By the Numbers

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

Volume 2, Number 2 March, 1990

COMMITTEE NEWS

SABR XX, Cleveland. The Statistical Analysis Committee's meeting at the SABR convention in Cleveland will be from 12:15-1:15 PM, Saturday, July 28. If you are planning to attend the convention, try to make the committee meeting. The complete list of committee meeting time is in the following table.

Committee	Meeting Time
Bibliography	9:00 AM, Friday
Publications	10:00 AM, Friday
Computerization	10:00 AM, Friday
Oral History	12:00 Noon, Friday
Negro Leagues	12:00 Noon, Friday
Collegiate	7:30 AM, Saturday
19th Century	7:30 AM, Saturday
Latin America	8:30 AM, Saturday
Ballparks	8:30 AM, Saturday
Biographical	12:15 PM, Saturday
Stat. Analysis	**12:15 PM, Saturday**
Minor Leagues	1:15 PM, Saturday
Records	1:15 PM, Saturday

The Statistical Analysis Committee is sponsoring a research session at the convention; presentations in the session will be published in the September newsletter. We still have time for a few more presentations, so if you are working on something, if you want some feedback on it, and if you are planning to be in Cleveland, let me know.

Data Bases and Research Resources. I am planning to compile a listing of data bases which are available to researchers. I want to be able to provide people with information on the type and extent of data in the data base, the type of computer programs the data can be used with, the cost of the data, and any comments or feedback from researchers on the ease of use of the data base. I already have some information on the data available from Project Scoresheet. If you have information on other data bases or if you have a data base you would be willing to make available (sell) to other researchers, please let me know (or send me the information). I hope to publish a list of what's available in the June newsletter.

Research Ideas. I have been thinking of things which people might be interested in working on. I wanted to share these with you, in the hope that one of you might take up one of these challenges.

A. Fantasy Leagues. The number and variety of fantasy leagues is growing. For example, Bill james is involved in one which is new this year. I personally have some difficulty with the whole concept, but that's just me. One interesting question is the sorts of data used in these leagues, and the use which is made of this data. Do they use things which are important factors in explaining winning percentage in reality? Which league structures make use of the "best" data in the "best way?

B. Cooperative Projects With Other Committees. SABR now has 14 committees. I think it would be interesting if some cross-committee research projects got started. For example, using the data base which the biographical committee has developed on dates and places of birth and death for virtually all major league players, a researcher could look at any of a number of interesting issues. Among these are the following:

(1) Changing Geographic Distribution of Players by Place of Birth. Over time, how has the geographic distribution of player places of birth changed? In the US, has this simply followed population flows, are there other factors operating as well?

(2) Changing Ethnic/Racial Composition of the Major Leagues. What changes have occurred, during what time periods? How, if at all, are these linked to the socio-economic status of the various demographic groups involved?

(3) Mortality By Place of Birth, Birth Cohort, and Position. Over time, have there been changes in mortality rates and/or life expectancies, by place of birth or by position?

C. Evaluating Measures of Performance. This is a somewhat timely subject, since MLB recently threw out a performance measure (game-winning RBIs). What criteria can be used to evaluate new or proposed performance measures? How would long-standing performance measures stand up to a similar evaluation? How would "esoteric" measures of player performance such as Bill James's Runs Created or Peter Palmer's Linear Weights do in a formal evaluation?

D. Clutch Performance Versus Clutch Ability. This issue goes on forever. We have two more pieces in this issue dealing with it, and I am planning a "theoretical" piece for the June issue. But we could use more. Are there new ways of looking at this issue?

Reviews of Baseball Annuals. I had once hoped we could provide annual reviews of baseball annuals. However, the publication dates of that material (and its timeliness--people buy it when it comes out, not six months later) and the quarterly (more-or-less) schedule of the newsletter amke that difficult. If anyone would like to look retrospectively at a series (e.g., Elias), to see if there are any on-going strengths or weaknesses, I'd be interested. But we won't be doing annual reviews of annual publication.

Support the Newsletter. This is my pitch for material. I'm sort of living issue to issue here, and I could do with a little inventory. So let me see what you're doing.

Donald A. Coffin
Indiana University Northwest
3400 Broadway
Gary, IN 46408
219/980-6646

****

HITTER OR PITCHER?

By Rob Wood

Virtually all breakdowns (home-road; day-night; grass-plastic; "clutch-overall; batter-pitcher) for individual players yield differences which are not statistically significant within single seasons. This does not mean that formal statistical analysis has only negative results to offer. On the contrary, we know that on a higher level several splits <u>are</u> significant. Teams do roughly eight games a year better at home than on the road; every hitter's batting average probably has an inherent left/right platoon differential; a <u>left-handed</u> hitter's batting average is helped most by Fenway Park; etc.

Although we cannot glean much knowledge from the lowest levels of statistics (the sort which Elias has made famous), at one step higher knowledge is there for the taking.

One issue about which we can come to some conclusions, using the sorts of data and methods I used in my pieces in the August and December newsletters, is the importance of the batter and the pitcher in the typical match-up. To make the issue clearer, suppose there were a hypothetical league in which all pitchers were identical. In this league, Wade Boggs would hit .350 against each pitcher, Eddie Murray would hit .300 against each pitcher, Chet Lemon would hit .275 against each pitcher, etc.[1] Looked at from the point of view of the (identical) pitchers, each pitcher would surrender an identical (equal to the league average) opponent's batting average.[2]

Clearly, in this league, the outcome of an at-bat depends only on who the hitter is, and is independent of who the pitcher is. The question is whether this is true of baseball in reality.

---

1. Subject to a random error. More precisely, no hitter's batting average against any individual pitcher would be significantly different from his underlying "true" hitting ability level.
2. Subject, again, to some random error term.

What is the best way to answer this question? Statisticians use the concept of the variance to address such issues, and this concept can be applied here. The (true) variance across all pitchers' opponents batting averages in our hypothetical league is zero (since all have the same "true" opponent BA, equal to the league average), while the variance in BA across all hitters in our hypothetical league is positive. Thus we can see that in this hypothetical league the proportion of the total variance (the sum of the batters' variance and of the pitchers' variance) that is attributable to hitters is 100%, which agrees with the way we set this league up.

This is a use of the statistical tool of analysis of variance, which is used to apportion total variance among the relevant factors. The method used is to identify the total variance and then to determine what proportion of the variance is attributable to each factor.

Three important points must be made here. First, the variances among hitters and among pitchers are to be compared, not the standard deviations.[3] Incorrectly using the standard deviation leads to hitters' and pitchers' proportions of the total variance being too close to 50%. Second, as we are trying to estimate the degree to which the outcome of each at-bat is hitter-dtermined or pitcher-determined, we must weight each hitter's statistics by his at-bats (or plate appearances) and each pitcher's statistics by the number of batters faced. Third, the "weighted variance" must be calculated using the same sample period for hitters and for pitchers.

Table 1 gives the breakdown of total variance for several key offensive statistics for the American League in 1987. Use of the American League raises the importance of pitchers in explaining total variance, because pitchers hit, usually very poorly, in the National League. Including the performance of pitchers as hitters raises the proportion of the variance explained by hitters.

Percent of Variance Attributable to Hitter or Pitcher, 1987 AL Statistics

Category	Hitter	Pitcher
Home Runs*	77%	23%
Slugging Average	66%	34%
On-Base Average	65%	35%
Strikeouts*	64%	36%
Walks**	59%	41%
Batting Average	58%	42%

*Per at-bat.
**Per plate appearance.

The table is self-explanatory. All the offensive categories in the table are more a result of the abilities or actions of the hitter than of the abilities of actions of the pitcher. Home runs are the category which is most strongly affected by hitter performance. Sabrmetricians have long argued that walks are more batter-determined than commonly believes, and this analysis of variance provides additional evidence that this is true. Perhaps even more surprisingly, these results suggest that strikeout rates are more batter determines than is batting average.

Clearly this does not mean that pitching is unimportant. Regardless of who is hitting, you'd rather have Orel Hersheiser pitching than Ted Power (other things equal). What it means is that given who the pitcher is, the outcome of the hitter-pitcher confrontation is more hitter-determined than many people have thought.

(Editor's note. This concludes Rob Wood's three-part piece on statistical significance and the explanatory power of formal statistical analysis. Your comments on any part of it, or on the thrust of the entire piece are welcomed. One question someone might want to address is how this sort of analysis can be used in making decisions which may matter to a team. Any takers?)

3. The standard deviation (usually denoted as σ) is the square root of the variance (σ^2).

****

NOTES AND COMMENT

Pete Palmer had the following comments on the articles in the December, 1989 issue of By the Numbers.

(1) On Pete DeCoursey's piece on pitcher run support: "Since the run support includes all runs and ERA just earned runs, a typical pitcher will have about a 0.50 run margin. It would be better to use runs allowed. Each run difference in run differential per game amounts to an increase in winning percentage of 0.100. This is based on 10 runs per win. A change of 0.89 in run differential per game for Rawley would account for a difference of 0.089 in win percentage, out of a total change in win percentage of 0.274, or about 1/3 of the change. The difference due to chance alone from year to year, assuming all skills both for the pitcher and for his team are unchanged, for a pitcher with 25 decisions each season, is, according to the binomial distribution,

$$[(.5^*.5)/25 + (.5^*.5)/25]^{.5} = 0.141$$

This means Rawley's declining winning percentage is not significantly different from expectations at the 95% level. Still, the overall point that pitchers with a high won-lost percentage are usually on good-hitting teams that give them strong run support as well as having low ERAs is a good one. There is no pitcher in history who would have a projected winning percentage much over 0.600 on an average team. About half the lifetime 0.600 pitchers played for teams that were at least 0.550 without them. It is rare for a pitcher wo be worth more than three extra wins a season to his team over a period of several years."

(2) On my piece on Mitch Williams, he writes: "I hadn't realized how lucky Williams was last year. It turns out that my clutch pitcher index measures this pretty well.

"CPI = (Runs/Predicted Runs), where predicted runs are calculated from the linear weight method with the assumption that doubles and triples allowed per non-homer hit were at the league average. Since it takes into account a few more factors than your simplified method, like homers allowed and hit-by-pitch and also has a negative value for outs and gives a higher value to hits than to walks, it would be expected to get more accurate results. For 102 pitchers in 1989 with 60 to 100 innings pitched, it got a standard deviation in runs allowed of 5.3, compared to 6.9 for you regression result of 0.346*(Baserunners per 9 IP)..." Pete provides a list of all pitchers giving up 1/3 (or more) less runs than expected from 1969 to 198, using linear weights. According to his method, Williams would have been expected to surrender about 40 or 41 runs, about the same as my expectation of 42 runs allowed. These pitchers would be expected to experience a jump in ERA of about 1.4 in the next year.

Murray Browne writes to let us know that he has published a 30-page booklet of pitcher Game Scores for 1989, with additional data and some team-by-team comments. If you'd like a copy, send $2 to Murray L. Browne, 236 Schilling, West Lafayette, IN 47906

Rob Wood sent a copy of an arti... from the New York Times, February 27, 19..., titled "1-in-a Trillion Coincidence, You Say? Not Really, Experts Find," which deals with the likelihood of improbable events occuring when there are a (very) large number of trials. It's a useful piece when you think about how many opportunities there are for something unusual to occur in baseball. If you'd like a copy, send a SASE to Don Coffin, Indiana University Northwest, 3400 Broadway, Gary, IN 46408.

****

ON FORMER PLAYERS AS INFORMATION SOURCES; OR, DOES LARRY BOWA REMEMBER THE REAL PAST?

By Don Coffin

This year's Bill Mazeroski's Baseball annual contained a bound-in sample copy of a new baseball magazine (The Show). This sample had a discussion by three former shortstops (Larry Bowa, Jim Fregosi, and Tony Kubek) about today's shortstops. In the course of that discussion, Larry Bowa commented on what he sees as the rise of offense as a major part of the shortstop's responsibilities, at the expense of defense.

It's not clear to me to whom Bowa refers when he talks about the rise of shortstops who are around primarily for their offense. Ozzie Smith? Kevin Elster? Shawon Dunston? Jose Lind? Spike Owen? Dickie Thon? Alfredo Griffin? Gary Templeton? Jose Uribe? Barry Larkin? Rafael Ramirez? Andres Thomas?

Barry Larkin had the best (part-year) numbers last year, and Ozzie Smith has been the best offensive shortstop in the NL for a number of years. Does Bowa think Ozzie in in the league for his offense? Last year's starting NL shortstops combined for the average offensive production shown in Table 1. (Without Larkin and Smith, it's truly awful.)

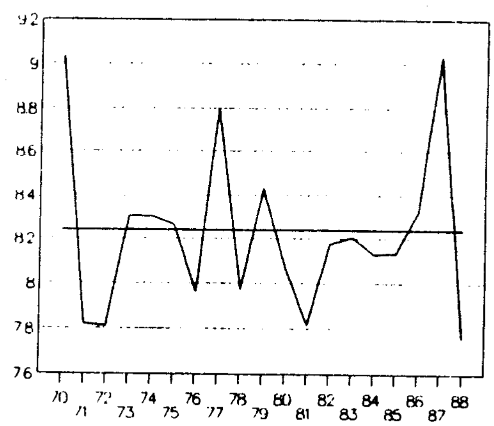
Bowa also commented that in today's game, there are more 6-5 games and fewer 3-2 or 2-1 games than when he played. Now, for this to be true, the average runs scored per game would have to have increased. At least in the NL, at least since 1970 (the year Bowa broke in), average runs scored per game (both teams) has not increased (the average has been 8.24). As the chart in Figure 1 shows, there have been ups and downs, but there is no trend (if anyone wants to see the regression results, I'll be happy to provide them).

Now maybe Larry Bowa doesn't remember what happened when he was playing very well. Or maybe he remembers things so that he looks better in retrospect than he was at the time (he did say that he thought he was the best shortstop he'd ever seen--which means he never watched Ozzie Smith). Or maybe we need to remember to check out everyone's memories when they start telling us how things were in the "good old days," even when the "good old days" were just last week.

Table 1: Average Offense, NL Shortstops, 1989

	Overall Average	Average Minus Smith and Larkin
G	141	144
AB	488	497
R	50	47
H	122	120
2B	22	20
3B	4	3
HR	6	7
RBI	47	47
SB	9	7
BA	.251	.242
SA	.350	.338

Figure 1: National League Runs Scored Per Game (Both Teams), 1970-1988



****

CLUTCH HITTING ONE MORE TIME

By Pete Palmer

(Editor's Note: In the last issue of the newsletter, Rob Wood addressed the single-season batting average differences required for a "clutch" BA to be significantly higher or lower than the overall BA for a hitter. Here, Pete Palmer uses the career data from Elias to examine a very similar question.)

For the first time, the Elias Baseball Analyst has published at-bat figures for their leaders in various game situations. This allows for an analysis of the variance between normal and special situations. There is a handy formula which allows you to find what kind of difference would be expected by chance and whether the actual differences found are greater than what would be anticipated. The formula assumes that batting is a binomial distribution where each event has a constant probability of success or failure. The probability of success (p) is simply the batting average, while the probability of failure (q) is simply (1-BA). The only other number needed is the number of at-bats (n). The formula is

$$\sigma = [(p*q)/(n)]^{.5}$$

For example, for a BA = .250, with 600 at-bats,

$$\sigma = [(.25*.75)/(600)]^{.5} = 0.018.$$

The binomial theorem tells us that, with a sufficiently large sample, about 2/3 of the values in the distribution will fall within one standard deviation (σ) of the mean, and that about 95% of the values will vall within 2σ of the mean. In the example, 2/3 of the values in the distribution would be BAs between .232 and .268, while 95% of the BAs would fall between .214 and .286.

When comparing samples (x and y) of two different sizes, the formula used is the formula for a pooled standard deviation, or

$$\sigma_{xy} = [(p_x q_x/n_x) + (p_y q_y/n_y)]^{.5}$$

For example, if a player has 236 hits in 671 at-bats in late-inning pressure situations (a .352 BA) while going 1083 for 3660 (.296) otherwise, we can calculate a pooled standard deviation as

$$\sigma_{xy} = [(.352*.648/671)+(.296*.704/3660)]^{.5}$$

$$\sigma_{xy} = [.000340 + .000057]^{.5}$$

$$\sigma_{xy} = [.000397]^{.5} = 0.020$$

The next step is to figure out the expected difference ($BA_x - BA_y$) between the two situations. This is something which is often left out by analysts, who assume that the expected difference is zero. Because pitchers bear down more in clutch situations and because ace relief pitchers are more apt to be used in late inning pressure situations, the overall batting average is about 10 points lower in these situations. League figures as shown in the Elias Analysts indicate only a six point drop in BA, but the error in this number is about two points. Also we do not have league data for the full 10-year period covered by the data presented. Using a difference of 10 points balances the number of extremes between the high and the low figures.[4] If we used a drop-off in BA of six points, the number of extreme cases would be the same, but more of them would be negative, rather than positive differences. If you assumed that there is no drop-off in clutch situations, then you would conclude that the number of hitters with higher BAs in clutch situations is smaller than that expected by chance.

The data I used in the example of calculating a pooled standard deviation is for Tim Raines, the leading late-inning pressure hitter. His difference is (.352-.296), or +56 points in "clutch" situa-

4. Also, using a larger drop will tend to make positive differences more significant, which biases Palmer's conclusions in favor of the hypothesis that there are more "clutch" hitters than would be expected by chance (Ed.).

tions. The expected difference is -10 points, so his overall difference is +66. If you divide this by the 20-point σ (calculated on the basis of Raines' performance data), you get a Z-score of 3.3. Raines BA in late-inning pressure situations is 3.3 standard deviations higher than his BA in other situations.

Now a Z-score of 2.0 should occur, BY CHANCE, about 5% of the time, half on the positive side of the distribution and half on the negative side of the distribution. We can re-phrase this by saying that about 2.5% of all hitters should have clutch BAs which are 2 standard deviations (or more) above their BAs in other situations, while about 2.5% of all hitters should have clutch BAs which are 2 standard deviations (or more) below their BAs in other situations. A Z-score of ±3.0 or more should occur only about once in 800 observations.

The lifetime data in the Analyst covers all players in the past 10 years with at least 250 at-bats in late-inning pressure situations (or at least about 1500 at-bats overall for the average player in this group). There have been about 330 such players in the past 10 years, so one would expect about one player to have a Z-score of 3 or more (in either the positive or negative direction). There was, in fact, one such player--Tim Raines.

We would expect about 16 players to have Z-scores in excess of 2.0 (about 8 positives and 8 negatives); there were 14. Looking at the 1988 leaders, there were 10 players (out of 210 listed; 50 or more "clutch" at-bats or about 300 total at-bats) with Z-scores of +2.0 or greater (5% of the total), just as expected. (The highs and lows for the 10-year data run are listed in Table 1.) In short, over the past 10 years, and in 1988, the distribution of performances in late-inning pressure situations appears to conform to a random binomial distribution. It does not provide evidence for the presence of clutch ability.

****

Table 1: 10-Year Clutch Batting Averagesand Deviations from Expected Batting Averages

Player	Clutch BA	Other BA	Z
Raines	.352	.296	3.30
Sax	.318	.277	2.54
G.Iorg	.306	.252	2.42
R.Henderson	.319	.288	2.17
Newman	.269	.214	2.15
Fernandez	.336	.293	2.07
Manning	.277	.244	2.01
Hoffman	.287	.237	2.01
Oester	.290	.261	1.94
B.Diaz	.285	.253	1.90
Milbourne	.304	.259	1.84
Bosley	.310	.261	1.83
C.Brown	.309	.264	1.77
Coleman	.291	.258	1.66
R.Roenicke	.269	.228	1.65
L.Salazar	.286	.260	1.61
C.Moore	.289	.263	1.55
Tolleson	.272	.245	1.43
Wiggins	.284	.255	1.42
Foley	.283	.255	1.30
Staub	.295	.267	1.29
Yeager	.238	.212	1.25
B.Wills	.289	.262	1.24
G.Davis	.228	.269	-1.24
Scioscia	.228	.269	-1.36
Webster	.237	.283	-1.40
Bonnell	.235	.279	-1.46
Grubb	.225	.276	-1.47
Doran	.236	.278	-1.56
G.Brock	.204	.251	-1.57
Foli	.220	.269	-1.62
Benedict	.205	.252	-1.66
L.Smith	.248	.295	-1.68
Butler	.242	.287	-1.68
Morrison	.227	.270	-1.69
Bittner	.227	.293	-1.76
J.Davis	.213	.259	-1.83
Randolph	.235	.281	-1.95
Kittle	.186	.245	-1.95
Heath	.211	.259	-1.97
Lynn	.237	.285	-2.06
Gladden	.220	.281	-2.12
Burleson	.215	.282	-2.28
S.Owen	.184	.250	-2.41
Heep	.194	.265	-2.43
Rice	.245	.305	-2.88

OR DOES CLUTCH ABILITY EXIST?

By Tom Conlon

I have become increasingly alarmed by the number of statisticians commenting on "clutch ability" in baseball and concluding that "clutch ability does not exist." I will not, however, present data supporting tha actual existence of clutch ability. I wish to make the point that none of the studies (which I have seen), which purport to prove that clutch ability does not exist, have sufficient statistical power to be considered conclusive. Hence, I believe that the search for a bettwe statistical methodology must go further before we can arrive at a definitive answer to this question.

Generally, the reasoning used to demonstrate the "non-existence" of clutch ability is one or another version of using the odds of observed data occurring purely by random chance, and concluding that the odds against observing the actual data are not sufficiently high to conclusively conclude that the observed data are not due to random chance. In an article in Chance Magazine, Spring, 1989, Stephen J. Gould quotes Nobel (physics) laureate Ed Purcell as stating that "Nothing ever happened in baseball above and beyond the frequency predicted by coin-tossing models." In the December, 1989, issue of By the Numbers, Rob Wood presents a table of "how large the splits would have to be between clutch and non-clutch performance for these to be statistically significant at the 5% level."

To both authors, I would pose the following question: "If the model 'explains' the data, then the model is correct and unique. True of false?" (Clearly, the answer is "False.") Both authors jump to the conclusion that just because what was observed has a certain probability of being observed under a model of pure random fluctuation, and because the probability of observing the data under the "random model" is not sufficiently small (5% is often used in scientific studies as the threshold for significance, although it is not clear that this standard should apply here), that therefore this PROVES that the null hypothesis[5] is true and that the phenomenon does not exist. But any beginning student of statistics should understand that you cannot conclude that the null hypothesis should be accepted, unless you also can show the power of your data to reject the null hypothesis against various specific alternatives, if, indeed, the null hypothesis is false.

Mr. Wood tells us that the split between clutch and non-clutch performance needs to be bwtween 120 and 137 points of batting average for a 500 AB season in order to be statistically significant. Clearly, players do not obtain such a large difference, by any of the various methods of defining clutch situations. Is this observation, in and of itself, sufficient to conclude that no smaller (but real) differential could be the sign of a "clutch" hitter (as opposed to a pure random fluctuation)? In baseball, even a 50 point batting average differential[6] is very large--it alone carries you from mediocrity to being a star (.250 to .300, say). If a difference in a player's clutch/non-clutch average were as large 50 points, this would indeed be "very significant" to that player, that player's manager, and the fans following the player. I can only conclude by asking Mr. Wood the following question: What is the power of whatever test he used to compute the numbers presented to detect a 50 point difference over a 500 AB season? (I am sure it would be very small.)

I am still on the fence as to whether I believe in the existence of "clutch ability," either for hitters or for pitchers. I cannot prove its existence, but I cannot accept its non-existence either. I believe that current statistical methodology does not have sufficient statistical power to detect the amount of difference that we could reasonably expect to see,

5. (Editor's note.) In statistical tests, the "null hypothesis" would be the hypothesis that some observed difference--e.g., between clutch and non-clutch BAs is zero; the "alternative hypothesis" is that this difference is non-zero.
6. (Editor's note.) Over a full season.

and hence the continued claims that "what was observed was consistent with the null hypothesis," and that we must therefore accept the null hypothesis, are not acceptable. No statistical argument which concludes with the acceptance of the null hypothesis is complete without presentation of the power of the study, and I have not seen one article on this topic that did present a power computation. Since baseball statistical literature is read by statisticians and non-statisticians alike, I believe that the professional statisticians among us must endeavor not to present misleading statistical arguments which present only half the picture, and then to draw sweeping conclusions from them.

I look forward to more lively discussions of this topic, and to the development of new and powerful methods to answer this difficult, but interesting, question of the existence of clutch ability in baseball.

****

Editor's Note

The following article is excerpted from 1990 Baseball Woodview: A Sabrmetric Look at Baseball, a collection of eight essays by Rob Wood, whose work has been on display in By the Numbers. In addition to the essays, the book contains player rankings similar in content and tone to those in last year's Baseball Abstract. The essays are generally interesting and occasionally quite stimulating. If you want to order a copy of the book, send $7.95 (which includes postage and handling) to Robert O. Wood, 2101 California St., Suite 224, Mountain View, CA 94040.

****

SHOULD OZZIE SMITH BAT CLEANUP?

By Robert O. Wood

Bill James' Runs Created and Pete Palmer and John Thorn's linear weights formulas explicitly assert that, to a close approximation, a player's offensive contributions can be calculated with no knowledge of his runs scored and runs driven in totals. These pioneers have developed methodologies which rely solely upon the player's offensive components (singles, doubles, triples, home runs, walks, etc.) that do not depend upon the base-out situation or upon his teammates' prior or subsequent production.

This "situational independence" is one of the leading virtues of their formulas. However the idea that runs created should be the sole criterion of offensive production is controversial. Surely there is some information we can derive from runs or RBI totals. In this essay I hope to shed some light on a few issues related to the Runs Created methodology.

The following is a quote from a prestigious economics research journal that caught my eye, as it is just as applicable to much research in the sabermetrics field:

> The narrative approach allows a vast body of information that cannot be employed by conventional statistical methods to be brought to bear on the question. But the use of the narrative approach is fraught with dangers. It is subject to bias in the selection of evidence to present and in the interpretation of the historical record. ["Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz," Christina Romer and David Romer, NBER Working Paper No. 2966, May 1989]

The authors were beseeching their fellow economists to question the research of the renowned Milton Friedman, who is often able to win debates with the forcefulness and cleverness of his arguments, rather than with convincing statistical analyses of large samples of data. The research styles of Friedman and Bill James are similar in that they often employ the "narrative" approach: relating anecdotal evidence of a few selected instances to help them draw sweeping conclusions.

It is my view, as well as the above-referenced authors, that we must put these giants' research under even more stringent scrutiny than the research of others since we have naturally become accustomed to believing everything they say. I think sabermetrics has reached adolescence and is ready to question a few of the tenets of the field, even if they may have been introduced by Bill James. In this essay I hope to look closely at one aspect of James' runs created methodology.

It could well be that a player with a lower runs created mark is more valuable than another with higher runs created. To see that this is possible, consider the following. Suppose there are two outs in the bottom of the ninth inning with the bases empty and the home team trailing by a run. It is quite understandable that a manager would pinch-hit a home run hitter, say Dave Kingman, for the scheduled batter, say Bill Buckner, even if Kong has a much lower batting average, on-base average, slugging percentage, and runs created.

The reasoning is sound: even if Buckner gets a hit, it would likely be only a single, and would require at least one more hit to score him, whereas Kingman can tie the game "with one swing of his bat." In this case, pinch-hitting with Kong is essentially a one run strategy in that it increases the team's chance to score exactly one run in the inning, but decreases the chance to score two or more runs. Conversely, behind by two runs, a walk or a single is virtually identical to a home run, as both merely bring the tying run to the plate.

Once one admits the possibility that there are situations in which a low-average slugger is more valuable than a high-average banjo hitter, even though the slugger's runs created are less than that of the banjo picker, then the validity of runs created is called into question. For James used overall team totals in determining the relative values of singles, doubles, triples, home runs, etc., which by their construction are necessarily situation independent.

The value of a single is assumed not to depend upon how many runners are on base, and the same for every other type of outcome. There are several reasons why James (and Palmer and Thorn) devised their formulas this way. First, more detailed play by play data have only recently become compiled in a systematic fashion. Previously they only had team totals with which to work.

Second, the situational dependent aspects of baseball are assumed to "even out" over the course of the season. But to conclude this after assuming it at the outset is hardly valid reasoning. While these pioneering analysts trumpet the virtues of their "situational independent" statistics, the costs involved have never been studied.

While it is my belief that the situational-independent statistics of James and Palmer and Thorn have made significant contributions to sabermetrics, there is an air of finality to them that I feel may be unwarranted. Nowhere do they prove that their situational independent statistics are valid measures of individual player's contributions. They simply point out that their statistics can predict team runs scored totals fairly well. To my way of thinking it requires a leap of faith to go from teams to players.

What I propose in this study is to simulate thousands of games in order to get a fresh look a the situational dependence of runs scored totals. We heard how valuable Jack Clark was to the 1985 Cardinals (and over the years of several other such examples), since he was the one and only power hitter in a lineup otherwise made up of banjo hitters. "Take away Jack Clark and the Cardinal offense folds like a house of cards" was the refrain.

To get a credible contrast in hitting styles, I have constructed a slugger and a singles hitter to have identical runs created per 27 outs. The slugger will be called "Jack Clark" and the banjo hitter will be called "Ozzie Smith". Ozzie has a .300 batting average, but all of his hits are singles. Jack Clark has a .250 batting average, but with his doubles and home run power, has a .385 slugging percentage. Walks are not considered in my study (or, rather, they are treated as a special type of single). Each player creates exactly 3.5 runs per 27 outs.

To investigate issues regarding batting orders, I must have both a top and a bottom of the order. The simplest way to do this is to have a pitcher hit in the ninth spot. The pitcher is assumed to have a

.150 batting average, all of the hits being singles which advance base runners exactly one base. The pitcher is also assumed to be a fairly decent bunter.

As my control case, a lineup consisting of Ozzie Smith in spots 1 through 8, followed by the pitcher, averages 2.98 runs per game (in 10,000 games). We see that the weak-hitting pitcher costs the team roughly half a run per game.

This basically homogeneous lineup is contrasted to the lineup with Jack Clark in the cleanup spot. The preliminary belief is that this lineup will score more runs than the Ozzie lineup, since Clark's power will be of more value. This is borne out by the simulation, but to only a minuscule degree. Indeed the Clark lineup averages 3.02 runs per game (in 10,000 games).

The difference is small, only 4 runs over the course of the 162-game season, but is significant in a statistical sense. There is less than a five percent likelihood that the difference in runs per game in such large samples is due to pure chance. The Clark lineup is actually better than the Ozzie lineup, but the difference is minimal.

In a sense this differential of four runs gives a measure of the variability of the runs created formula. Given the player's statistics (singles, doubles, etc.), the range of his true offensive contribution is seen to be small.

James has previously demonstrated that the variability of the runs created formula in team runs is small (on the order of 25 runs per team per season). My result is more persuasive in that it demonstrates that each player's offensive contribution is suitably captured by the runs created formula within 4 "team" runs.

Using James' Pythagorean projection, Clark's team winning percentage versus Ozzie's team would roughly be .507, or an 82-80 record for the season. Thus the impact of Jack Clark is roughly 1 game in the standings. This finding is consistent with previous researchers results, including some of my own work, on the (lack of) import of the batting order.

As a test of the robustness of my result, I ran the simulation once again, but improved the quality of the representative player. When Smith and Clark each create roughly 4.25 runs per game, adding Jack Clark as the cleanup hitter in a lineup of Ozzie Smiths improves team run totals by 0.05 runs per game (over 10,000 games).

By symmetry we would expect a parallel simulation to find that hitting Ozzie Smith leadoff in a batting order of Jack Clarks would improve the team's runs scored by a small but significant amount. Gary Fletcher has previously conducted such a simulation ["Simulator-II: The Leadoff Man and his Effect on the Lineup," Baseball Analyst, August 1988]. Gary finds that a lineup headed by a high on-base average type (Ozzie) scores more runs relative to its runs created than a lineup headed by a player who creates the same number of runs as Ozzie but with a lower on-base average. As in my simulation, the differential in runs is small.

The original idea for this study was to test the "validity" of James' runs created methodology. Runs created, like Palmer and Thorn's linear weights and several other of the New Statistics, are trumpeted as "situational independent". Thus a player having a good season on a poor offensive team is not "penalized" for his lack of runs and RBI (the two leading "situational dependent" statistics).

Situational independence is truly a virtue of runs created. However, to my knowledge, no one had previously attempted to assess the cost of this virtue. "Everyone knows" that the role of a leadoff hitter is to get on base and the role of the middle of the lineup is to hit for power. Why? Because, simply, the team will then score more runs.

These two considerations are at loggerheads. If lineup construction is important, then James' runs created formulation is invalidated (or at least the "standard errors" of his estimates would be large). If, on the other hand, runs created is valid, then lineup construction must be relatively unimportant.

In the present study, I explicitly tested the impact of putting a "slugger" in the middle of the lineup. This study is an important confirmation of James'

runs created methodology. I have shown, at least to my satisfaction, that runs created and linear weights are indeed valid measures of offensive production, and that the cost of their situational independence is typically minuscule.
****

THE STATISTICAL ANALYSIS COMMITTEE BIBLIOGRAPHY PROJECT

One major difficulty for people trying to find out what research has already been done on baseball is the lack of any index or bibliography of works. We are going to try to correct this through the **Statistical Analysis Committee's Bibliography Project**. Here's how the project will work.

(1) We have developed a list of "key words" to identify the subject(s) of a piece of research. That list is published at the end of this outline of the project.

(2) We are asking members of the Statistical Analysis Committee, and of SABR in general, to identify periodical or newspaper articles, book chapters, books, etc., which apply statistical analysis to baseball. Statistical analysis, for out purposes, involves more than simply creating and reporting on a new statistic or method of measuring performance. We need articles which have applied some formal method to see how well some event of set of events in baseball can be explained using some measure of some event or set of events in baseball. Examples of formal analysis include, but need not be limited to, the following:

a. Analysis of the distribution of an event (is it a normal distribution, positively skewed, etc?).
b. Analysis of the difference between an observed and an expected value.
c. Correlation analysis.
d. Analysis of variance.
e. Factor analysis.
f. Regression analysis.
g. Simulation studies.

(3) Once the research has been identified, provide complete bibliographic information about it. For articles this will include the following:

a. Author.
b. Article title.
c. Place and date of publication.
e. Pages.

For a book or book chapter, this includes the following:

a. Author.
b. Book title and chapter title (if relevant).
c. Publisher and date of publication.
d. Page numbers (if relevant.

Once the bibliographic information has been listed, you should then examine the list of key words and attach one primary key word code number and up to three secondary key work code numbers. The primary key word is the one which best describes the major emphasis of the piece in question. Use secondary key words to describe subordinate points in the analysis.

We need a lot of help on this. Several committee members have suggested that for some sources we ask for volunteers to do a series (e.g., the Elias Analysts, the Bill James Baseball Abstracts, the SABRmetric Review, etc.). Among the publications that we know we need to index are the following:

The Elias Baseball Analysts.
The Bill James Baseball Abstracts.
The Baseball Abstract Newsletter.
The Baseball Analyst.
The SABRmetric Review.
The Baseball Research Journal.

We also would like to be able to index team-specific newsletters, including those which have gone out of business. If you have any of these, please help us out.

Additional sources of articles include academic journals (such as the American Economic Review, the Southern Economic Journal, the Journal of Accounting, etc., in which I know of articles), newspapers, self-published pamphlets, etc. Use what you have to help us out.

Finally, we would like to be able to maintain a file of out-of-print or hard-to-find articles, which could be made available to researchers for the cost of duplicating and postage. I have the space and the support to do this, so if you have something that needs to be preserved, you might send me a copy.

The list of key words, which follows, is based on a preliminary list developed by Clem Comly. This list can be modified at any time, and if you have suggestions for improving it, let me know. I have modified it somewhat and provided a grouping. This grouping involves nesting key word categories, with associated key words and code numbers. With each key word I have provided a code number. In indexing articles, please use the code numbers. In each category, the highest level is a one-digit code number (1 for Team Winning Percentage). Under that are related key words and code numbers (2, 3, or 4-digits long). Please classify things at the lowest level you can.

Code Number	Key Word(s)
1	Team Winning Percentage
11	Offense and Winning Pct.
111	Runs Scored
112	Batting Average
113	Slugging Average
114	On-Base Average
115	Baserunning and
12	Defense and Winning Pct.
121	Runs Allowed
122	Pitcher Performance
123	Fielding
2	Team Offense/Runs Scored
21	Batting Average
211	Situation-Specific BA
22	Slugging Average
221	Situation-Specific SA
222	Isolated Power
23	On-Base Average
231	Walks
232	Hit-by-Pitch
24	Baserunning
241	Stolen Bases
242	Hit-and-Run
243	Outs on the Bases
25	Advancing Baserunners
251	Sacrifice Bunts
252	Sacrifice Flies
26	Misc. Team Offense

Code Number	Key Word(s)
3	Team Defense/Runs Allowed
31	Pitching
311	ERA
312	Opposition BA
313	Opposition SA
3131	Home Runs Allowed
314	Opposition OBA
3141	Walks Allowed
315	Shutouts
316	Complete Games
317	Relief Pitching
3171	Relief Pitchers/Game
3172	Relief IPs
3173	Saves
32	Fielding
321	Fielding Average
322	Assists
323	Range
324	Def. Efficiency Ratio
325	Defensive Average
33	Misc. Team Defense
4	Individual Offense
41	Batting Average
411	Situation Specific BA
42	Slugging Average
421	Situation Specific SA
422	Isolated Power
43	On-Base Average
431	Walks
432	Hit-by-Pitch
44	Baserunning
441	Stolen Bases
442	Hit-and-Run
443	Outs on the Bases
45	Advancing Baserunners
451	Sacrifice Bunts
452	Sacrifice Flies
46	Special Measures of
461	Runs Created
462	Linear Weights
463	Other
47	Misc. Indi. Offense

Code Number	Key Word(s)
5	Individual Defense
51	Pitching
511	ERA
512	Opposition BA
513	Opposition SA
5131	Home Runs Allowed
514	Opposition OBA
5141	Walks Allowed
515	Shutouts
516	Complete Games
517	Pitcher Game Scores
518	Relief Pitching
5181	Relief Appearances
5182	Relief IPs
5183	Saves
52	Fielding
521	Fielding Average
522	Assists
523	Range
524	Def. Efficiency Ratio
525	Defensive Average
53	Special Measures of
54	Misc. Indiv. Defense
6	Pay and Performance
61	Player Salaries
611	Performance
6111	Performance Measures
612	Salary Arbitration
613	Free Agency
614	Discrimination
62	Managerial Salaries
621	Performance
6211	Performance Measures

Code Number	Key Word(s)
7	Team Profitability
71	Revenue
711	Attendance
7111	Ticket Prices
7112	Incomes
7113	Population Characteristics
7114	Player Characteristics
7115	Team Winning Pct.
7116	Televised Games
712	Broadcast Revenue
7121	Local Broadcast Rev.
7122	National Broadcast Rev.
713	Concessions/Parking
714	Licensing Fees
72	Costs
721	Player Salaries
7211	Salary Arbitration
7212	Free Agency
7213	Deferred Salaries
722	Player Development Costs
723	General/Admini. Costs
724	Depreciation
73	Team Profitability
731	Winning Percentage
732	Market Size
733	Franchise Values
734	Trends in