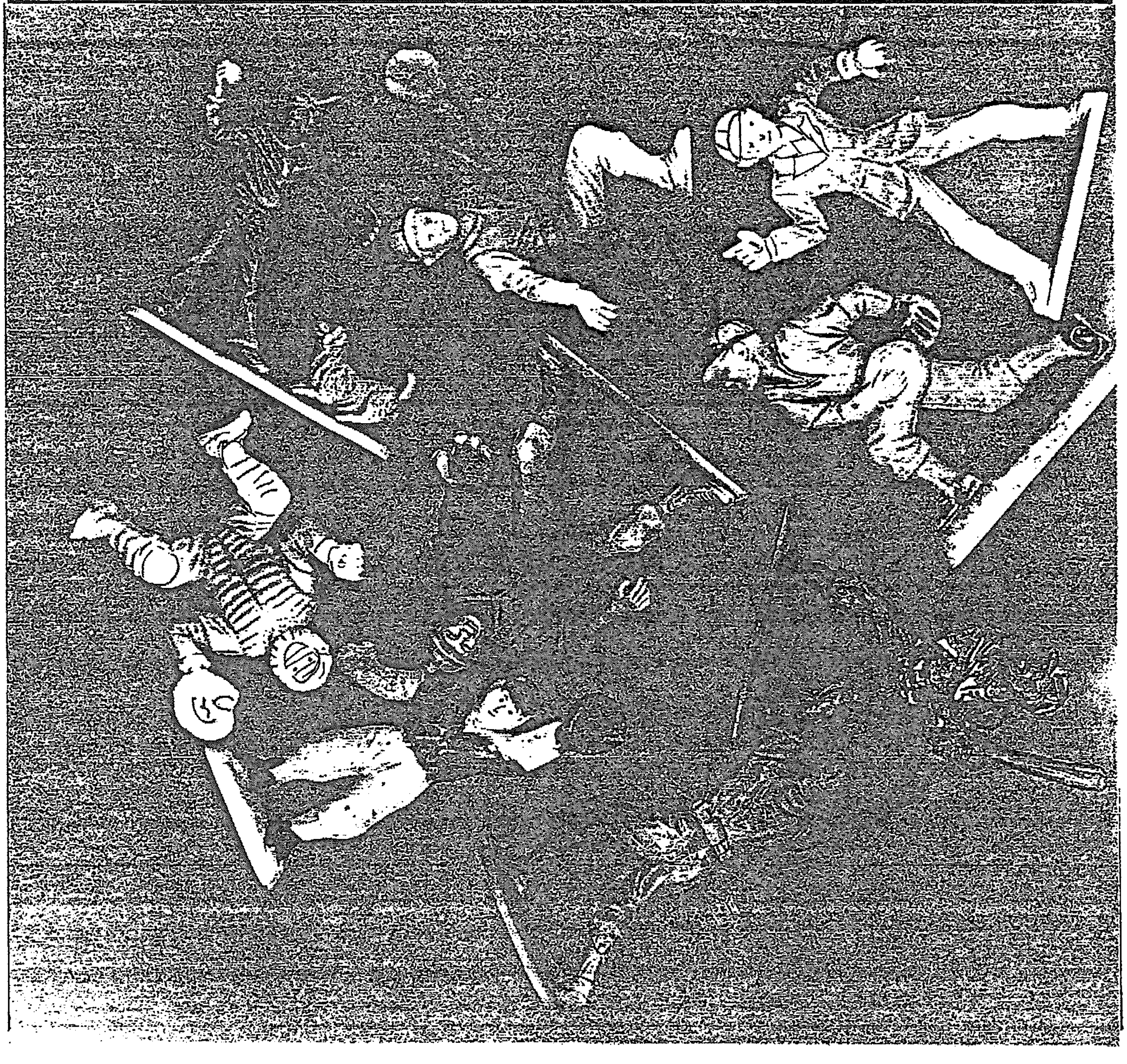
baseball ANALYST

Volume 18 "The Journal of Sabermetrics" June 1985



SAY WHAT, JOE?

Joe Filler

(Detroit) There are things that I like. What are they? They are things that I think up when I put my mind to it. And when I put my mind to it and make my brain work really hard I can think of all kinds of things that I enjoy better than other things that I don't enjoy so much. The things I don't enjoy I don't like to think about because that wouldn't be nice and I certainly wouldn't want to torment myself or my readers with lists of things that I myself can't even bear to think about, let alone write down on a napkin at my favorite bar and then later, if the swelling isn't bothering me too much, pound out onto the old keyboard for publication. There are any number of things that I love and you know, I can't begin to tell you how much I like to list them for all my reader-friends to see and talk about later and maybe get inspired to make lists of their own of things that they like thinking about and doing and seeing and hearing about and tasting--we can't forget tasting.

As Bobby Doerr once said to this reporter, "Joe, I know what I like and I know what I don't like." And anything that a genuine, good old-style ballplayer like Mr. Doerr has to say can't be all wrong in my eyes. The trick is sitting down and actually doing it. According to an editor I once had, thinking about things just isn't enough, you've actually got to write 'em down. I think one after another on a simple piece of white paper is a good way to start. And for you kids out there, the little shavers still trying to figure out what a box score is, don't be afraid to just up and do it. The first one is always the hardest. Yes sir. Knowing that directly underneath you is going to be an ad for a car company is a mighty intimidating thing. And when you think of all the things that have been written over the years by all the people who ever wrote then it gets even more worrisome and that first thing you have to say can begin to look like it's never ever going to get there out on the paper, but sonofagun if I don't open my magazine every Thursday or Friday and there it is, the column I never thought I'd even get started!

But there I go thinking about things that I don't like. I really do like lots of things and I'm not ashamed to admit it. Nor am I ashamed to list them in bold print in a national sports publication. I'm an up-front man about my fondness for things and I don't think it's important to go into why I like them either. I think it's just enough to admit to it and get onto the next thing on the list. Who needs a bunch of descriptions and puncuation? Not this reporter's reporter. Leave that for the young cubs who are still wet behind the ears and think they have to follow everything they learned in their fancy college sports page. Back when baseball writers were baseball writers, they didn't even have colleges. If a man couldn't think of a word, he'd just make one up. Those were better days indeed. I've been around too long to be bothered with sentence structure and puncuation. I've seen too much and know too much. These young punk editors put all that flotsam in anyway. What do they know? Do they have any idea what a royal pain in the keester it is to slog out a column EVERY week of your life?

But I do like things and am very secure in what I like. I decided what I liked when I was eleven and haven't changed my mind since. If something new comes along I either ignore it or don't like it out of principle. Things were great when I was eleven, except for the war (and who says that was so bad anyway?), and nothing has happened since then that I could possibly like.

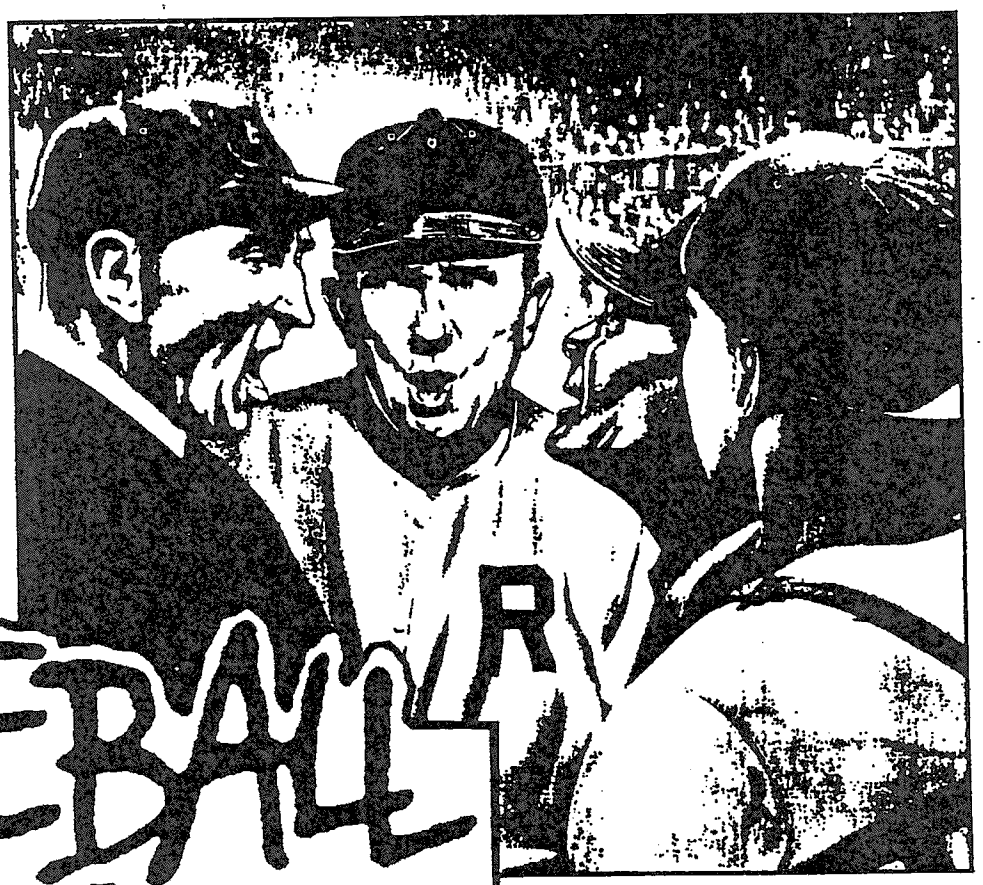
Here is what I like that comes to my mind right now:

Sleeping late...Train rides to St. Louis...Day games...Holidays...Pay Day...Day Games on Holidays...Day Games on Holidays right after pay day...Sleeping late on a train to St. Louis for a holiday day game right after pay day...Newspaper strikes...Vacation...Calling in sick...Retirement...Death

VOLUME 18
June 1985

Editor-Jim Baker
Founder/publisher--
Bill James
Business Manager--
Susie McCarthy
Statistical Consultant--
John Borkowski

Editorial Office
20 Lincoln Street
Jersey City, NJ 07307
Business Office
Box 171
Winchester, KS 66097



BASEBALL Analyst

NOTES: In this issue...Alden Mead; "Comparing Statistics From Different Eras"...Dan Greenia's Freak Show...Dallas Adams continues the discussion on the importance of getting the lead-off runner on, started by Chuck Waseleski....Dan Heisman; "Quantity vs. Quality: A Look at Linear Weights Per Game"...Scott Segrin contributes "Divergence of Won-Loss Records" and also a letter on the last page...Mike Kopf has written "New Worlds to Conquer," a book review of Kevin Kerrane's Dollar Sign on the Muscle which has just come out in paperback at a price we can all afford.

I hate to beg, but we're running short on articles again. It's kind of weird given that the availability of relevant statistical information has been increased by 500% in the last couple of years. Perhaps we're all too busy digesting the stuff to comment on it? Let's have an article on the paucity of offensive production in the NL. Those among you who have read Bill Conlin's column in the June 3 TSN know there is a better explanation for the decrease than the one he offers. (But then, there is usually a better explanation for anything that he comments on.) Things are happening, games are being played, men are putting on uniforms every day and compiling statistics so that we can have something to write about. Come on: LET'S DO IT!

ABOUT THE COVER: There is no explanation for the cover.

JOE FILLER comes to us courtesy of THE SPITTING NEWS. He is the first of a planned series of "Guest columnists" that we will be bringing your way.

COMPARING STATISTICS FROM DIFFERENT ERAS

by Alden Mead

(Based on a talk given at Minnesota Regional SABR Meeting, 1984)

Everyone interested in Baseball history enjoys arguing about the relative merits of players who performed in different periods, and under different conditions. For example, does Ty Cobb's incredible .366 lifetime batting average prove that the Georgia Peach was superior to the best modern players (at least in hitting for average), or does it merely prove that hitting for a high average was easier in Cobb's time? It's great fun, of course, just to argue about such matters without having to reach any conclusion; but as Sabrmetricians we feel happier if we have some more or less objective means of comparison. In this article, I'd like (except for some short remarks at the end) to confine myself to the problem of comparing <u>batting averages</u>. I'll discuss the best known approach used up to now (the relative batting average, or RBA), indicate why I think it is not necessarily the best way to do it, and suggest another simple approach which I think has some advantages. I'll also make some comparisons of the results of my approach with those of other methods.

In comparing averages from different years, it seems obvious that a relevant factor to consider is the <u>league</u> batting average. To take a classic example, Bill Terry's .401 average compiled in 1930 would seem to be at least somewhat tainted by the fact that the National League that year, including pitchers, batted .303. Because of the juiced-up ball, or for whatever reason, it seems likely that high averages were easier to come by in 1930 than at other times. One might well ask whether Terry's achievement was really any greater than (say) Carl Yastrzemski's .326 average in 1967, a year in which the American League batted .236. It is just this kind of question which a mathematical comparison approach must try to answer. In this article (again, apart from some short remarks at the end), we'll take it for granted that <u>comparison with the league average</u> is the approach to take in answering such questions. The problem that will concern us is the way in which the comparison should be made.

A simple method which rather obviously suggests itself is simply to divide the player's batting average (BA) by the league BA. This yields the relative batting average, or RBA, defined by:

RBA = (player's BA)/(league BA)

Published discussions of the RBA which I have read are the interesting 1976 article by David Shoebotham [1], and the excellent book by John Thorn and Pete Palmer [2]. Thorn and Palmer also refer to an article by Merritt Clifton [3], which I, however, haven't seen.

If desired, the RBA can also be refined in various ways, such as omitting the player's own at-bats and hits from the league averages; omitting pitchers' at-bats and hits in the same way; or making a "park adjustment" [4]. I won't make any of these refinements here, either for RBA or other methods of comparison, not because I disapprove of the refinements, but because they are irrelevant to the main point I m trying to make.

The RBA provides an immediate and simple answer to the Terry-Yaz controversy discussed above. Terry's RBA in 1930 was .401/.303 = 1.32; Yaz's in 1967 was .326/.236 = 1.38. On this basis, it looks like Yaz should be declared the winner. If one assumes that RBA's are constant, e.g., that the Terry of 1930 would have had an RBA of 1.32 in any year, one can estimate how Terry and Yaz would have performed if magically transported into each other's times. On this assumption, the Yaz of 1967, if transported into the NL of 1930, would have compiled an average of 1.38 x .303 = .418, winning the batting title in a walk. The Terry of 1930, if forced to play in the AL of 1967, would have done no better than 1.32 x .236 = .312, barely nosing out Frank Robinson for second place in the batting race. The matter would appear to be settled.

But is it <u>really</u> settled? Is it really clear that Yaz outperformed the average of his league to a greater degree than Memphis Bill? It is if one accepts uncritically the notion that the thing to do is to <u>divide</u> the averages. But, after all, <u>why divide</u>? Why not, for example, <u>subtract</u>? As far as I am aware, no one who has advocated the use of RBA has given any reason for preferring division over subtraction. For that matter, there are any number of other things one <u>could</u> do, such as subtracting the square roots, etc. If subtraction is the correct approach, then Terry beats Yaz after all, having exceeded his league's average by .098 to Yaz's .090. Based on this, Terry would have batted .236 + .098 = .334 in the 1967 AL, winning the title handily; Yaz in the 1930 NL would have stroked the ball at a clip of .303 + .090 = .393, tying Babe Herman for second place. The question is, which comparison method is correct, or is there a third method better than either?

I'm afraid there is no objective answer to the question of who has eclipsed the average of his league to a greater extent, since the word "extent" has not been given a precise meaning. What one has to do is make a precise definition, chosen judiciously, of what it is one is trying to measure, and then go ahead, accepting the fact that someone else might find another quantity more interesting for another purpose.

In comparing the BA of a superior hitter with the league average, it seems to me that a reasonable basis for comparison is the improvement in BA that the average player might reasonably hope to achieve (in the same number of at-bats as the superior player) <u>without</u> actual improvement in his ability. This, however, is well known from probability theory, and is given by the standard deviation, or SD, defined as:

$$SD = (\text{Sq Root})\left\{\frac{(LG\ BA)\times[1.000-(LG\ BA)]}{(PLAYER'S\ AB)}\right\}$$

For example, to calculate the SD for Terry in 1930, we first have to look up his AB total, which was 633. The league BA was .303, and 1.000 minus this is .697. So we multiply .303 by .697 and divide by 633, getting .0003336. The SD is then the square root of this, or .0183. For Yaz in 1967, who had 579 at-bats, a similar calculation gives an SD of .0176.

The significance of the SD is that it is a measure of the number of points by which an average player might hope to exceed the league BA <u>just by chance</u>. Put another way, it measures the number of points by which a table game card set to the league average might exceed that average in the given number of at-bats. Such a card would have about a 15.7% chance of exceeding the league BA by one SD or more, only about a 2.3% chance of beating the league BA by two or more SD, and still less for greater margins. Further details are to be found in another article of mine [5], in which the SD is applied to some other problems. One gets a measure of the degree to which a given batting average is out of reach of the average hitter by subtracting the league BA from the player's BA and dividing the result by the SD. This defines the standard deviation distance, or SDD:

$$SDD = \frac{(PLAYER'S\ BA) - (LEAGUE\ BA)}{SD}$$

The greater the SDD, the more overwhelming the odds against an average hitter equaling or surpassing the given BA in the same number of at-bats. The SDD is thus a measure of the degree to which the player's performance is unachievable for the average player. I think it is a more appropriate measure than either RBA or difference. To return to our sample problem, Terry had an SDD of 5.38 in 1930, Yaz had 5.12 in 1967. On the basis of SDD, Terry beats Yaz.

A few comments about the SDD, its relation to other methods, and the results:

(i) The <u>mathematical fact</u> represented by the SDD is the degree of improbability of an average hitter (or a table game card set to the average) equalling or surpassing the league leader by chance. Thus, there is simply no denying that such a card would have less chance of equalling Terry than of equalling Yaz. [The odds are overwhelming in both cases, of course: about one chance in 10,459,000 in Terry's case, one in 2,056,000 in Yaz's; but even though both numbers are huge, the difference between them is significant.] Is it, then, a mathematical fact that Terry's achievement was greater than Yaz's? <u>No</u>, because "magnitude of achievement" is not mathematically a precisely defined quantity. I do think, though, that the SDD is highly relevant to the question of which achievement was greater, considerably more so than RBA.

(ii) The RBA essentially measures the difference

between player and league averages in units of the league average, the subtraction approach in units of absolute points. The unit used in the SDD is the SD, which goes roughly as the square root of the league average, and is thus in between the other two. Compared to SDD, RBA tends to favor players in low-hitting leagues, while subtraction favors players in high-hitting leagues.

(iii) The SDD depends not only on the player's average, but on his AB total: with the same BA, the more AB the greater the SDD. I think this is entirely appropriate: it is a more solid accomplishment to maintain a high BA over a large number of AB than with a smaller number. The mathematical reason for this is that large chance fluctuations are more probable with a smaller AB total, so that more AB mean greater degree of certainty that the high BA was no fluke but due to really superior stickwork.

(iv) Because the SDD depends on AB as well as player and league BA, any attempt to calculate what a player of one era would have batted in another must include an assumption about AB. If the Yaz of 1967 had played in the 1930 NL, and had batted 633 times, as did Terry, while maintaining his SDD of 5.12, he would have batted .397, good for second place. The same calculation for Terry in 1967 with Yaz's total of 579 AB gives a BA of .331 and the batting title. I think it would be a mistake to take these numbers too seriously, but it might be a bigger mistake to take seriously the analogous projections based on RBA.

(v) The table on the next page lists the top thirty since 1900 in both RBA and SDD. There are a number of similarities and differences between the two lists. The similarities show that the truly outstanding performances will be recognized as such by any system that is not absolutely grotesque.

(vi) Ty Cobb dominates both tables, appearing 10 times in the RBA table, 8 times in SDD. In general, both tables are dominated by American Leaguers, largely of the pre-WWI era. Cobb, Lajoie, Jackson, Speaker, Sisler, and later Williams certainly outclassed their contemporaries, by any measure, in hitting for average. The only National Leaguers to make either list more than once are Honus Wagner (twice in RBA) and Stan Musial (twice in SDD).

(vii) Both lists discriminate against American Leaguers playing under the DH rule, because I haven't subtracted pitchers' stats from the non-DH leagues and years. Despite this, Rod Carew and George Brett make the RBA list and Carew the SDD list as well.

(viii) For the controversial year of 1910, the table shows the "official" stats for Cobb and Lajoie. Note that Lajoie still beats Cobb in SDD because he had more AB. According to modern research, one should subtract three AB and two hits from Cobb's total. This would still round off to an RBA of 1.58, but would put him behind Lajoie. The revised SDD would be 7.36.

THE TOP THIRTY IN RBA AND SDD

	PLAYER, YEAR	RBA	PLAYER, YEAR	SDD
1	Ty Cobb, 1910	1.58	Ty Cobb, 1911	8.02
2	Nap Lajoie, 1910	1.58	Nap Lajoie, 1910	7.99
3	Tris Speaker, 1916	1.56	Ty Cobb, 1912	7.73
4	Ty Cobb, 1912	1.55	Ty Cobb, 1917	7.58
5	Nap Lajoie, 1904	1.55	Nap Lajoie, 1901	7.55
6	Ty Cobb, 1909	1.55	Ty Cobb, 1910	7.47
7	Ty Cobb, 1917	1.54	Tris Speaker, 1916	7.47
8	Ty Cobb, 1911	1.54	Ty Cobb, 1909	7.41
9	Ted Williams, 1941	1.53	Nap Lajoie, 1904	7.32
10	Nap Lajoie, 1901	1.52	George Sisler, 1922	7.30
11	Ty Cobb, 1913	1.52	Rogers Hornsby, 1924	7.25
12	Ted Williams, 1957	1.52	Joe Jackson, 1911	7.24
13	Ty Cobb, 1918	1.50	Joe Jackson, 1912	7.04
14	Rogers Hornsby, 1924	1.50	George Sisler, 1920	6.91
15	Ty Cobb, 1916	1.50	Rod Carew, 1977	6.85
16	Joe Jackson, 1911	1.49	Ted Williams, 1941	6.77
17	Joe Jackson, 1912	1.49	Cy Seymour, 1905	6.75
18	Ty Cobb, 1915	1.49	Ty Cobb, 1915	6.65
19	Ty Cobb, 1914	1.48	Ty Cobb, 1916	6.63
20	Honus Wagner, 1908	1.48	Stan Musial, 1948	6.47
21	George Sisler, 1922	1.48	Tris Speaker, 1912	6.44
22	Cy Seymour, 1905	1.48	Joe Torre, 1971	6.44
23	Rod Carew, 1977	1.46	Honus Wagner, 1908	6.43
24	Joe Jackson, 1913	1.46	Ty Cobb, 1913	6.35
25	George Brett, 1980	1.45	Ted Williams, 1957	6.25
26	Tris Speaker, 1912	1.45	Stan Musial, 1946	6.24
27	Stan Musial, 1948	1.44	Joe Jackson, 1913	6.16
28	Joe Torre, 1971	1.44	Harry Heilmann, 1923	6.16
29	Honus Wagner, 1907	1.44	George Stone, 1906	6.08
30	George Sisler, 1920	1.44	Roberto Clemente, 1967	6.04

(ix) Where there are apparent ties in the table, the computer has carried the calculation to more decimal points and ordered them correctly.

One can do similar calculations, of course, with other stats, but I think a great deal of caution is always in order. A low league average in some statistic does not always mean greater difficulty; it can often just reflect varying tactical fashions. An extreme example is stolen bases. When George Case stole 51 bases in 1939, this represented a percentage of .273 of the estimated times he reached first base (singles plus walks), compared to a league average of .015, leading to an SDD of 15.05. When Tim Raines led the National League with 78 thefts in 1982, this represented a percentage of .371 compared with a league average of .096, and an SDD of 13.50. Should we rank Case ahead of Raines as a base stealer? I don't think so. The low league average in 1939 does not prove that base stealing was more difficult then. If anything, it was probably a bit easier. It was just not fashionable. In all such comparisons, one must carefully watch out for this sort of thing. One reason for choosing the batting average for the main emphasis in this article was that it is less subject than other stats to varying fads

and fashions. While the willingness to sacrifice average for power has varied greatly from era to era, there has never been a time when managers forbade their players to try to get base hits, as they forbade them to steal bases in the thirties.

I think that the SD can be a very useful tool in many aspects of the analysis of baseball statistics. This article takes up one such aspect, and I hope it will provoke some discussion.

References

[1] David Shoebotham, "Relative Batting Averages", Baseball Research Journal, 1976, pp. 37-42.

[2] John Thorn and Pete Palmer, The Hidden Game of Baseball (Doubleday, 1984), Chapter 6, pp. 102-121.

[3] Merritt Clifton, "Relative Baseball", Samisdat, 1979.

[4] Thorn and Palmer, Reference 2, Chapter 5, pp. 82-101.

[5] Alden Mead, "Probability and Fluctuations in Baseball, scheduled for Baseball Research Journal, 1984.

Pitcher Burnout(?)

By Daniel Greenia

As Oakland Manager from 1980-82, Billy Martin allowed his starters to pitch an extraordinary number of complete games to compensate for a weak bullpen. In 1980, 94 CG, with rest of AL averaging 35; 1981-60 (league 21); 1982- 42 (league 31). Subsequently, each pitcher's performance declined drastically.

1979		1980					1981		1982	
IP	ERA		Age	IP	ERA	GS	CG	IP Proj.	ERA	IP	ERA
219	4.27	R Langford	28	290	3.26	33	28	195(290)	3.00	237	4.21
146*	4.81	M Norris	25	284	2.54	33	24	173(257)	3.75	166*	4.76
177	5.03	M Keough	25	250	2.92	32	20	140(208)	3.41	209	5.72
186	4.21	S McCatty	26	222	3.85	31	11	186(276)	2.32	129*	3.99
113	4.30	B Kingman	26	211	3.84	30	10	100(149)	3.96	123	4.48

	1983		1984	
Langford	20*	12.15	9*	8.31
Norris	89*	3.76	--*	----
Keough	100	5.33	AA ball	
McCatty	167	3.99	180	4.76
Kingman	5	7.71	AAA ball	

My conclusion? Not that overwork ruined young arms, rather that Martin and Art Fowler (spitball tutor) got the most out of marginal talent. These guys weren't much before Martin nor after.

* spent time on disabled list.

SOME ADDITIONAL COMMENTS ON THE BENEFIT OF GETTING THE LEADOFF BATTER ON BASE

BY DALLAS ADAMS

IN ISSUE 11 OF THE BASEBALL ANALYST, CHUCK WASELESKI PRESENTED DATA FOR ALL 1983 BOSTON RED SOX GAMES. THE DATA SHOWED THAT THE AVERAGE HALF-INNING IN WHICH THE LEADOFF BATTER REACHED BASE WAS WORTH 0.36 RUNS, WHILE THE AVERAGE SCORING IN HALF-INNINGS WHERE THE LEADOFF BATTER MADE OUT WAS ONLY 0.28 RUNS. WASELESKI'S IMPLICATION WAS THAT IT WAS THE SUCCESS OR FAILURE OF THE HALF-INNING'S FIRST BATTER WHICH CAUSED THIS DIFFERENCE IN SCORING.

CHARLES HOFACKER, IN ISSUE 13 OF THE ANALYST, DEMURRED SOMEWHAT:

> 'IT IS TEMPTING TO THINK OF THE RUNNER AT FIRST AS BEING THE SOLE CAUSE OF THOSE EXTRA RUNS. IF SUCH WERE THE CASE, WE COULD ESTIMATE HOW MANY MORE RUNS WE WOULD GET IF WE HAD A LEADOFF BATTER WHO GOT ON, SAY, 30 MORE TIMES PER YEAR WHILE LEADING OFF. ... BUT THE HALF-INNINGS IN WHICH THE FIRST BATTER REACHES FIRST MAY NOT BE REPRESENTATIVE OF ALL HALF-INNINGS. IN FACT, I WOULD EXPECT THESE HALF-INNINGS TO BE RELATIVELY POORLY PITCHED COMPARED TO THE AVERAGE INNING. WHEN WE COMPARE HALF-INNINGS IN WHICH A LEADOFF BATTER REACHED AGAINST INNINGS IN WHICH THE LEADOFF BATTER WAS OUT, WE ARE ALSO COMPARING PITCHERS WHO GOT THE FIRST BATTER WITH PITCHERS WHO DID NOT. NO DOUBT THE LATTER CATEGORY INCLUDES MORE BAD PITCHERS AND MORE TIRED PITCHERS. CONSEQUENTLY THE DATA PROBABLY OVERSTATE THE CAUSAL BENEFIT OF GETTING THE FIRST BATTER ON BY SOME UNKNOWN DEGREE.'

A VERY PERCEPTIVE OBSERVATION.

HOFACKER SUGGESTED THAT THE TRUE DIFFERENCE COULD BE OBTAINED BY COMPARING SCORING IN INNINGS WHERE THE LEADOFF BATTER REACHED BASE BY MEANS OF AN ERROR (I.E. IN A MANNER INDEPENDENT OF THE PITCHER'S SKILL AND/OR FATIGUE) AGAINST THOSE INNINGS WHERE THE LEADOFF MAN MADE OUT.

UNFORTUNATELY, THIS APPROACH ALSO SUFFERS FROM THE 'HOFACKER EFFECT': THOSE HALF-INNINGS WHERE THE FIRST BATTER MADE OUT ARE PROBABLY CHARACTERIZED BY HAVING MORE GOOD PITCHERS AND MORE FRESH, RESTED PITCHERS, AND THUS ARE NOT TYPICAL OF ALL HALF-INNINGS.

MY PURPOSE IN WRITING THIS PAPER IS TO SHOW THAT THERE EXISTS AN ANALYTICAL METHOD BY WHICH THE TRUE SCORING ADVANTAGE OF GETTING AN INNING'S FIRST BATTER ON BASE CAN BE CALCULATED. THIS ANALYTICAL METHOD IS INDEPENDENT OF THE SKILL OR FATIGUE OF THE OPPOSING PITCHER.

OVER A FULL SEASON, THE TOTAL NUMBER OF RUNS SCORED EQUALS THOSE SCORED IN HALF-INNINGS WHERE THE FIRST BATTER REACHED BASE SAFELY PLUS THOSE SCORED IN HALF-INNINGS WHERE THE FIRST BATTER MADE OUT. HENCE THE EXPECTED NUMBER OF RUNS IN AN AVERAGE HALF-INNING EQUALS THE WEIGHTED AVERAGE OF (1), RUNS PER HALF-INNING WHEN THE FIRST BATTER REACHES BASE PLUS (2), RUNS PER HALF-INNING WHEN THE FIRST BATTER MAKES OUT: THE WEIGHTING FACTORS BEING, RESPECTIVELY, THE PROBABILITY OF THE FIRST BATTER REACHING BASE, AND OF HIS NOT REACHING BASE. SYMBOLICALLY:

$$A = (Y)(OBA) + (N)(1-OBA)$$

A = NUMBER OF RUNS SCORED IN AN AVERAGE HALF-INNING
Y = AVERAGE NUMBER OF RUNS PER HALF-INNING WHEN THE FIRST BATTER REACHES BASE SAFELY

N = AVERAGE NUMBER OF RUNS PER HALF-INNING WHEN THE FIRST BATTER FAILS TO REACH BASE SAFELY
OBA = ON BASE AVERAGE (I.E. THE PROBABILITY OF THE FIRST BATTER REACHING BASE SAFELY)

THE EQUATION ABOVE CAN BE REARRANGED TO:

$$Y = (A - (1-OBA)(N))/OBA \qquad \text{(EQN. 1)}$$

NOW, OVER THE COURSE OF A FULL SEASON A TEAM SCORES A KNOWN NUMBER OF RUNS AND DOES IT AT THE EXPENSE OF A CALCULABLE NUMBER OF OUTS. THEREFORE IT IS A SIMPLE MATTER TO COMPUTE THE TEAM'S RUNS/OUT RATIO FOR THE SEASON.

GENERALLY, THE NUMBER OF OUTS A TEAM MAKES WHILE BATTING IS NOT PUBLISHED PER SE, BUT IT CAN BE EASILY DETERMINED. THE TOTAL NUMBER OF PLATE APPEARANCES COMPILED BY A TEAM'S BATTERS EQUALS AB+BB+HBP+SH+SF+INT (WHERE 'INT' IS BATTERS REACHING BASE ON INTERFERENCE). FOR EVERY ONE OF THOSE PLATE APPEARANCES, ONE OF THREE THINGS EVENTUALLY HAPPENS:

(1) THE BATTER SCORED A RUN
(2) THE BATTER WAS LEFT ON BASE
(3) THE BATTER WAS PUT OUT (EITHER BEFORE OR AFTER REACHING BASE)

HENCE FOR THE TEAM, $PA = R + LOB + OUTS$

WHICH REARRANGES TO, $OUTS = PA - R - LOB$

THUS: $$OUTS = AB + BB + HPB + SF + SH + INT - R - LOB \qquad \text{(EQN. 2)}$$

AND ALL THE TERMS ON THE RIGHT SIDE OF EQUATION 2 ARE REPORTED IN THE OFFICIAL STATISTICS; THEREFORE THE NUMBER OF OUTS A TEAM MAKES WHILE BATTING CAN BE CALCULATED.

'N' WAS DEFINED AS THE AVERAGE NUMBER OF RUNS SCORED PER INNING BATTED WHEN THE FIRST BATTER FAILED TO REACH BASE. WHEN THE FIRST BATTER DOES SO FAIL, THE TEAM HAS TWO-THIRDS OF AN INNING REMAINING IN WHICH TO SCORE. ITS EXPECTED RUN PRODUCTION IN TWO-THIRDS OF AN INNING IS TWO TIMES ITS SEASONAL RATIO OF RUNS/OUTS. THAT IS:

$$N = 2(RUNS/OUT) \qquad \text{(EQN. 3)}$$

SIMILARLY, 'A' WAS DEFINED AS THE AVERAGE NUMBER OF RUNS SCORED PER INNING BATTED (WITHOUT CONSIDERATION AS TO WHAT THE FIRST BATTER DID). 'A' THUS REPRESENTS THE TEAM'S SCORING EXPECTATION WHEN IT HAS ALL THREE OUTS (A FULL INNING) TO WORK WITH. THUS,

$$A = 3(RUNS/OUT)$$

EQUATION 1 CAN NOW BE REWRITTEN AS:

$$Y = (3(RUNS/OUT) - 2(1-OBA)(RUNS/OUT))/OBA$$

WHICH SIMPLIFIES TO:

$$Y = (2(OBA)+1)(RUNS/OUT)/OBA \qquad \text{(EQN. 4)}$$

EQUATION 3 YIELDS THE EXPECTED NUMBER OF RUNS WHEN THE FIRST BATTER IN A HALF-INNING FAILS TO REACH BASE SAFELY. EQUATION 4 GIVES THE EXPECTED RUNS WHEN THE FIRST BATTER DOES REACH BASE. NOTE THAT THESE TWO EQUATIONS USE THE TEAM'S SEASONAL TOTALS OF RUNS SCORED AND OUTS MADE (AND OBA, ALSO); TOTALS WHICH ARE COMPILED IRRESPECTIVE OF

WHETHER OR NOT THE LEADOFF BATTER REACHED BASE. THUS THE 'Y' AND 'N' VALUES ARE INDEPENDENT OF THE SKILL/FATIGUE OF THE OPPOSING PITCHER. THE RUN VALUE OF GETTING THE LEADOFF BATTER ON BASE EQUALS 'Y' MINUS 'N'. EQUATIONS 3 AND 4 COMBINE TO GIVE: Y-N = (RUNS/OUT)/OBA.

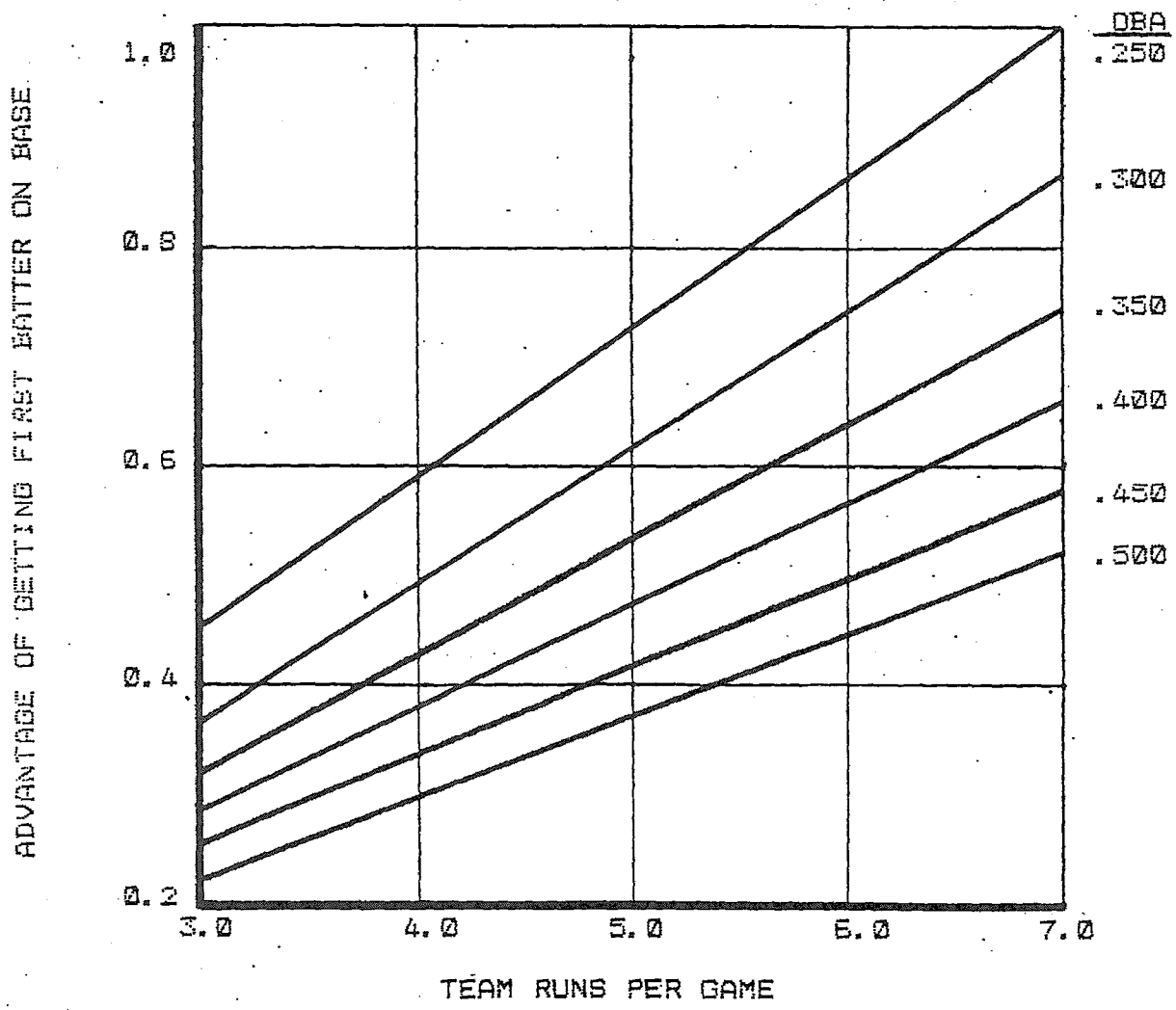
FOR THE 1983 RED SOX AND THEIR OPPONENTS: THE SOX SCORED 724 RUNS IN 4344 OUTS (CALCULATED PER EQUATION 2); THEIR OPPONENTS, PER SOX PITCHING STATS, SCORED 775 RUNS FOR 4339 OUTS (SOX PITCHERS WORKED 1446.1 INNINGS, WHICH EQUALS 4339 OUTS). THAT'S A COMBINED RUNS/OUT RATIO OF .17264; PER WASELESKI'S DATA, THE COMBINED OBA OF FIRST BATTERS WAS .350 (INCLUDING REACHING BASE ON ERROR). THESE NUMBERS WERE SUBSTITUTED INTO EQUATIONS 3 AND 4 TO FIND THAT IN THEORY:

.83854 RUNS/HALF-INNING RESULTED WHEN THE LEADOFF MAN GOT ON
.34527 RUNS/HALF-INNING RESULTED WHEN THE LEADOFF MAN WAS OUT

THE DIFFERENCE, .49327 RUNS PER HALF-INNING, CAN BE TAKEN AS THE THEORETICAL VALUE IN 1983 RED SOX GAMES OF GETTING THE LEADOFF BATTER ON BASE. SO, THE 'HOFACKER EFFECT' REDUCES WASELESKI'S VALUE BY 27%.

TO COMPLETE THIS PROJECT I MADE A PARAMETRIC STUDY, EVALUATING EQUATIONS 3 AND 4 FOR A WIDE RANGE OF RUN SCORING RATES AND ON BASE AVERAGES. THE RESULTS SHOWN GRAPHICALLY, ARE FOR 'Y' MINUS 'N': THE RUN VALUE OF GETTING THE FIRST BATTER ON BASE IN AN INNING. THE GRAPH USES THE MORE FAMILIAR PARAMETER 'RUNS PER GAME' IN PLACE OF 'RUNS PER OUT'; FOR THE CALCULATIONS, I ASSUMED 27 OUTS PER GAME.

ADVANTAGE IN RUNS/HALF-INNING OF GETTING THE FIRST BATTER ON BASE



DAN HEISMAN:

Quantity vs. Quality: A Look at Linear Weights Per Game

In 1983 I did an article for The Baseball Analyst called "Quantity versus Quality", whose basic theme was that players that stick around for long careers seem to have a better "superstar" recognition than others of comparable statistics who have played shorter careers. This tendency seems even more pronounced today, with the proliferation of record keeping. For example, the stock of such players as Jim Kaat, Gaylord Perry, Willie Stargell, Carl Yastrzemski, Phil Niekro, etc. gained considerably in value past their 35th birthday as they became the grand old men of baseball, while such players as Tony Oliva, Eddie Mathews, Ron Santo, Richie Ashburn, and Juan Marichal never quite achieved this stature due to shorter, though no less qualitative careers. As an example, I showed that Ashburn and Rose had about the same number of hits at the age Ashburn retired (and was still hitting over .300), but that Ashburn went 15 years without election to the Hall of Fame by the BBWAA, while Rose has gone on to play til his mid-40's, achieve over 4000 hits, and demi-god status in some eyes.

Last year John Thorn and Pete Palmer published their excellent book, "The Hidden Game of Baseball", in which they showed the high correlation of Palmer's statistic, Linear Weights, to offensive productivity. I immediately wondered how this information could be used to compare qualitative and quantitative statistics. In the back of the book the authors present a list of "Overall Games Won", a quantitative number of games won by players over their career. I thought it would be interesting to convert this quantitative information to a "per game" value, a qualitative statistic. For example, while Joe DiMaggio is rated 23rd in overall wins, he rises to 16th in wins per game. The entire set of top twenties for both non-pitchers and pitchers is given in the accompanying table.

What conclusions can be drawn from the comparison of the two lists? The non-pitcher lists contain many of the same names in the quantitative and qualitative lists. It is interesting to note that Mike Schmidt rates a spectacular third in the qualitative list (after the 1983 season) behind Babe Ruth and Ted Williams. But all in all, it seems that the best quantitative non-pitchers were also the best qualitative non-pitchers, so my comparisons would mean little.

However, on the pitcher lists there is much discrepancy, so it is possible to draw some conclusions about quantity vs. quality. If the list of better known and Hall of Fame pitchers appear on one list but not another, then an indication of which is weighted more in HOF elector's (and the general public's) minds can be made.

And, indeed, this is the case! The cumulative list reads like the Who's Who of Baseball pitchers: Johnson, Young, ALexander, Mathewson, Grove, Seaver, Gibson, Spahn, etc. But at the top of the per game list there are different orderings: Grove, Walsh, Wilhelm, Johnson, Newhouser, Trout, Alexander, Lemon, Brecheen, etc. I think any impartial observer with a baseball background would agree that the cumulative list contains the more conventional (and thus "higher weighted") order of pitching skill. Yet the proficiency, not the cumulation, shown in the per game list theoretically should have the higher correlation. Thus, from this set of lists I am forced to conclude that this impartial data set well backs my thesis of quantity given more weight than quality in HOF selections or other "All-Star" considerations. The subject is a living one, and I would be interested in hearing of others on the subject.

DIVERGENCE OF WON-LOSS RECORDS

by Scott Segrin

For his comment on the Milwaukee Brewers in the 1983 edition of the Baseball Abstract, Bill James counted, for three teams, the number of times they were ahead, behind or tied after each inning of their games. He used this data to 'study the divergence of a team's won-loss record from .500 throughout the progression of a ball game'. For example, if after one inning of all of its games, a team was ahead 35 times, behind 25 times and tied the other 102 times, their divergence from .500 for the first inning would be +10 games (actually it's +10 half games, but I'm going to call them games for this study). This figure is calculated by subtracting the number of games behind from the number of games ahead; ties count half each way so they don't enter into the equation. What this says is that if you were to go back through the season's records and terminate all of this team's games after one inning, they would have a won-loss record of 10 gaves over .500. Once this divergence is calculated for all nine innings and the end of the game, a graph can be drawn which shows how this divergence progressed or regressed as the game went on. Once we have this information for all teams, we can look at the shape of the graphs and see which were good late inning ball clubs, and which played better in the early innings. We can see if there are any characteristics of good or bad clubs, or clubs with good or bad bullpens, or clubs with good or bad starting pitching. This is probably only part of the complete list.

What I have done is to repeat this study for the entire American League, 1984. In Table I, I have presented the data and on the following page, the graphs. The purpose of James' study, I think, was to determine whether more games are won in the early or late innings, and then to determine whether teams with a good bullpen tend to do well in the late innings. I will examine each of these topics seperatly as well as present a few other observations I have made. But before doing so, I will explain how I have presented the data. In Table I, the first two columns for each team are the number of times they were ahead (A) and behind (B) at the end of each particular inning. You can figure out how many times they were tied if you'ld like. The third column, the divergence (DIV), is simply column A minus column B. On the following page I have graphed the divergences for each team. The horizontal represents the same number of times being ahead as behind. A team that finishes on this horizontal is, of course, a .500 club. The graphs aren't labeled for two reasons, 1) I didn't feel like labeling all of them, and 2) The numbers aren't all that important. What we are really concerned with is the shape of the graphs. Do they diverge early or late? Do they curve up or down or are there some that don't curve much at all? Although there are many places where the graphs jump up or down quickly and for no apparent reason, if you just hold the page back a bit and maybe squint your eyes, you can see the general shape of each graph. This is the information we are really interested in. Besides, if you really want to see the graphs labeled, all of the numbers are right there on the previous page--just fill 'em in.

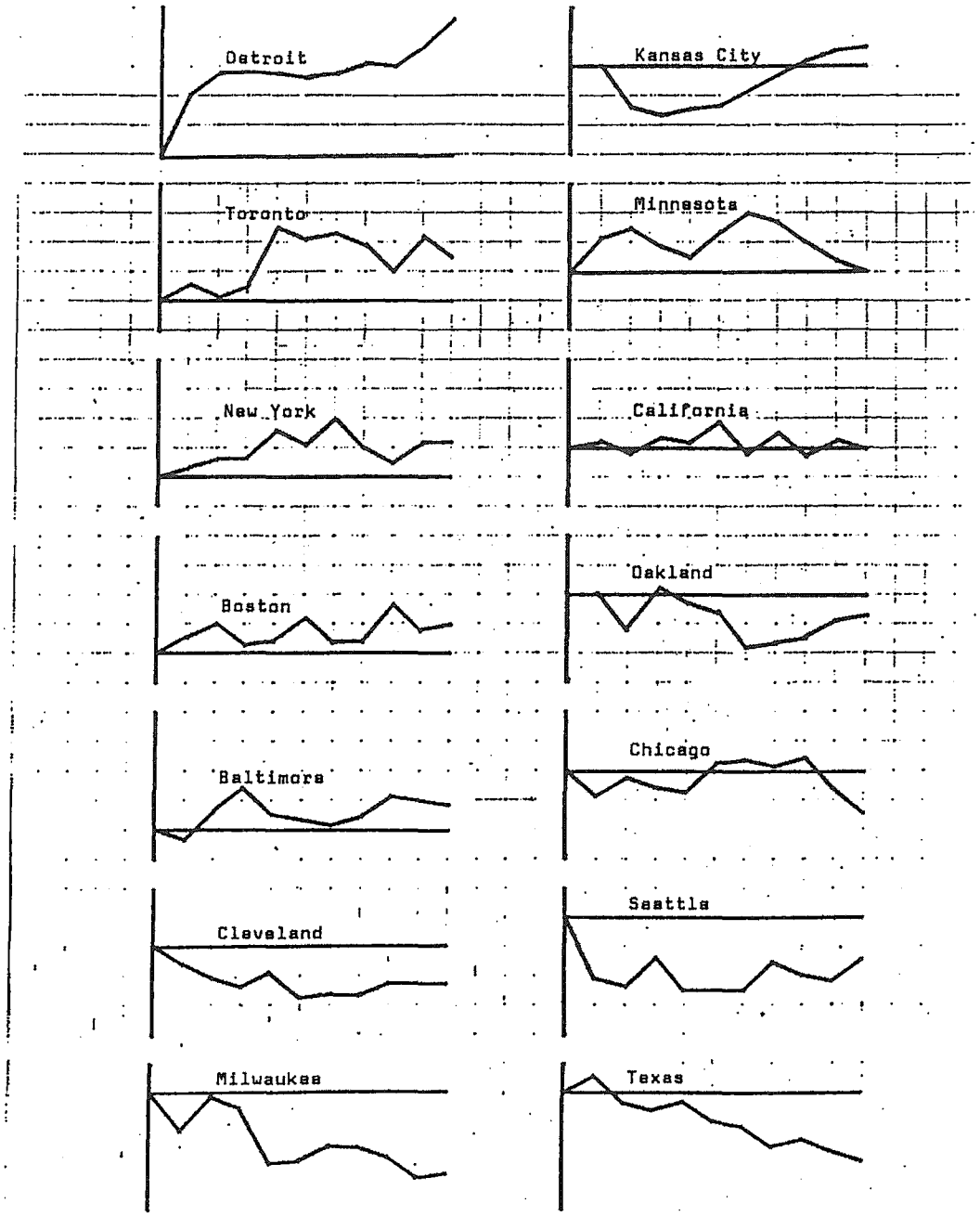
Table I

Inn	DETROIT A	DETROIT B	DETROIT DIV	TORONTO A	TORONTO B	TORONTO DIV	NEW YORK A	NEW YORK B	NEW YORK DIV	BOSTON A	BOSTON B	BOSTON DIV
1	50	30	+20	36	31	+5	38	35	+3	43	37	+6
2	71	43	+28	47	46	+1	54	48	+6	58	48	+10
3	75	47	+28	58	53	+5	58	52	+6	62	59	+3
4	80	53	+27	77	51	+26	70	54	+16	72	68	+4
5	81	55	+26	78	57	+21	70	59	+11	78	66	+12
6	85	58	+27	79	56	+23	79	59	+20	77	73	+4
7	89	58	+31	76	57	+19	75	65	+10	75	71	+4
8	88	58	+30	74	64	+10	75	70	+5	82	65	+17
9	93	56	+37	83	61	+22	76	64	+12	78	70	+8
F	104	58	+46	89	73	+16	87	75	+12	86	76	+10

Inn	BALTIMORE A	BALTIMORE B	BALTIMORE DIV	CLEVELAND A	CLEVELAND B	CLEVELAND DIV	MILWAUKEE A	MILWAUKEE B	MILWAUKEE DIV	KANSAS CITY A	KANSAS CITY B	KANSAS CITY DIV
1	31	34	-3	41	48	-7	27	40	-13	34	34	0
2	53	46	+7	49	60	-11	47	48	-1	45	58	-13
3	67	54	+13	53	66	-13	53	58	-5	51	68	-17
4	70	65	+5	61	70	-9	53	77	-24	61	75	-14
5	73	70	+3	62	80	-18	56	79	-23	64	77	-13
6	73	72	+1	61	77	-16	59	78	-19	67	75	-8
7	76	72	+4	61	78	-17	64	83	-19	72	75	-3
8	80	69	+11	67	79	-12	62	84	-22	74	72	+2
9	81	71	+10	68	80	-12	61	90	-29	78	73	+5
F	85	77	+8	75	87	-12	67	94	-27	84	78	+6

Inn	MINNESOTA A	MINNESOTA B	MINNESOTA DIV	CALIFORNIA A	CALIFORNIA B	CALIFORNIA DIV	OAKLAND A	OAKLAND B	OAKLAND DIV	CHICAGO A	CHICAGO B	CHICAGO DIV
1	42	31	+11	39	37	+2	39	39	0	37	46	-9
2	59	45	+14	48	49	-1	52	64	-12	50	52	-2
3	69	61	+8	57	60	-3	61	60	+1	62	67	-5
4	69	64	+5	66	64	+2	66	68	-2	65	73	-8
5	74	61	+13	73	64	+9	66	73	-7	73	71	+2
6	81	61	+20	71	72	-1	62	80	-18	73	70	+3
7	79	62	+17	76	71	+5	61	78	-17	73	71	+1
8	79	69	+10	71	73	-2	63	78	-15	73	70	+3
9	76	73	+3	72	70	+2	71	80	-9	71	78	-7
F	81	81	0	81	81	0	77	85	-8	74	88	-14

Inn	SEATTLE A	SEATTLE B	SEATTLE DIV	TEXAS A	TEXAS B	TEXAS DIV
1	27	48	-21	41	34	+7
2	40	63	-23	51	54	-3
3	50	64	-14	58	65	-7
4	53	78	-25	62	65	-3
5	56	81	-25	64	74	-10
6	57	82	-25	65	76	-11
7	64	80	-16	62	81	-19
8	63	83	-20	62	79	-17
9	62	84	-22	62	82	-20
F	74	88	-14	69	92	-23



IMPORTANCE OF EARLY AND LATE INNINGS

In James' study, he noted that the three teams he checked all diverged more in the early innings than they did in the late innings, supporting the theory that most games are decided in the early innings. I also notice this to be true. The average divergence from .500 through the first five innings is 14 games, whereas the average divergence from where the teams were after five innings to the end of the games is only 9 games. Further, a good part of that 9 comes from a few clubs that were either particularly good or particularly bad at the end. James also noted that the three teams he studied, identified in the early innings which direction they were going. With those results and the results of my study, it is safe to say that the good teams go ahead early and the bad teams fall behind early. In almost every case, (except Kansas City) the winning teams were winning and the losing teams were losing by the second inning.

I'm going to stray a bit from my current path, but there's something that seems to fit in here. As I was collecting the data for this study, I wrote down the inning in which the GWRBI (yes--that thing) occured for each American League game. Here are the results: (I know, three too many games. But you get the general idea.) Almost amazingly, one in five games are decided in the first inning, and over half are decided by the 4th. The analysis of this data might make an interesting study in itself, but I'm going to leave it as it is for now. From all of this, then, I'm going to conclude that, all other things being equal, a team has a better chance to win games by playing well in the early innings than it does by playing well in the late innings. Until someone shows me enough evidence to prove me wrong, I will believe this conclusion to be true. I'm not exactly why it is true. I'll save that for another day. I can offer one possibility though. Suppose your home team gives up four runs in the second inning of a game. The game pretty much remains that way; maybe they even come back and score a run in the fifth. Then in the ninth, down by three, they get a couple of runners on and your star hitter, who's up next represents the tying run. After he flies out to end the game you say "Boy, they were in it right to the end". You think the game was decided in the ninth because, after all they WERE still in it until the end--and in some sense you're right. But in hindsight the game was over after two innings. The feeling then, that most games are decided in the late inngings may be merely an illusion.

Table II

Inning	GWRBI's	Percentage
1	243	21.4%
2	122	10.7
3	118	10.4
4	112	9.9
5	89	7.8
6	90	7.9
7	73	6.4
8	91	8.0
9	97	8.5
XTR	101	8.9

EFFECTS OF HITTING AND PITCHING

I'm going to make an assumption before I start talking about pitchers. Without it there really isn't much you can say about pitchers in this study. If you believe it, you might believe the rest of what I have to say here. If not, you probably won't.

The assumption is that a team's hitting is fairly consistent throughout the course of its games. That is to say that over the course of a season a team will score roughly as many runs in one inning as the next. I have, written down on about 70 sheets of paper, the line scores for every American League game. I have looked at those sheets for a long time. Not nearly long enough to count the number of runs scored in each inning, but long enough to say that I would be very surprized if there were significantly more or fewer runs scored in the 4th inning than, say, the 8th. There are indeed more runs scored in the first inning than the second. That's because the top of the order bats in the first inning and the middle or bottom of the order bats in the second. But after that I don't think there's very much difference. If this assumption is true, then the effect of batting on the teams graphs shows up only as a straight line. Let me explain what I mean. If there were no pitchers in baseball, but only pitching machines, all of the graphs would look something like straight lines. They might slope up or down but they would still look like lines. This is because hitting is consistant and no one inning means any more than another. So what makes the graphs bend? The pitching. The pitching is not consistant as the game goes on. Pitchers don't play every day, they have rough starts and then settle down, they start out strong and then tire, they come out of games for relief pitchers who do well and for relief pitchers who do poorly. Get the idea?

RELIEF PITCHERS

It has become common belief in baseball these days that it's imposible to win a pennant without having a stopper in the bullpen. "After all, just look at the four playoff teams" the argument goes. I think I'd made a strong case earlier for the opposite train of thought. Since the majority of games have already been decided by the sixth inning, how important can a strong bullpen really be? There is, however, some evidence that suggests a relief ace is indeed a valuable asset to a ballclub. The three teams whos graphs increased most after the 7th inning--Detroit +15, Kansas City +9, Oakland +9--had arguably the three best relief pitchers in the league. This trend doesn't really hold for the rest of the league as shown in Table III, but three out of three isn't bad.

Table III

TEAM	DIVERGENCE AFTER 7TH INN	SAVES TOP RELIEVER	SAVES TEAM
Detroit	+15	32	51
Kansas City	+9	44	50
Oakland	+9	36	44
Boston	+6	22	32
Cleveland	+5	23	35
Baltimore	+4	17	32
New York	+2	31	43
Seattle	+2	8	35
Toronto	-3	10	33
Texas	-4	12	21
California	-5	11	26
Milwaukee	-8	23	41
Chicago	-15	12	32
Minnesota	-17	29	38

There were also cases where the lack of a quality reliever might have cost a team a chance at a pennant (Toronto), did cost a team a chance at a pennant (Chicago), and right out blew a pennant (Minnesota). Before I did this study I felt that a good relief pitcher does not make a good team, nor does the lack of a relief ace make a bad team. I still feel that way. But I also felt that the winning and losing of pennants should not be blammed on bullpens as it so often is. Now I'm not so sure. I am convinced that most games are won or lost in the early innings, but having a relief ace in there at the end of the close ones might be the ingredient need to put a team on top.

AN EXPERIMENT

Care to tread out onto some thin ice with me? Let's fool around with some of the graphs. Detroit had a divergence of +20 after the first inning. I will call this their first inning gain. In the second inning they went from +20 to +28, so their second inning gain is +8. Their third inning gain is zero, their 4th inning gain is -1, and so on. I have done this for Detroit and Kansas City and presented the results in Table IV. The bottom line, which I called Detroit's Inning Advantage, is Detriot's gain minus Kansas City's gain for that particular inning (for Kansas City's Inning Advantage just subtract the other way). But what does this line mean? This is where the ice begins to get thin. Perhaps we can predict the complexion of a Tigers-Royals game. From those numbers it looks like Detroit should jump out to a (sometimes big) lead early, then Kansas City should come back a bit before Detroit comes on to win the close ones. (I hope you can swim.) Well, look what really happened! In their 15 games (including the playoffs), The Tigers outscored the Royals 23-11 in the first three innings, were outscored 23-25 in the middle three, and then scored 29 to the Royals 20 in the end. The Tigers were also 3-1 in one-run games and won both extra-inning games. I'm certantly not suggesting that every case works this well. I'd be mildly surprized if even a few more worked this well. But it does make for something interesting to think about.

Table IV

Inning	1	2	3	4	5	6	7	8	9	F
Detroit gain	+20	+8	0	-1	-1	+1	+4	-1	+7	+9
Kansas City gain	0	-13	-4	+3	+1	+5	+5	+5	+3	+1
Detroit Inning Adv.	+20	+21	+4	-4	-2	-4	-1	-6	+4	+8

MINNESOTA TWINS

If you don't get anything else out of this study, please notice how the Twins blew the pennant in the West. Had the AL West games ended after six innings, the Twins would have won the pennant by 8½ games over Chicago and would have finished 14 games ahead of Kansas City. So that's why they were booing Ron Davis.

NEW WORLDS TO CONQUER

by Mike Kopf

There came a day, so historians tell us, when Alexander the Great surveyed his empire and wept that there remained for him no new worlds to conquer. Well, sabermetrics hasn't exactly conquered the world of major league baseball, but it's certainly established a firm beachhead; what with sabermetricians writing many of our era's best best baseball books, sabermetricians fighting the major salary battles (on both sides) and Earl Weaver achieving semi-celebrity status as the unofficial manager of sabermetrics. Why, today it is even possible to enter a biker's bar, interrupt a baseball argument with a disparaging remark about the shortstop in question's range factor, and emerge alive--sometimes. But even if we can now contemplate a day in the not too distant future when Sparky Anderson will embrace Bill James, sabermetricians are unlikely to suffer Alexander's fate, as there is every reason to believe that another baseball world will remain to be conquered; a world, moreover, that is likely to resist the invading sabermetric hordes with all the tenacity of the Spartans at Thermopylae. (Although, as Rhett Butler once sardonically pointed out, the aforementioned heroic defenders died to the last man.) The world I'm referring to is that of baseball scouting, and it is the subject of a wonderful book by Kevin Kerrane entitled Dollar Sign on the Muscle (Beaufort Books, $15.95).

As far as I know this is the first book ever written about scouting, and frankly I cannot imagine a better job.

There's no doubt that Mr. Kerrane knew his way around a diamond before he began his research, but he must have spent innumerable hours talking to scouts and winning their confidence before he brought out the tape recorder and gave posterity some of the finest baseball monologues since The Glory of Their Times, and in a way superior because raunchier; as much as I admire Mr. Ritter's book, didn't you occasionally get the feeling that his interviews were sanitized? Didn't it bother you that at least one salty, grizzled verteran never burst out with, "Ty Cobb was a sonofabitching bastard, and I don't care who knows it?" (Peripherally, did you ever fantasize that, given the right subject and a reliable Sony, you could transcribe a first-rate book? Maybe you can, but I'm inclined to believe that getting people to speak eloquently into a machine is a talent no more common that the ability to create memorable characters out of pure language that distinguishes our finest novelists.)

If Dollar Sign on the Muscle consisted only of the oral history of scouting, it would still be a valuable book. But happily, Mr. Kerrane is as much skeptic as oral historian. He obviously admires the veteran scouts he interviews, yet he does not accept their complaints about the Major League Scouting Bureau (all teams are now compelled to belong to this combine), their nostalgia for the good old days before the Amateur Draft, or their almost unanimous contempt for anything resembling scientific measurement of a prospect's potential (some are still disdainful of even the stopwatch, not to mention the JUGS gun) at face value. Indeed, he contrasts these old guard, fundamentallist views with the modernist, quasi-sabermetrical theories of former Orioles and Reds Scouting Director Jim McLaughlin, who argues that scouting can be objectivized far beyond what the average fan--and especially the average scout--would consider possible or desirable. In fact, this book is on one level a debate between the front line scouts who, if they can't exactly define what makes a standout baseball prospect, know, like a Supreme Court Justice scouting pornography, one when they see him, and men like McLaughlin and Dave Ritterpusch (another ex-Scouting Director) who believe that a player's potential can be, if not precisely defined, at least to a certain extent measured. Mr. Kerrane takes no side in the debate, but it doesn't take a genius to conclude that, on the scouting level, sabermetrical types are getting knocked out of the box (McLaughlin and Ritterpusch are not ex-scouting Directors by choice). And after all, you can't categorically state that the intuition and "good face" (a standout ballplayer will look like a standout ballplayer, so runs the conventional wisdom) system doesn't work; I mean Mickey Mantle and Al Kaline and George Brett did make the majors, but A.) could any system be so faulty as to obscure the mighty talents of superstars? and B.) what about journeymen ballplayers like Jamie Quirk, who make the majors and hang on--sometimes--by the skin of their teeth? How many Jamie Quirks are still toiling on the sandlots because they didn't fire someone's intuition, or lacked the "good face?"

Something else struck me. In their monologues, all the scouts naturally like to talk about the major leaguers they discovered. Leon Hamilton is justifiably proud of finding, among others, Doyle Alexander and Bobby Richardson. But he passed on Henry Aaron. Jocko Collins signed Del Ennis and Granny Hamner for the Phillies, but gave the back of his hand to Bobby Shantz. And Superscout Howie Haak, fabled discoverer of Roberto Clemente, once saw Eddie Murray and Mitchell Page work out on the same field, and was impressed enough to draft in the third round...Mitchell Page. Okay, we all make mistakes, and when a scout signs a prospect, he is essentially placing a bet on how that player will perform four years from now. That's asking a lot. I'm glad I don't do it for a living. But the fact remains: 92 per cent of the ballplayers who sign professional contracts never have even a cup of coffee in the bigs. Is that an unacceptable rate of failure? Would the universal application of the series of tests and measurements--both physical and mental--proposed by Jim McLaughlin cut into that ungodly percentage of non-success? We'll never know until sabermetricians get a toehold in scouting.

But I should apologize for dwelling at such length on the sabermetrical angle, because this book should have tremendous appeal to anyone who cares about baseball. As I noted, the monolgues are splendid, and an early chapter neatly recapitulates the history of scouting from earliest times to the present. And--I should've mentioned this earlier--as the book progresses Mr. Kerrane zeroes in particularly on the Phillies scouting system, and follows it from earliest spring training evaluations right through to the June 1981 Amateur Draft, apparently sitting in on some high level pre-draft skull sessions in the process. He also records the turmoil that rocked the organization four months later when Dallas Green moved to the Cubs and, on his way out the door, lashed out at Phillies scouting. (If you read the book, I think you'll conclude he had a point.)

Most of all, though--and I haven't emphasized it nearly enough--this is a book about a way of life. Acutally, two ways of life, becuase scouting prior to the Amateur Draft was different. If you found a prospect, and could get his name on a contract, he was yours. Think how Indians' scout Cy Slapnicka must have felt when he became the first baseball man to watch Bob Feller throw. Nowadays, with the draft, plus the Major League Scouting Bureau scouring even the boonies, where is the thrill in discovery? Who first spotted Dwight Gooden's blinding speed? What difference does it make? He couldn't do anything about it. Somehow that strikes me as wrong. If the Mariners can't find their own talent, don't they deserve the second division? (That's where they'll end up anyway.) Or, who

would you rather have as the last bastion of laissez faire commerce, the major league scouts or the drug dealers? Not that you have any choice.

So the modern scouts must perservere with the knowledge that they'll never get that chance at the legendary "big score." What they will get is bone tired from the endless travel and months away from their families. In exchange they'll receive mediocre pay, precious little job security, and constant second guessing from the top. Yes, I can hear the cynics saying, "just like the average American worker," and they're right; I don't mean to romanticize the scouts; certainly Mr. Kerrane doesn't. There's plenty of reason to believe that they're particulary resistant to change even by baseball standards, that they wear prejudices conspicuously on their sleeves (amazing in the Dick Allen-Reggie Jackson era how many scouts will not go near a ballplayer who wears glasses), and will accept any system that attempts to objectively analyze talent when ballgames are being snowed out at the Helldome. But still they're out there on the front lines, and when they do their job right, it's somebody else who gets rich. I guess what I'm saying is that maybe it's possible to admire the tenacity of Robert E. Lee and his Confederates without actually wishing that they had won the war.

A note to fellow sabermetricians:

It is evident to me, from reading the introduction and other parts of the book, that the motive behind the writing of THE 1985 ELIAS BASEBALL ANALYST was not solely to advance the knowledge and understanding of baseball. If you read between the lines, it says: 'Well, it's high time we put you imitators in your place. Think your pretty good, do ya?, with your sabermetrics and Project Scoresheets and all. Well, take this! ... And that's not all. We were just flexing our muscles. What you know, we could fit in our pinky...

OK, I got a little carried away, I know. However, try to look beyond all of that. What this book is is one of the most valuable and vast sources of information that I have ever seen. Not only because the information is there, but because of the potential it has. Do power hitters do poorly when leading off an inning? Do power pitchers do well at night? What types of players perform best during the warm weather months?, on artificial turf? What types of players have the greatest left/right differentials? These are the types of questions we've been asking for years, that can now be easily answered. Whatsmore, Elias has done all of the dirty work for you! I sat down and, in less than an hour, wrote a two page list of studies that could be done using only the information found in this book. I'm sure I haven't exausted all of the possibilities. So let's get to it!

We, as sabermetricians, are involved in a discussion of baseball. The ELIAS BASEBALL ANALYST has not ended that discussion, but instead has given us the oportunity to discuss at a higher level, where we can talk about about things that took too many words to talk about before.

Sincerely,

Scott Segrin