pitchers are never thought to possess it) should perform well in "clutch" statistics; and, second, hitters should be good or bad in the clutch to a similar extent from year to year. Studies have found neither to hold to a significant degree, and the second criterion is universally violated.

1. Richard D. Cramer, "Do Clutch Hitters Exist?" Baseball Research Journal, 1977, pp. 74-79.

What has not been properly understood is the very large splits that would have to be achieved within a season for a batter to have demonstrated clutch ability to a degree which exceeds the likelihood that pure chance could have caused them. I do not propose to lecture the reader on the meaning of "statistical significance." Everyone has at least a grasp of what the concept attempts to address. Instead I will simply present the following table which lists the amount a player's clutch batting average must exceed his non-clutch batting average for the difference to be statistically significant at the 95% level (which means that pure chance would lead to such a split less than 5% of the time.)

Clutch Versus Non-Clutch Performance Differences Required for Performance Differentials to be Statistically Significant

At-Bats	Non-Clutch BA	Required Clutch BA	Required Differential
300	.200	.356	.156
300	.250	.418	.168
300	.300	.477	.177
400	.200	.335	.135
400	.250	.395	.145
400	.300	.453	.153
500	.200	.320	.120
500	.250	.380	.130
500	.300	.437	.137
600	.200	.310	.110
600	.250	.368	.118
600	.300	.425	.125

The table assumes that 10% of all at-bats are "in the clutch," which roughly corresponds to the Elias definition of "late inning pressure situations with runners in scoring position." Ultimate clutch at-bats (e.g., two-out, winning and/or tying runs in scoring position in late innings) are much rarer indeed. I use the 10% assumption in order not to be stacking the deck too much at the outset.

The table is easy to understand. For example, the first row states that a .200 hitter in 300 at-bats would have to hit .356 before his clutch performance differential were deemed statistically significant--a whopping differential of 156 points of batting average. The other entries are similar. The effect of the various factors is self-evident. The fewer the number of at-bats, the larger the differential would have to be. Similarly, the better the hitter, the larger the differential would have to be. If clutch situations were defined more liberally (say, 15% of all at-bats), the differentials in the above table would be smaller.

If we examine the Elias splits, we would find that very few players would qualify as clutch hitters using this formal statistical framework. (Indeed, using a 95% significance level implies that roughly 5% of all hitters would be so classified is clutch performance were due entirely to random factors and ballplayer's abilities were "normally" distributed.) Perhaps even more persuasive is that the names of the clutch performers seem to be randomly chosen, having little to do with our preconceptions of what would make up a good clutch hitter. The final kicker is that virtually no player has net the formal statistical criterion in back-to-back seasons. Eddie Murray is one of the few reputable clutch hitters whose clutch performances pass the formal statistical tests.

Cramer and Palmer remind us that we do not know what goes into clutch performance. We hope we are somehow measuring the ability of a player to rise above the competition at the most crucial moments. When the game is on the line, who can perform the best (even better than their norms) and who has a tendency to "choke"? However, we may be measuring something else entirely. Some batters have the knack of only bearing down when the game is on the line, or a runner is in scoring position, etc. Should we really call such an observed differential between pressure performance and non-pressure performance "clutch ability," or rather "extreme

laziness and/or selfishness that has no place in team sports"? Jack Clark and Perdo Guerrero are but two players who have been accused of this over the years.

It is my view that clutch performance is actually a measure of each particular player's weakness as a hitter. If a batter has a specific "blind spot" (e.g., cannot lay off the high-and-tight fastball), then in almost all crucial situations, the pitcher will pitch to that weakness (and the opposing manager will make sure to have a pitcher who can take advantage of that weakness). Another way of looking at this is to consider a clutch at-bat as being when a pitcher is trying his utmost to strike out the batter. Eddie Murray is a switch hitter with good command of the strike zone, who can foul off a pitcher's best pitch, and who will not get called out on strikes on the borderline pitch. Thus he will perform very well "in the clutch."

****

BETTING ON STREAKS

By William Nolan

(In the last issue of By the Numbers, Keith Karcher wrote an analysis of the probable value of using recent winning or losing streaks as a basis for making bets on baseball games. He showed statistically that such an approach is not likely to be a winner. In this issue, William Nolan reports the results of his tracking of hypothetical bets based on the streak system. -Ed.)

Danny Sheridan's recommendation that you bet on teams on winning streaks of three games or more, and bet against teams on losing streaks of three games or more can be tested against the actual game results. Using games from after the All-Star break for the 1989 season, I kept a record of hypothetical bets placed on teams with three-game or longer winning streaks and against teams with three-game or longer losing streaks. I would place no bet if the teams were both on streaks (i.e, two teams on winning streaks or two teams on losing streaks) and would place two bets on games in which one team had a winning streak and the other had a losing streak.

For odds, I used the lines printed in the Boston Globe, until I moved in late August, at which time I used the lines printed in the Columbus Dispatch and, on a few occasions, the Globe, USA Today, the Detroit Free Press, and the Chicago Tribune. There were vew variations in lines, but I would have preferred to use a single source. A vacations and the vicissitudes of the Dispatch made that impossible.

I placed a total of 480 hypothetical bets and finished with a winning record of 260-220. However, by my calculations, this would have amounted to a net loss of $47 had I actually placed minimal $5-win bets.

Of course, I had expected that even a winning record could result in losing money, but I was also forced by the "system" into bets on long-odds favorites that I would never have made with real money. In particular, the odds-makers seem to me to be overly enamoured of Roger Clemens and the New York Mets. All in all, my sense is that one could fare pretty well betting on long-odds underdogs and I plan on testing that hypothesis next season.

The results from this study are summarized in the following table.

Record in	
Double-bet games:	55-37
Winning streaks:	122-115
Losing streaks:	138-105
3-game streaks:	111-114
4+-game streaks:	149-106

Only a longer study would tell if betting losing streaks (or betting only streaks longer than three games) is really more profitable. In this sample, for example, Detroit (twice), Cincinnati (twice), and Seattle all had losing streaks of 10 or more games, but no one had a winning streak of that length.

Finally, a thought for extending this analysis: If we really believe that a streak is an indicator of the outcome of a

game, it would really be more accurate to have some index that indicates the relative win/loss streak between the two teams. For example, there is no good reason, given this system, to refrain from betting on a game in which a team with a seven-game winning streak (W-7) is playing a team with a 3-game winning streak (W-3), if the system requires a bet on a W-3 team playing a W-2 team.

Nonetheless, these results also suggest that simply following Sheridan's streak betting system is a good way to lose money.

****

PITCHER'S GAME SCORES: AN UPDATE

By Murray Browne

I have just completed my second year of tracking game scores[2] for starting pitchers, and the first year in which I covered both leagues (in 1988, I tracked the AL only. In 1989, game scores ranged from 94 to 5 in the AL and from 103 to -1 in the NL. The average score for all AL starters was 49.96; in the NL, the average was 52.55. The following table reports all game scores of 90+ in 1989.

Date	Pitcher	GS	Opp
8/30	DeLeon(StL)	103	Cin
6/16	Clemens(Bos)	94	Chi
9/30	Ryan(Tex)	94	Cal
4/12	Ryan(Tex)	93	Mil
4/28	McCaskill(Cal)	92	Tor
6/3	Ryan(Tex)	92	Sea
9/21	Fernandez(NY)	92	StL
4/14	Key(Tor)	91	KC
7/6	Ryan(Tex)	91	Cal
8/16	Tanana(Det)	91	Bal
4/23	Ryan(Tex)	90	Tor
6/7	Swindell(Cle)	90	Cal
8/4	Steib(Tor)	90	NY
9/29	Witt B(Tex)	90	Cal

2. For a review of how game scores are calculated, see either The 1988 Bill James Baseball Abstract or the last issue of By the Numbers.

unlike no-hitters and perfect games, little fanfare accompanied DeLeon's 103 (11 IP, 0 R, 1 H, 0 BB, 8 K),[3] even though it is probably the first game score greater than 100 since Tom Seaver's 106 in 1974. St. Louis lost in extra innings despite DeLeon's masterpiece. The other oddity is that while there were twelve performances in the AL with game scores of 90+--5 by Nolan Ryan alone--, there were only 2 in the NL.

For those who prefer the "men of steel" approach to their hardball, the worst performances of the year were recorded by Bob Knepper of Houston (4.1 IP, 10 H, 10 ER, 5 BB, 3 K, for a -1) and by Bill Wegman of Milwaukee with a 5. Honorable mention goes to Andy Hawkins of the Yankees who scored 6s in a pair of outings.

Following up a suggestion in the previous issue of By the Numbers, I began to break down game scores by team and by home-and-away. This is by no means a complete analysis, but it is a beginning. The table on the following page has these results. (The numbers in bold face are league-leading scores; Totonto had a average home game score of 48.85 in 26 games at Exhibition Stadium and 54.42 in 55 games at the Skydome).

Several things worth mentioning show up in this table. First, team game scores do not match up excatly with team ERAs, since game scores track only the performance of starters. But they sometimes are an indicator of problems in the bullpen. For example, Texas finished fourth in team game score, but only seventh in team ERA. Texas has a top closer (Russell), so that at least we know that Texas seems to have a weakness in middle relief. Montreal has a similar discrepancy (third in game score, seventh in team ERA.

Second, although it is well-known that Los Angeles and Oakland are great pitching parks, LA starters had the highest road game scores in the NL as well as the best at home. Oakland starters, on the other hand, were only fifth in away game scores.

3. Except in the editor's home, where several By the Numbers readers called to tell him about it. Scott Flatow was the first.

Third, some teams performed better at home although the park has a reputation as a hitter's park (the Cubs and the Braves), while Houston, with a home pitcher's park, did better on the road.

I now have compiled the individual pitcher game scores, including offensive support, and pitchers' game score averages. If you would like a list, send me a SASE (236 W. Schilling, West Lafayette, IN, 47906) and I'll drop it in the mail. Also, I'd welcome any comments about the use of game scores.

(Editor's note: The simple correlation between team game scores and team ERA is .904, suggesting that these two numbers are really measuring the same thing. One issue is whether that is true at the individual level, as well as at the team level. Anyone out there want to look at this?)

****

Team	Total	Home	Away	ERA
American League				
Cal	**53.37**	53.80	**52.93**	3.28
Oak	52.99	**55.14**	50.84	3.13
KC	52.21	54.21	50.21	3.55
Tex	52.18	52.14	52.22	3.95
Cle	51.78	52.21	51.36	3.64
Tor	50.90	52.62	49.17	3.59
Bos	49.96	49.37	50.51	4.01
Mil	49.41	51.16	47.67	3.80
Sea	49.35	48.73	49.98	4.01
Min	49.07	47.04	51.10	4.28
Bal	48.15	48.53	47.78	4.01
Det	47.49	49.38	45.59	4.54
Chi	46.81	48.20	45.21	4.26
NY	45.70	45.99	45.41	4.54
National League				
LA	**56.19**	**58.80**	**53.47**	2.94
NY	55.52	58.04	53.00	3.30
Mon	55.35	57.70	53.00	3.48
SD	53.28	54.77	51.79	3.40
StL	53.04	53.57	52.52	3.36
SF	52.79	55.68	49.91	3.30
Chi	51.93	52.15	51.72	3.44
Hou	51.24	50.76	51.91	3.63
Pit	51.27	53.72	48.87	3.64
Atl	51.19	53.43	49.02	3.71
Cin	50.27	48.19	52.35	3.75
Phi	48.58	50.19	47.01	4.08