BASEBALL ANALYST

JOURNAL OF SABERMETRICS

June, 1988 Vol. #36

This is the third month of our catch-up program; the February issue made the mail in June, the April issue in July, and this, the June issue, is out in late August. I'm pleased that the submissions are holding up as well as they are; I think that we have finally cracked a barrier here. For years I more or less begged people to send me there research, and we'd publish anything. I have enough material now to be a little bit selective, and the general quality of what I am receiving is so much improved from two years ago that there is no comparison. Believe it or not, I wound up cutting two of my own articles this time, and didn't really have the space to respond properly to some of the issues raised in this issue. Volume 36 and counting. .

This issue has eight elements:

Bill James, Editor and Publisher Address: Box P, Winchester, KS 66097

BASEBALL'S HALL OF FAME VOTING SYSTEM

Robert O. Wood

Everyone has at one time or another criticized the selections for the Hall of Fame. Criticism is often directed at specific votes, either a favorite player is not elected, or a player deemed unworthy is. Particular vote outcomes should not be confused with the voting system as a whole.

The current system provides two avenues by which a player can reach Cooperstown. He can be elected either by the Baseball Writers Association of America or by the Veterans' Committee. Every player who was active for ten years is placed on the BBWAA ballot. Each year he must be named on five percent of the ballots to merit further consideration. Of course, if he is named on 75 percent he is elected. A player must have used up his 15 years of BBWAA eligibility before he can be considered by the Veterans' Committee.

Of the 149 members in the Hall of Fame elected for their achievements as players in the major leagues (up to and including Willie Stargell's election in 1988, but excluding Negro League selections), 75 were elected by the Veterans' Committee. This 51% ratio does not necessarily imply that the Hall of Fame is overrun by poorly qualified members who snuck in the back door. For example, certain types of players may be notoriously overlooked by the BBWAA.

History of Hall of Fame Voting

The first BBWAA voting took place in 1936. In write-in balloting, the "Immortal Five" were duly elected as their vote exceeded 75%. A five year retirement period was not required, evidenced by the election of Babe Ruth who had retired the year before. The original Veterans' Committee was formed in 1936 to honor 19th century stars such as Cap Anson and Hoss Radbourne. Later incarnations honored early 20th century players such as Tinker to Evers to Chance who played before many members of the BBWAA were born. The 51% ratio seems to be due to the Veterans' Committee being charged with honoring the early stars of the game.

In 1951 Joe DiMaggio announced his retirement with what was considered to be a couple of good seasons left in him. Previously the BBWAA had been hesitant to elect just-retired players. DiMaggio was beloved and could have been immediately elected. A faction within the BBWAA feared a situation in which an active player was a member of the Hall of Fame. There were several owners who would have tried to sign DiMaggio and tout him as a Hall of Famer.

For this reason the BBWAA passed the rule that a player must be retired for five years before he can be voted upon. A second rule required that all eligible players appear on a formal pre-written ballot, and that eligibility would last for 15 years. The Veterans' Committee, in turn, was prohibited from considering anyone not retired for 25 years.

As time passed, players who retired between 1931-1949 fell off the formal ballot. Since the BBWAA was previously under the assumption that their deliberations had no time limit, it was feared these players may not have received full consideration. The Veterans' Committee saw a possible injustice and went overboard in the attempt to correct it.

In an earlier study by myself and Robert K. McCleery ("Ranking the All-Time Greats," Baseball Analyst, volume 28, March 1987), we introduced a methodology to evaluate ball players. In brief, we combined a hitter's offensive and defensive contributions, taking into account the environment and context (era and park effects) in which they were achieved. The result was two numerical formulas (one for hitters and one for pitchers) tailored so that a player with a "value" of 230 meets the minimum standard for Hall of Fame enshrinement, a player with a value exceeding 300 is fully qualified, and a player with a value in the range 230-300 is in the gray area in which he deserves some consideration in making it to Cooperstown.

Of the 14 so-called "fake" Hall of Fame members (players with Wood-McCleery career values below 230), all were elected by the Veterans' Committee, and 9 retired between 1931-1949. Not all from this period were doubtful selections however. The Veterans' Committee deservedly rescued notables such as Goose Goslin (Wood-McCleery value 288), Arky Vaughan (282), and Billy Herman (279).

The list of players who were elected by the Veterans' Committee after being passed over by the BBWAA under the current rules consists of five players: Johnny Mize (327), Enos Slaughter (281), Bobby Doerr (276), Pee Wee Reese (254), and George Kell (242). Three of the five are deserving selections (which we defined as players with values exceeding 265, the midpoint of the Hall of Fame gray area). Indeed we rate Mize as the 24th greatest position player of all-time. Although the other two recent selections are perhaps questionable, in no way does the Veterans' Committee resemble an accident waiting to happen.

Statistical Analysis of Hall of Fame Voting

Bill James has devised a point system to measure a player's chances of making the Hall of Fame. Points are awarded for 300 career home runs, for every .300 season, and so on. The system mirrors the fact that voters do not take into account a player's ballpark nor the era in which he played. There were many .300 batting averages in the 1920-1940 period, so a player of the lively ball era receives more points than a player from the 1960's. Similarly, a hitter who played in a band-box receives points solely due to his favorable home park.

This study will analyze BBWAA voting from 1973-1988. I converted vote totals to the more meaningful percentage of votes received. A player's entire vote history is not needed to predict how well he will do. The key determinant is his vote tally of the previous year. Only a small fraction of the variability remains to be explained by other factors.

Although it seems as if a "directional" variable can explain several individual vote patterns (i.e. a player whose vote totals have increased over time will likely increase in the future), over the entire voting history the directional effect is not statistically significant. The difference in a player's last two vote percentages, which can capture both the direction and magnitude of a trend, accounts for only a little more than 1 percent of a player's vote percentage.

We can divide the model's predictions into those for players on the ballot the first time versus players who have been on the ballot before. With respect to holdovers, the presence of a first-year superstar on the ballot costs a holdover about 3 percent. In the 15th and final year of eligibility, a player receives a bonus of 4 percent. Candidates with large point totals according to James' system have an upward trend of 3 percent per year. Additionally if a player in the previous year received more than 60 percent, he gets a bonus of 5 percent. [I call it the bandwagon effect — if it appears that a player is eventually going to be elected, everyone wants to vote for him.] The quality of the other holdovers on the ballot is irrelevant.

The statistical model is able to predict the vote percentage of a holdover within 5 percent. The voters' evaluation of the first-year crop introduces the most error. The model over-predicted each 1988 holdover by about 2 percent since Willie Stargell was considered to be a better Hall of Fame candidate than his statistics indicated. The difference was enough to make Jim Bunning, whose election was predicted, wait one more year.

With respect to first-years, James' point total is the most important explanatory variable. Voters are in general hesitant to vote for a first-year, reserving the status of first-ballot Hall of Famer for the true superstar. A first-year with a 50/50 chance according to James' point system would be named on 30 percent of the ballots, and a player at the "certain Hall of Fame member" point total would be named on 40 percent. However, the first-year with the most points receives an additional bonus of a whopping 30 percent.

There is a New York bias of 8 percent in voting for first-years (although the bias may be a reward of post-season play, since outstanding World Series performances are not taken into account by James' system, and the New York teams have won numerous pennants). A superstar on the ballot diminishes the vote of other first-years by about 2 percent, above and beyond the 30 percent bonus for the top first-year that the non-superstar will therefore not receive. The strength of the holdovers on the ballot is irrelevant to first-years.

The statistical model is able to predict the vote percentages of first-years within 15 percent. One possible explanation for this low degree of accuracy is that James' point system was designed to measure a player's long range status concerning election into the Hall of Fame, and not specifically his yearly vote percentages. A consensus regarding a player's Hall of Fame chances usually develops as he approaches the end of his playing career. Undecided voters can be swayed by such a consensus. The model, while restricted to numerical data, cannot take advantage of such useful information. For example, it was clear that Willie Stargell would easily be a first-ballot selection, yet the model predicted that he would receive only 65 percent.

Wood-McCleery values were generally worse predictors of a player's Hall of Fame chances than James' point system. Voters have notoriously ignored era and park effects in offensive statistics, and do not seem to be able to accurately combine defense with offense.

Predictions of Future Hall of Fame Voting

It may be of interest to put the model to the ultimate test, that of predicting who will be elected to the Hall of Fame in the next few years. As you may know, there is a surplus of superstars who have retired in the past five years so will be making their first appearance on the BBWAA ballot shortly. The model's yearly predictions follow, where an asterisk denotes a first-year electee. 1989: Yastrzemski*, Bench*, and Bunning; 1990: Palmer*; 1991: Carew* and G. Perry; 1992: Rose*, Seaver*, and Morgan; 1993: R. Jackson*. Notables who are predicted to fall short of the magic 75% figure in 1993, but may eventually get elected, include (with their predicted 1993 vote percentage): Cepeda (58%), Oliva (54%), Santo (33%), Tiant (39%), Jenkins (59%), Kaat (47%), Oliver (31%), Fingers (37%), Perez (22%), P. Niekro (71%), Garvey (35%), Blue (21%). As you can see, the barrage of superstars coming on to the ballot will likely suppress the votes received by current holdovers and non-superstars.

The Veterans' Committee will likely also be electing several members during the next few years. The list of possible inductees includes N. Fox, J. Gordon, T.

Lazzeri, B. Johnson, R. Ashburn, G. Hodges, P. Rizzuto, H. Newhouser, R. Schoendienst, and M. Vernon. It is probable that 5-7 players will be elected by the Veterans' Committee in the 1989-1993 period. In total, therefore, the next five years are likely to witness 15-20 players entering the hallowed walls of Cooperstown. Perhaps the overwhelming qualifications of several of these selections will put to rest the recent debate on the ever-lowering standards for enshrinement into the Hall of Fame.

Concluding Remarks

The BBWAA has generally done a fine job of selecting Hall of Fame members. Its worst picks by far (according to Wood-McCleery value) have been Lou Brock and Ralph Kiner, both of which have outstanding achievements. While I would have voted for neither, these cannot be construed to be idiosynchratic selections.

The Veterans' Committee, on the other hand, has made several questionable selections. The majority came during the transition from the original informal voting system to the current formal system. The Veterans' Committee took the stance that it would rather let in ten undeserving players than keep out one deserving player. We can hope that past egregious errors were due to this isolated situation, and that future selections will maintain the standards of the most recent five selections.

The most glaring weakness of both sets of voters is an inability to understand era and park effects. While this flaw has led to 1930's players being over-represented in the Hall of Fame, it may have the opposite effect for 1960's players. In that event, the Veterans' Committee may serve as a white knight coming to the rescue of Santo, K. Boyer, Cepeda, Allen, Torre, and W. Davis.

Additionally, voters are enamored by specialists at the expense of players who were very good at all phases of the game. Finally, voters have lower standards for Hall of Fame membership than I would have. Whereas I think the gray area should lie between 230-300 on the Wood-McCleery value scale, past voting implies that it actually resides between 220-290. Although the difference may not seem to be great, each player's probability of election is increased by roughly 15%. Standards are a matter of personal taste, however, and one cannot argue about tastes.

While I have been a vociferous critic of many past selections, I would not recommend radically changing the present Hall of Fame voting system. I would, however, recommend fine-tuning by further limiting the number of players a BBWAA member can vote for. The number of votes per ballot has averaged a staggering 7.6, and is largely independent of the strength of the ballot. The number of players that the Veterans' Committee can select should also be further limited.

Although the two-tiered Hall of Fame voting system is considered by many to be inherently flawed, this study suggests that the potential for deserving players to be overlooked by the BBWAA exists, and the potential for undeserving players to be elected by the Veterans' Committee in the future has been exaggerated. What is needed is a healthy dose of education. The importance of era and park effects, as well as the proper way to evaluate a player's statistics, must be better understood. It is this duty that sabermetricians should consider paramount. When accomplished, the quality of Hall of Fame selections will be elevated to a level befitting baseball's highest honor.

A BRIEF LOOK AT LEFT-HANDED PITCHING EFFECTIVENESS

Dick O'Brien

In the Thorn and Holway book, The Pitcher, (Prentice Hall, 1987), an examination (pp 180-182) of why left-handed pitchers have greater success than right-handers comes to no definitive conclusion as to why this is so. They report that a Bill James study showed first-division teams using 44% more lefties than second-division clubs over a recent ten-year period. They also mention a Pete Palmer study showing lefties having a lower ERA 72% of the time since the turn of the century. Palmer's study went on to show that lefties are increasing in numbers through the years and are getting a bigger percentage of staff decisions.

Between the years 1876-1900, only 2 of 27 pitchers winning 150 plus games were southpaws. Although records are incomplete during this period, it's reasonably safe to assert that at no time did the league carry more left-handers than 20% of the total staff. Shown below are the enties by decades of (A) yearly average number of LHP, (B) yearly average LHP percentage of staff, (C) total LHP W/L record, (D) LHP W/L percentage, (E) yearly average LHP percentage of staff decisions.

		(A)	(B)	(C)	(D)	(E)
1901-10	AL	21	.23	1403-1359	.508	.22
1901-10	NL	20	.22	1142-1148	.499	.21
1911-20	AL	29	.24	1542-1510	.505	.22
1911-20	NL	27	.23	1667-1611	.509	.25
1921-30	AL	31	.28	1592-1557	.506	.26
1921-30	NL	29	.28	1570-1568	.501	.25
1931-40	AL	25	.25	1338-1356	.497	.22
1931-40	NL	25	.24	1356-1355	.501	.22
1941-50	AL	29	.27	1438-1509	.488	.24
1941-50	NL	32	.26	1657-1712	.492	.26
1951-60	AL	39	.31	1624-1679	.490	.27
1961-70	AL	49	.33	2516-2443	.507	.29
1961-70	NL	49	.33	2430-2421	.501	.30
1971-80	AL	64	.33	3444-3245	.511	.33
1971-80	NL	56	.30	2932-2856	.507	.30
1981-87	AL	73	.33	2452-2376	.508	.34
1981-87	NL	60	.31	1950-1851	.513	.29
1901-87	Al			17349-17084	.504	
1901-87	NL			16366-16158	.503	

Thorn and Holway suggest that perhaps one of the reasons for lefty success is the infrequency of their appearance. Such an argument would be more persuasive in the early years and sure enough, as their numbers increased their success tapered off (1931-60), but the argument vaporizes from the 1960s onward.

It has also been suggested that the traditional power positions (first base, outfield) have been filled predominantly by left-handed batters, and since they have better marks against right-handed pitching, this accounts for the increased effectiveness of the port-siders. This has been true historically (up to a point) but has become increasingly less so with the rise in number of right-handed power hitters. I find this explanation tenuous, at best.

How to account for the nadir in left-handed pitching success from 1931-1960 is, I believe, simply an overall reflection of the decline of pitching during this time. Pitchers were throwing fastballs and curves 95% of the time and it wasn't until bigger ballparks were built and a larger pitch selection came about that the whole art of pitching improved.

One can't help but wonder about the appropriateness of the truism that right-handed hitters are better high ball hitters and left-handed batters better low ball hitters applying here. Right-handed batters see predominantly low balls from good left-handed pitchers and lefty swingers see mostly high balls. This increasingly evident fact is to me the most persuasive argument thus far as to why lefty pitchers are more effective. They are simply pitching to most batters weaknesses.

Lefties vs Lefties
As Time Goes Bye

Tom Locker

About 9% of the general population is lefthanded. About 30% of Major League pitchers are lefthanded. It has always been an unsubstantiated opinion of mine that it takes a year or two for young lefthanded hitters to adjust to the increased frequency of lefthanded pitchers which must be faced in the Majors, even in the high minors there aren't quite as many lefties as in the Majors. Young righthanded hitters don't face this difficulty, they are actually facing more pitchers against who they have a platoon advantage.

My original idea was that as a young lefthanded hitter became accustomed to facing lefthanded pitchers, his batting average would climb for a year or two, then stabilize at a level representing his actual ability.

So I went through my old Baseball Abstracts and compiled the records of lefthanded batters vs lefties. I wanted to study players who had at least 3 seasons with 100 at bats, with no more than 2 seasons in the Majors before their first 100 at bat season. Of course I fudged on those figures a little if it seemed appropriate. Here's the data.

Base Year at least 100 at bats, years with less than 60 at bats deleted

	Base Yr	+1	+2	+3	+4	+5	+6
Baines	.284	.231	.283	.336	.262		
Boggs	.333	.281	.235	.347	.352		
A. Davis	.286	.239	.246	.240			
M. Davis	.246	.205	.277	.299	.278		
Durham	.294	.322	.252	.282	.227	.257	
Gwynn	.312	.288	.331	.361			
Hayes	.244	.236	.229	.235	.234		
Hrbek	.314	.271	.297	.248	.270	.225	
Mattingly	.331	.288	.356	.302			
Milner	.286	.160	.221	.224	---		
Moseby	.237	.194	.219	.296	.252	.256	.229
O'Brien	.273	.219	.225	.309	.247		
Strawberry	.232	.222	.256	.209	.248		
Average	.282	.243	.264	.284	.263	.246	

I considered a change in batting average of less than 10 points to be no change. The thing that jumps out is that virtually every player's average drops in his second "Fulltime" season (the 2 "Sames" were both drops of less than 10 points). Of the 13 qualifying players in the study, only 1 player (Bull Durham) had a substantial increase, 28 points, and he only batted 87 times. Two of the players did about the same in their 2nd season, the others all declined.

The chart shows the players, as a group, declined against lefthanders from .282 to .243 in their 2nd season. This is by far the strongest and most surprising observation which can be made.

Obviously the pitchers were finding weaknesses; by the 3rd season things begin to stabilize somewhat. But it still takes 3 more seasons to get back to the level of their "rookie" year.

The only conclusion I can offer is that you can bet that any lefthanded hitter who got 100 at bats against lefties for the 1st time in '87, will have a lower average against them in '88.

Seasons with at least 50 at bats

	Base Yr	+1	+2	+3	+4	+5	+6
Baines	86-.151	201-.284	195-.231	198-.283	220-.336	206-.262	
Boggs	105-.333	185-.281	166-.235	213-.347	182-.352		
A. Davis	206-.286	176-.239	122-.246	208-.240			
M. Davis	134-.246	73-.205	155-.277	117-.299	126-.278		
Durham	126-.294	87-.322	107-.252	131-.282	128-.227	74-.257	
Gwynn	65-.231	202-.312	205-.288	236-.331	249-.361		
Hayes	168-.244	46-.304	110-.236	157-.229	196-.235	167-.234	
Hrbek	185-.314	144-.271	185-.297	222-.248	152-.270	138-.225	
Mattingly	181-.331	264-.288	243-.356	199-.302			
Milner	119-.286	94-.160	77-.221	47-.213	56-.224	--	
Moseby	114-.237	144-.194	116-.219	185-.296	210-.252	234-.256	188-.229
O'Brien	143-.273	151-.219	169-.225	165-.309	186-.247		
Strawberry	99-.232	167-.222	156-.256	187-.209	230-.248		

All Seasons

	80	81	82	83	84	85	86
Baines	86-.151	24-.292	201-.284	195-.231	198-.283	220-.336	206-.262
Boggs	--	--	105-.333	185-.281	166-.235	213-.347	182-.352
A. Davis	--	--	--	--	206-.286	176-.239	122-.246
M. Davis	--	--	--	134-.246	73-.205	155-.277	117-.299
Durham	47-.213	--	126-.294	87-.322	107-.252	131-.282	128-.227
Gwynn	--	--	--	65-.231	202-.312	205-.288	236-.331
Hayes	--	--	168-.244	46-.304	110-.236	157-.229	196-.235
Hrbek	--	--	185-.314	144-.271	185-.297	222-.248	152-.270
Mattingly	--	--	--	--	181-.331	264-.288	243-.356
Milner	--	--	119-.286	94-.160	77-.221	47-.213	56-.224
Moseby	114-.237	144-.194	116-.219	185-.296	210-.252	234-.256	188-.229
O'Brien	--	--	--	143-.273	151-.219	169-.225	165-.309
Strawberry	--	--	--	99-.232	167-.222	156-.256	187-.209

Players not used in study

M. Hall	--	--	--	70-.114	26-.192	no data	26-.154
R. Law	185-.238	--	69-.304	91-.297	103-.204	24-.375	--
Scioscia	--	15-.333	25-.080	--	39-.128	87-.253	107-.234
Sheridan	--	--	--	39-.205	62-.194	14-.071	--
Van Slyke	--	--	--	54-.241	35-.257	55-.109	116-.207
Walker	--	--	--	46-.217	84-.179	183-.230	89-.236
Wynne	--	--	--	61-.279	166-.259	111-.243	70-.186

DEBUNKING A PITCHING CLICHE

BY Dallas Adams

One of the common baseball cliches is that good pitching stops good hitting. (The unspoken implication being that this is relative to the normal league rate at which average pitching stops average hitting.) This statement is often heard or read but that's no proof that it is true. Indeed, I have data which demonstrates it to be false.

On a team basis, the job of the offensive side is to score as many runs as possible; which means that the proper measure of team batting ability is the number of runs it scores. Similarly, the intent of the team in the field is to allow as few runs as possible.

If good pitching does stop good hitting, then there will be fewer runs scored in such games, as compared to the runs scored in games matching average pitching against average hitting or poor pitching against poor hitting.

There is a very simple way to sort teams into categories expressing qualitatively the skill levels of both pitching and hitting: Make the assignment on the basis of where the team finishes in its divisional pennant race. A team finishing first in its division, for example, is assumedto have the best hitting and the best pitching in its division; a team finishing second is assumed to have the second best pitching and the second best hitting; and so on.

Sure, for a given team these assumptions can be erroneous. For example, the divisional winner can do it with very strong hitting and only average pitching, or one division can be much stronger than the other. But the following year the reverse can be true and over a multi-year period these variations will cancel out. In the long run, the levels of batting and pitching skill for the average first place team will be greater than those of the average second place club; and the skill levels of the average second place team will exceed those of the average third place squad; etc.

Hence we can examine the number of runs scored in games (over a period of several years) where the pennant winner in one division played the pennant winner in the other division to see how effectively "first place pitching" stopped "first place hitting". We can do the same for games where second place teams played each other in order to find how "second place pitching" fared against "second place hitting". And similarly for third place teams, fourth place teams, fifth place teams and sixth place teams.

As I mentioned above, if the cliche is true and good pitching really does stop good hitting, then there will be fewer runs scored in games between first or second place clubs than are scored in games between lower ranked teams.

The annual Baseball Guides published by The Sporting News contain individual game scores. I tabulated these scores from 1969 (the first year of divisional play) through 1981 for head-to-head play between the teams with equal position of finish in their divisions. The data is shown in the table.

INTER-DIVISOONAL HEAD-TO-HEAD PLAY BETWEEN TEAMS OF EQUAL POSITION OF FINISH, 1969-81

POSITION OF FINISH	TOTAL GAMES	TOTAL RUNS	AVERAGE RUNS/GAME
1	295	2511	8.51
2	294	2451	8.34
3	306	2636	8.61
4	296	2408	8.14
5	295	2463	8.35
6	299	2516	8.42

The small variations in runs/game appear to be random and probably are due to having only about 300 games (12 scheduled games per year in each league) in each category. There very definitely is no indication that good pitching stops good hitting better than other levels of pitching stop the corresponding levels of hitting.

****

NOTE ON LEFTIES VS. LEFTIES

--Bill James

One of the principle theories of the origin of the platoon differential is that it results from the fear of being hit by a pitched ball. Assuming that rookies are battling intensely to prove themselves, that they are more "psyched up" than other players, so to speak, it might make sense that rookies would therefore overcome the self-preservation instinct urging them to bail out on an inside pitch more successfully than they would in later seasons, when there is less emphasis on producing immediate evidence of ability. Just a thought.

Predicting Major League Baseball Team Performance: Beyond the Pythagorean Formula

Robert Cramer, Robin Ellins and David Lutz

Bill James' pythagorean formula succinctly demonstrates that a team's win-loss ratio can be reliably estimated using its total runs scored and total runs allowed. The pythagorean formula's descriptive power serves as a logical point of departure for identifying the specific performance variables that contribute individually, or in combination, to a team's offensive and defensive success.

A. Brian Ault in the 1984 issue of the Baseball Research Journal provided a provocative analysis of the offensive and defensive variables that best predicted the number of major league baseball team victories. Using multiple regression techniques he found that total team wins were best explained by team earned run average (ERA) and the number of runs scored. Ault reported that team ERA and number of runs scored combined with team fielding percentage (pct.). to explain 89% of the variance in American League (AL) team wins and 85% of the variance in National League (NL) team wins. Ault summarized his results by suggesting that in the AL winning baseball was 58% pitching and defense, and 42% offense; in the NL the values were 54% and 46%, respectively. It is important to recognize that the percentages noted above refer to the explainable variance in team performance. Ault also reported that 11% of the variance in AL team wins and 15% of the variance in NL team wins remained unexplained.

Team ERA and the number of runs scored were among the five factors Wiley (1976) found to be correlated with a team's final standing. Wiley's list included fewest runs allowed, slugging pct., and batting average (avg.). Although Wiley reported that batting avg. and slugging pct. were related to a team's final standing, Ault found that team batting avg., slugging pct. and number of homeruns failed to predict team wins.

It was the difference in the research outcomes reported by Ault and Wiley that prompted our work. Specifically, the purpose of our research was to extend Ault and Wiley's work by: (a) determining the relative contribution selected team performance variables make to predicting team wins, (b) reporting the results of a regression analysis employing a more representative sampling strategy than the strategies used by Ault and Wiley, (c) comparing the results derived from the two samples we evaluated with each other and with previous results, (d) shedding some light on the unexplained variance in team victories identified by Ault by investigating the impact of several previously unexamined performance variables, and (e) validating the prediction equation developed from our analysis.

Because several excellent textbooks on regression are available (we can recommend: Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences by J. Cohen and P. Cohen, 1983, Erlbaum; and Using Multivariate Statistics by B. G. Tabachnick and L. S. Fidell, 1983, Harper and Row), we will not attempt an elaboration here. We began our research by collecting offensive and defensive statistics for all major league baseball teams competing during a 12 year period ranging from 1968-1980 using the Sporting News Official Baseball Guide. We have established the five year period from 1982-1987 as a population base for evaluating the predictive validity of our present findings and other ongoing projects; the strike shortened 1981 season was not included in our analysis for obvious reasons. The six offensive variables included in our analysis were team batting avg., slugging pct., number of doubles, number of homeruns, number of stolen bases and a measure of run scoring efficiency, the run-to-hit ratio. This list includes 3 offensive variables not examined by Ault: number of doubles, number of stolen bases, and the run-to-hit ratio. The five defensive variables included in our analysis were team ERA, number of base-on-balls allowed, number of strikeouts, number of double plays, and fielding pct. This list includes 3 defensive variables not examined by Ault: number of base-on-balls, number of strikeouts, and number of double plays.

Rather than use a single sample comprised of all competing teams we developed four samples; the results of two samples will be reported here. Sample 1 included the years 1968, 1972, and 1977 ($N = 70$), and Sample 2 included the years 1972, 1975, and 1980 ($N = 74$). The two samples were developed with the restriction that pairs of years in a sample be uncorrelated with respect to the criterion variable, team winning pct. We wanted to be sure that the inclusion of a case (team) more than once in a sample did not present interpretational problems.

The results of two stepwise regressions showed remarkable consistency (see Table 1). In Sample 1 90% of the variance in winning pct. was explained by our 11 variables; in Sample 2 it was 87%. You can see, however, that only four of the variables accounted for an overwhelming majority of the explained performance variance: in Sample 1, 88.56% and in Sample 2, 85.91%. In samples 1 and 2 team ERA was the most predictive with values equal to 42.73% and 36.72%, respectively. The run-to-hit ratio added another 29.42% in Sample 1 and 35.87% in Sample 2. In Sample 1 teams gave up an average of 3.45 earned runs per game and took 2.17 hits to score one run; in Sample 2 the values were 3.60 and 2.14, respectively. The variables batting avg. and fielding pct. added another 16.41% to explaining team winning pct. in Sample 1 and 13.32% in Sample 2.

The results reported here, together with the work reported by Ault and Wiley, indicate that the most important performance factors are accumulative team statistics - ERA, batting avg., number of runs scored and/or run-to-hit ratio, and fielding pct. - rather than variables representing a smaller group of players. Slugging pct. and number of homeruns were, once again, unimportant offensive factors. Our list of nonsignificant variables, like Wiley's list, included the number of doubles, stolen bases, base-on-balls, and double plays. It can be argued that these factors more than the significant factors represent the performance of a few specialists, and while they may contribute to a team victory on a given day their effect is diluted over the course of a full schedule.

Table 1: Variance Apportionment

	Percent of Total Variance ($R^2 \times 100$)		Percent of Explained Variance	
Variables	Sample 1	Sample 2	Sample 1	Sample 2
ERA	42.73	36.72	47.35	42.36
Run-to-Hit	29.42	35.87	32.59	41.38
Batting Avg.	14.38	12.22	15.94	14.09
Fielding Pct.	2.03	1.10	2.25	1.27
Remaining variables	1.69	.77	1.87	.90
Totals	90.25	86.68	100.00	100.00

One problem with the use of stepwise regression is that the procedure for selecting variables for inclusion in the equation takes advantage of random variations in sampling. Analyzing and comparing multiple samples instead of a single sample, as we did here, is one solution to the interpretational problems inherent in the use of the stepwise technique. Another solution is to empirically establish the validity of a set of variables for explaining team performance by testing their predictive power using a new sample. Therefore, we culled team statistics ($N = 26$) from the 1983 season (chosen at randon from our test population of seasons) and applied the following equation to the prediction of team winning pct.:

Predicted Team Winning Pct. = -3.58 + (-.097)**ERA** + (.76)**Run-to-Hit Ratio** + (2.28)**Batting Avg.** + (3.49)**Fielding Pct.**

We used a standard regression analysis to develop the prediction equation and the weighting coefficients are unstandardized values. The observed 1983 winning pcts. and the predicted values were highly correlated, $\underline{r}(24) = .89$. In general this means the superior teams in 1983 were predicted by the equation to be the top performing teams, and teams predicted to not do so well, did not. To put this finding in perspective we can say that the degree of correspondence found between the observed and predicted winning percentages would be expected to occur by chance only 5 times in 10,000 samples.

In summary, the variables most important to predicting team winning pct. are run scoring efficiency on the offensive side and team ERA on the defensive side. The difference between the two factors in explanatory power is most pronounced in Sample 1, but it is more important to note that in both samples the explanatory variables are similarly rank ordered. Averaging the percentages of explained variance found in our two samples for the predictive offensive and defensive variables we find that the explained variance in winning baseball is approximately 52% offense and 47% defense. The remaining 1% of the explained variance includes the statistically nonsignificant offensive and defensive variables. The relative importance of offense and defense found in our research is the opposite of Ault's results (56% defense and 44% offense when you combine AL and NL results). An explanation lies in our finding batting avg. an important predictor of team winning pct. whereas in Ault's research batting avg. did not predict team wins.

At this time we cannot argue that baseball is 50% offense and 50% defense. The percentages noted above apply only to the variance in winning baseball we can currently explain. Recall, that the unexplained, or yet to be explained, variance in winning is approximately 13% - 16%. Unfortunately, the addition of several theoretically important variables in the present research did not significantly reduce this figure. A comprehensive understanding of the elusive unexplained variation awaits the analysis of additional familiar and/or exotic variables, and the quantification of baseball's current intangibles such as managerial style and skill, the front office, team cohesion, etc. Some people may argue that luck or chance accounts for the unexplained variance in winning baseball and that additional research into reducing it is unwarranted. With all due respect to Albert Einstein, we cannot believe that God would choose to play dice with baseball. We strongly believe that additional research will result in greater predictive precision; the challenge lies in refining our ability to measure performance and then utilizing powerful statistical techniques to test our ideas.

While future work should be focused on understanding the unexplained variance in winning baseball, reserach should also be directed toward a more refined understanding of the currently explainable 84% - 87%. For example, it may be surprising to proponents of Earl Weaver's "three run homer" strategy that the homerun was not a more important contributor to winning. But alas, do not despair. When we used the run-to-hit ratio as the criterion variable partial support for the "power" strategy was indicated when slugging pct. accounted for approximately 49% of the variance in run scoring efficiency. This is not surprising, but once again homeruns contributed very little to prediction. In conclusion, we believe that regression models and other powerful inferential statistical methods can continue to combine with the wealth of available descriptive measures to reveal fascinating insights into the national pasttime.

References

Ault, A. B. (1984). Pitching, defense just slightly more important to team wins than offense. Baseball Research Journal, 53-54.

Wiley, G. T. (1976). Computers in baseball analysis. Baseball Research Journal, 104-112.

RESPONSE TO CRAMER, ET AL

--Bill James

While I appreciate the intelligence and academic maturity of the article, I am frankly puzzled by many aspects of the Cramer/Ellins/Lutz research. The Wiley study referred to throughout the article found that one of the best predictors of team performance was slugging percentage--better than batting average, better really than any other offensive stat except runs scored. My own studies of offensive indicators have also consistently found this to be true. The Wiley study reported that home runs were slightly more important in scoring runs than was batting average, a conclusion which I have also supported.

But Cramer et al toss this aside with a series of puzzling half-explanations, "slugging pct. and number of home runs were, once again, unimportant offensive factors. . .(these) represent the performance of a few specialists, and while they may contribute to a team victory on a given day their effect is diluted over the course of a full schedule. . . once again home runs contributed very little to prediction." Since everybody else that I know of has found that slugging percentage correlates more closely with runs scored than does batting average, I have to marvel a little at the ease with which this conclusion is overthrown. Once again?

Perhaps if I understood the methodology I would see the wisdom of it. . .is what is meant by "stepwise" that each new factor thrown in the hopper is examined for the ADDITIONAL explanatory value? If so, I would really have to question the appropriateness of considering "run-to-hit ratio" on the same basis as things like home runs and batting average. Hitting home runs is a skill, an ability. Run-to-hit ratio is a statistical construct which is influenced by a hundred different things--speed, batting average itself, power. One could make up limitless different ratios of a similar nature--run to at bat ratio, for example, or ratio of baserunners to opposition runs. Any of these, if measured at the right stage of the stepwise process and not forced to compete with more dominant factors, would of course show as tremendously important, but that raises a question: What is it, exactly, that is important? What skill is it that is revealed here as being important?

Well, there isn't one, is there? What it is is an OUTCOME being measured as if it were an ELEMENT. What the study does, I think, is splinter the effect of power into so many fragments that it no longer seems important. They measure two-thirds of the importance of power under a general heading of "run to hit ratio" (hitting home runs, of course, is one of the primary things which causes a run-to-hit ratio to be high), and then measure the other third at a late stage of the winnowing process.

I tried the formula which appears at the bottom of page 13, and came to the conclusion that there is a typo; I think they meant "-3.5" or "-3.508" or something rather than "-3.58". In any case, I applied the formula to the American League in 1987, and was quite surprised by how well it works. The formula seemed to be as accurate in predicting team wins as the Pythagorean method, perhaps more so (again using "-3.5" as the first factor.)

There may not be any way to completely separate the elements of play from the outcomes. Again, let me emphasize that there is a lot here that I don't understand. But I would suggest that unless the distinct elements of play can evaluated on something like (Gary Hart's contribution to the American cliché collection) a level playing field, or unless a tremendous degree of predictive accuracy can be demonstrated for the formulas resulting, then stepwise correlation probably does not have a great potential to help us understand why baseball teams win and lose, or how we can deal with those abilities in statistics.

ADVENTURES IN COMPUTER SIMULATIONS

By Gary Fletcher

In the 1988 Baseball Abstract, Bill James noted that "Several people...have done computer simulations of lineups and have all...reported that it doesn't make any difference, that one lineup is as good as another." He finds this hard to believe, especially since none of them described their methods.

I have written a program that simulates baseball offense. But before I begin to explain it, I think we should ask ourselves, "Just what are we trying to test, exactly?"

DOES THE LINEUP MAKE A DIFFERENCE?

There are three ways, as I see it, that the question can be answered positively:

A. Quality - A lineup of Babe Ruths will certainly outscore a lineup of Mario Mendozas.

B. Loss Of Plate Appearances - Over the course of a full season the leadoff spot will come to the plate roughly 150 times more than the number nine spot. The exact number will vary depending on just how good or bad the teams offense is, but for the sake of the argument, 150 is a good approximation. For each spot a player drops in the batting order he loses about 16 plate appearances. Jack Clark, batting 4th, will then lose about 50 PA on the season. Clark in 1987 created about .2 runs per PA. If the three batters ahead of him create an average of .1 runs per PA then the loss to the team will be about 15 runs.[((50x.2)/2)x3]

This effect could be more extreme in theory - what if the Yankees led off with Wayne Tolleson and batted Rickey Henderson ninth? You'd be looking at a loss of 35-50 runs at least.

Anyway, I think you'd be hard pressed to find a single team that did not have some of its' best hitters(according to RC/PA)batting behind some lesser ones. Even so, I wouldn't lead off Jack Clark either. The reason for that lies in the third possibility.

C. Construction - We all believe your power hitters should bat 3-4-5 and pretty well everyone but Chuck Tanner wants your on-base fellows up 1-2, even if the latter are poorer hitters than the former. Basically, we all figure Jack Clark will drive in Vince Coleman more often than Coleman will drive in Clark.

When some people - Pete Palmer and John Thorn said as much in "The Hidden Game" - say that the batting order doesn't make any difference they are, I assume, saying that this point is overrated, even trivial.

I wonder. It seems to me that the loss of production due to less plate appearances is logically irrefutable. Now, say you simulate a lineup that starts with your best hitter and ends with the worst(the Earnshaw Cook proposal)and the lineup scores about 750 runs. Then you try a more Classic lineup and they score 730 or 765 or something. Do this ten times and maybe you end up with an average difference of just a couple of runs. You might well conclude that the batting order makes no difference.

But it could be that the lost PA production of the classic lineups best hitters might have been 350 runs in ten seasons. If the two lineups still came out even, then it seems to me that the classics were making up those runs a year because of the internal dynamics of their lineup construction.

MY OWN METHODS

As I said earlier, I wrote my own program. I don't relish explaining it step-by-step. If you do any programming, you know how inscrutable even basic code can be, sometimes even to the one who wrote it. So rather than just listing several pages of gibberish, I'll try to explain what I did in a question and answer form.

1. How Are The Results Of Each Plate Appearance Determined?

Suppose a player had 693 plate appearances. If he had 180 singles, the numbers 1 through 180 would represent singles. If he had 14 doubles the numbers 181 through 194 would represent doubles and so on through triples, homers, strikeouts, walks, flyouts, groundouts and GiDPs. The program then generates a random number from 1 to 693 and that determines the offensive result for that particular plate appearance.

2. How Strictly Do You Adjust The Base/Out Situation?

At the start of every inning the bases are empty and there are no outs. I decided that 27 outs ended a game 75% of the time, 24 outs in the other 25%, which would happen in home games the team led in the ninth inning.

3. What About Advances On Base Hits?

The only tricky stuff happens with singles and doubles with men on base. Homers always leave the bases empty and score all baserunners. Triples score all baserunners and always leave the hitter on third.

A single always puts a man on first base. But a runner at first could end up at second, third or home on a single. A runner at second will either stop at third or score.(In reality, he sometimes stays at second. But I decided that that sort of thing is rare. My simulated runners always advance at least one base on a hit.)

A double always leaves a man at second base and scores any runner at second or third. A man on first will either score or stop at third.

I was lucky that the Great American Baseball Stat Book carried base-running tables this year that showed how often, on average, each teams runners went second to home or first to third on singles, first to home or doubles and third to home on sac flies. Thanks to Project Scoresheet.

4. What About Players Who Are Faster Or Slower Than Average?

I gave each player a crude speed score which, at the extremes, either increased of decreased the chances of taking an extra base.

But, to be honest, I didn't let the players speed affect his chance of taking an extra base very much. I don't really know how individual players perform in this area, so I pretty much left it at the league averages.

5. Did You Make Any Other Major Assumptions?

Yeah. Of Vince Colemans' batting outs in 1987 that were not strikeouts, how many were groundouts and how many were flyouts? Beats me.

Working with league fielding data, I figured that about 33% of the batting outs - that were not strikeouts - were flyouts. In next years GABS I'd love to see a breakdown on the batting outs for each player and individual baserunning tables as well.

6. What About Right/Left, Grass/Turf and Day/Night breakdowns?

All these things affect the quality of an offense. If you think your leadoff man should have a high OBP, but against a lefty he doesn't, you won't lead him off. Fine. What I'm trying to test here is the interaction of different types of players in a batting order. If I run a Yankee lineup for a season with Henderson leading off, then batting ninth, it does the experiment no good to have his abilities changing all the time.

7. So You Didn't Adjust For It?

No, but the program will, if you like. If you want to test your lineup against a lefty, you just plug in the stats of the players vs lefthanded pitching. Same for grass/turf and day/night.

8. How Do You Know If The Simulation Is Accurate?

When the program completes a run of 162 games the simulated runs scored match up very well with the runs created estimate derived from the simulated stats. Anybody out there who is using a simulation program should use that test. If the runs created estimate - taken from a seasons' worth of games - doesn't match up well with the actual runs scored, there must be something wrong with the simulation.

A simulation, being an imitation of reality, can never be completely accurate. All I can do is sketch out the broad outlines and larger details.

9. What About Stolen Bases?

Since I don't know how many of a players SB were of third or home, the simulation only considered SB with a man at first. I computed a stolen base attempt percentage and a stolen base success percentage for each player. A player had one opportunity to attempt a steal before each batters PA was resolved.

Granted, the timing of a stolen base was not controlled. My simulated players are just as likely to attempt a steal with George Brett at bat, or a pitcher. No batter was in any way diminished by a stolen base attempt.

I might someday decide to look at the stolen base issue with the simulator. If I do, I'll add these and other modifications.

And no, I didn't forget double plays, either.

WHAT I FOUND OUT

I apologize for leading you on. I don't have anything to say about lineup construction just yet. If you deal with just nine players there are still 3,662,374,391 ways to put them together. To test them all would be an impossible task. (Editor's note: actually there are only 362,880. Nothin' to it.)

So you must have an idea as to what makes a good lineup and why another one is flawed. I've got some ideas, my own and other peoples and I hope to report on them in the future.

But don't go away. I do have two things I'd like to say, things I have learned from this project.

#1. HOW MUCH IS BASERUNNING WORTH?

As I was writing the simulation program, at one point I had something that would simulate a station-to-station offense. In that version, no baserunner ever advanced any more bases than were indicated by the value of a base hit. Stolen bases were still considered.

What would you guess? How many less runs would a team score if no base-runner ever went first to home, first to third or second to home on a single? If no baserunner ever went first to home on a double, or ever tagged up and advanced on a flyball or a groundout?

The answer is 50% I would put a team through a season and if their runs created estimate was around 750, they would actually score around 360-390 runs. After the program was refined to the point where these things were allowed in the proper amounts, the runs created estimate was rarely off by more than 3%

Granted, a lot of second to home on singles and the like are routine. But if we credit an extra base taken in excess of the value of each hit to be due to baserunning, then baserunning is therefore responsible for one run in every two.

Since the runs created formula works so well with team data, what does that mean? It means, first of all, that very good and very bad baserunning is rare and so doesn't upset the runs created formula when dealing with a reasonably large group of players.

It also means that the runs created formula is assuming an average level of baserunning skill. If future reference sources would include complete baserunning tables for individual players, I imagine that the formula could be made extremely accurate for them, and considerably different from what we now have.

#2. JACK CLARK IS EVEN BETTER THAN YOU THINK

A few Abstracts ago, a man named Paul Johnson was given a chance to introduce an adjustment to the runs created formula. The basic thrust of his research was that players who combined high SLG% with high OB% were credited with too many runs created. Bill James agreed with that, and even mentioned the idea again in this years Abstract.

I've run a number of seasons on the simulator using lineups composed of just one player. What I've found is just the opposite.

A lineup of Jack Clarks consistently beat their runs created estimates. A lineup of Cory Snyders consistently fell short. The error was generally in the range of 5-8%

This phenomenon turns out to have predictive value. Players with high SLG% and OB% can be confidently predicted to exceed their runs created estimates, while any player with a poor OB% alone will consistently fall short.

Now, it could be that since Jack Clark is constrained by the batting order to bat just once every nine times that much of his value cannot be utilized by his teams. Perhaps in the real world, such a players offensive contribution, taking place in the context of players much weaker than himself, is diminished.

Maybe. I think it is just as likely that Paul Johnson made a correction that improved the runs created estimates for teams at the expense of the wrong players. I think the runs created formula overrates Cory Snyder and Kirby Puckett and Jim Rice, not Jack Clark and Eddie Murray.

A FINAL WORD

One of the real problems with studying anything with computer simulations is that you are always wondering if you are measuring anything except some misguided programming. That is not a comfortable thought.

But, you know, I can watch a display that shows actual runs and runs created that are current with each at bat, along with all the other stats. Sometimes there will be a rather large difference between the runs and the runs created - but then the gap will close up. Sometimes a big error is just being caught without a chair when the music stops. That's why I have had to run ten seasons worth of games sometimes, just to make sure the error wasn't real, or that the accuracy wasn't just coincidence.

Because of this, it seems almost pointless for me to print up the results of these test runs. It would be impossible for me to prove, long distance, that I didn't just make up all of those numbers.

Well, no. I will run the charts with the next article. I mean, this is the ANALYST.

I sent a copy of my program along with the article, but I doubt that anybody would want to print it here. I wouldn't. It's boring.

But if you want a copy, send an SASE and $2.00 to

Gary Fletcher
#308-2251 Pitt River Rd.,
Port Coquitlam, B.C.,
Canada,
V3C - 1R7

And I might as well warn you that I am interested in programming that works. I'm not implying that this is commercial software by any means.

Also, the program was written in Extended Basic on a TI994A with memory expansion and disk drive(s). Thanks.