By the Numbers

Volume 8, Number 1 The Newsletter of the SABR Statistical Analysis Committee October, 1998

RSVP

Phil Birnbaum, Editor

Well, here we go again: if you want to continue to receive *By the Numbers*, you'll have to drop me a line to let me know. We've asked this before, and I apologize if you're getting tired of it, but there's a good reason for it: our committee budget.

Our budget is $500 per year.

Our current committee member list numbers about 200. Of the 200 of us, 50 have agreed to accept delivery of this newsletter by e-mail. That leaves 150 readers who need physical copies of BTN. At four issues a year, that's 600 mailings, and there's no way to do 600 photocopyings and mailings for $500.

In the worst case, we'll start asking for a little bit of extra money. But we might not need to. If, for instance, 50 more of us switch to an e-mail subscription, and 25 are no longer interested in receiving BTN, that leaves only 75 paper subscribers, which just fits into our $500 budget.

And so: apologies again, but if you want to keep on receiving BTN – and I know that's most of you – please let me know. I know it's inconvenient, and I know you've been asked this before, and I know you're probably frustrated that BTN isn't published as much as you'd like, but those potential 25 dropouts and 50 e-mail recipients are critical to our budget, and I just need a quick note or e-mail for you to say you're not one of them. All I need is your name and a "yes."

If you already replied to my September e-mail, you don't need to reply again. If you didn't receive my September e-mail, it means that the committee has no e-mail address for you. If you do have an e-mail address but we don't know about it, please let Neal Traven (our committee chair – see his remarks later this issue) know, so that we can communicate with you more easily. Giving us your e-mail address does *not* register you to receive BTN by e-mail. Unless you explicitly request that, I'll continue to send BTN by regular mail.

As our 1998 budget has not been touched until now, we have sufficient funds left over for one more full issue this year. But whether that issue gets published depends on whether we get enough article submissions. I have corresponded with a few of you who have told me that you're working on preparing submissions for BTN; I hope you'll have a chance to get them to me by December, so we can get another BTN out the door in '98.

Thanks for your patience. Welcome to BTN, and I hope you enjoy this issue.

You can e-mail me at birnbaum@magi.com. Or, you can write me at #608-18 Deerfield Dr., Nepean, Ontario, Canada, K2G 4L1. ♦

In this issue

Editor's Comments

Phil Birnbaum

Sabermetrics doesn't have much of a public profile. You won't find "Sabermetric Weekly" next to *The Sporting News* on your local newsstand, or "This Week in Baseball Statistics" on ESPN. You won't even find a sabermetric column in Baseball Weekly, a publication devoted to baseball fanatics, and I'd bet that the phrase "linear weights" has never appeared in your local newspaper.

Nobody cares about sabermetrics–except us. And that's why I think it's so important that this newsletter be active in the field.

Heard any interesting developments in baseball analysis lately? I haven't, and I'd bet most of you haven't either. And that's a problem.

Suppose you're a reporter, or a baseball researcher, or just a baseball fan who stumbled across a Bill James or Pete Palmer book in your local library. Your interest piqued, you want to find out more. What's the state of the science? What's new in sabermetrics? The field must have advanced quite a bit since Craig Wright's book came out, mustn't it? What does the world know about baseball that it didn't know back in 1988, or 1991, or even last month?

Where would you go to find out? To us, of course. I mean, if you wanted to know the state of sabermetrics, wouldn't your first stop be the Society for American Baseball Research? Wouldn't you think that perhaps the *Statistical Analysis Committee* for the Society for American Baseball Research would be active in the field? If anyone should be keeping tabs on what's happening in baseball analysis, it should be us. If you call the American Medical Association, and they have a Cancer Research Chapter, wouldn't you expect them to know something, as an organization, about what's been happening in cancer research?

But if someone came to be, as a member of our committee, and asked me what's new in the field, I wouldn't be able to answer.

There's a lot happening in sabermetric research, I'm sure. I've seen some very nice research-oriented websites; there are baseball annuals, many of them home-published, that come out every year, and, less formally, there are informative postings to internet newsgroups and our very own SABR-L.

Let's start keeping track of this stuff.

When a new movie comes out, you expect to find a review in Entertainment Weekly. When a new sports car is built, you expect an article in Car and Driver. And if there's new sabermetric work happening, this committee should expect it to hear about it in By the Numbers.

And so, a request: if you see new research, or a new book, or an interesting internet posting, and if you learned something from it, let us know. Even better: summarize it for us, or write a review, or a letter of criticism. But at the very least, drop me a line to let us know it's out there.

And if it's you doing the research, by all means, send it along. Ever since the demise of the Baseball Analyst in the late 80s, sabermetrics has had no journal, no place for us to take our research to share with the sabermetric community. Though officially termed a "newsletter," I'd like to see BTN take over as the Journal of sabermetrics. We've tried serving that purpose in fits and starts, but the constant committee turnover and inconsistent publication dates have hurt us immensely. Speaking personally, I have research papers I wrote ten years ago looking for a place to publish, and one paper I sent to BTN several years ago is long lost. I have done close to zero over the past while–why spend the time if nobody seems to be interested?–and I can't be the only one in that situation. I hope we start getting some of our members' research in here.

To that end, I promise a deal: if you do the research, I promise that BTN will be a reliable place to get it reviewed and published.

It takes more than me and my computer, of course–it takes articles, and those are by no means guaranteed. Back in the last days of the Baseball Analyst, publication was spotty because of a lack of submissions, and it would be foolishly optimistic to think that more members would want to send articles to me than to Bill James. But you never know. There might possibly be a five-year backlog of frustrated members out there, and I hope you'll consider BTN as a place to submit your research.

Phil Birnbaum, 18 Deerfield Dr. #608, Nepean, Ontario, Canada, K2G 4L1; birnbaum@magi.com ♦

Committee News

Neal Traven, Committee Chair

Welcome back to By The Numbers. It's been far too long since the last issue of our newsletter, an absence that I take full blame for. To address that deficiency, and to get committee news and views out to you on a more regular basis, I've decided to accept Phil Birnbaum's offer to become the editor of BTN. As Phil put it himself, this way I'll have someone to fire if the newsletter doesn't get published!

From now on, short articles and similar materials that you're offering to the newsletter should go to Phil, though I'll also happily accept and review them before passing things along to him. Since I'm still the holder of the "official" mailing list of the committee, please send changes of address and the like to me; I'll give Phil the updated list of newsletter addressees before each issue goes out. Of course, Phil and I will be in regular communication about these matters.

For my own part, I'll contribute a column or discussion points to every issue of the newsletter. I'll also continue to be the point-of-contact between SABR in general and the Statistical Analysis committee, which means that I'll pass along Society news of importance to members, report on what transpires at the annual national meeting, etc. In addition, if a committee member wishes to raise a committee-related issue with the SABR Board, officers, or administrative staff, I can act as conduit in that direction as well.

Committee adminstrivia–Phil will undoubtedly give you his snail address, email address, phone number, and whatever additional contact information he can offer. In the meantime, please note the following changes in contact information for both of the committee co-chairs:

Clem Comly now has an email address–ccomly@erols.com.

All of my contact information is now different from what's in the SABR Directory, and from what's in the list of committee chairs in the September-October 1998 SABR Bulletin. They do have my correct info in their mailing list and on SABR-L, but not in the place where it probably matters most Please update your records from the contact information at the end of this article.

Convention report

As was reported in the September-October 1998 SABR Bulletin, this year's national convention drew 437 registrants, making it the largest West Coast convention in SABR history. By my estimate, 50 or more were members of the SAC. Of the 39 research presentations during the meeting, 11 (28.2%) were by members of the committee–Tony Blengino (two talks), Clem Comly, Bill Gilbert, Michael Hoban, Herman Krabbenhoft, Steve Krevisky, Mark Pankin, Doug Pappas, Dave Smith, and Jim Vail. Special plaudits to Pappas ("Thirty Years of Collective Bargaining") and Smith ("The Effectiveness of Platooning"), finalists for the Baseball Weekly award given to honor the best research presentation at the convention. For the second year in a row, Mark Kanter was on the winning trivia team.

In contrast to previous conventions, in which the SAC meeting was invariably scheduled for the crack of dawn, we met at the very reasonable hour of 10 AM this year. Unfortunately, it came on Sunday, so several easterners had already left on their losing-time-zones flights home. As usual, the meeting consisted largely of descriptions of areas of research being pursued by the attendees, punctuated by impromptu digressions into the subject matter itself. Several of the research presentations given during the convention also became brief discussion topics during the committee meeting.

Perhaps the most valuable item at the SAC meeting came from SABR Publications Director Mark Alvarez. Mark strongly encourages and welcomes committee members to submit research papers with a statistical orientation. If your subject matter is too extensive for submission to By The Numbers, I encourage you to submit it to Mark Alvarez for the Baseball Research Journal. He really wants to increase the statistical and analytic content of BRJ. Mark recognizes the disconnect that built up over the years between BRJ and statistical analysis–as fewer articles were published, fewer were submitted, so fewer were published, so fewer were submitted, ...

Alvarez would be the first to admit that he has neither expertise nor great interest in statistical analysis. In sharp contrast to past Publication Directors, however, he is reaching beyond his personal viewpoint, and turning to the SAC for assistance in evaluating statistical submissions. Thus, he has asked me to supervise the "vetting" or critical review of statistically-oriented papers, to assure that those which make it into broad-based SABR publications are valuable to the analytic community within the Society, and (more importantly) scientifically well-founded and well-reasoned. At the committee meeting, several long-time "distinguished" members expressed willingness to participate in this peer-review process; they will be called on when needed.

Neal Traven, 500 Market St. #11L, Portsmouth, NH, 03801; 603-430-8411; baseball@ttlc.net ♦

A Flaw in the "Big Bad Baseball Annual"

Clem Comly

It is paradoxical but true that a better hitter can have a lower "offensive wins above replacement" than a lesser hitter. The author explains the two reasons why, and, from that, why a criticism of the statistic "runs created per game" is misguided.

The 1998 Big Bad Baseball Annual contained G. Jay Walker's article "Runs Created per Game: the Beauty with a Blemish." That article's conclusion is not correct.

Mr. Walker gave this example of three player lines:

	PA	AB	H	2B	3B	HR	W	H+W	Out	BA	SP	OB%	RC/G	OffW%	G	OWAR
Iteration+1	600	498	198	33	11	54	102	300	300	.398	.833	.500	18.6	.941	11.1	6.57
Start	600	500	180	30	10	50	100	280	320	.360	.760	.467	15.0	.912	11.9	6.66
Iteration-1	600	502	162	27	9	46	98	260	340	.323	.687	.433	11.9	.866	12.6	6.50
Iteration =	0	-2	+18	+3	+1	+4	+2	+20	-20	= turn 20 outs into hits and walks						

- OffW% is Offensive Winning Percentage (derived from RC/G per Bill James)
- G is offensive games (outs/27)
- OWAR is Off. Wins above replacement = (OffW% -.35)*G.

He then went on to say that it makes no sense for the OWAR to be worse for "Iteration +1", which is obviously better than Start. Therefore, he concludes, the formulae are flawed.

Walker made two mistakes. One is the missing games between start and "Iteration+1" don't disappear: they are used by other players in the lineup. Assuming the rest of the team plays .500 ball:

		Rest of Team Plays at .500						
		Player		Rest of Team			Team Totals	
	OffW%	Games	OWAR	OffW%	Games	OWAR	Games	OWAR
Iteration+1	.941	11.1	6.57	500	150.9	22.64	162	29.21
Start	.912	11.9	6.66	.500	150.1	22.51	162	29.17
Iteration-1	.866	12.6	6.50	.500	149.4	22.41	162	28.91

Allowing for the lost games going to the rest of the team shows that that "Iteration +1" is indeed better than "Start".

The second problem is combining WL% by weighting per game. This does not reflect not normal baseball conditions. It is equivalent to either:

- having a great starting 8 get injured (or worse) in an April plane crash with an 11-1 record and have them replaced with .500 level bench players, or

- the rules of baseball change so "Start" takes every at bat in the first game for his team, then (average) player 2 bats all of the next game, then (average) player 3 bats all of the next game, etc. In the course of the season, "Start" bats in 12 games.

So what you have is 12 games where the average score is 15-5 and 150 games where the score is 5-5. The right way to look at the situation is to go back to RC/G (that Mr. Walker derived OffW% from):

	Rest of Team Plays at .500								
	Player			Rest of Team			Team Totals		
	RC/G	Games	Runs	RC/G	Games	Runs	Games	Runs	Wins
Iteration+1	18.6	11.1	206	4.66	150.9	703	162	909	95.9
Start	15.0	11.9	179	4.66	150.1	699	162	878	93.1
Iteration-1	11.9	12.6	150	4.66	149.4	696	162	846	90.2
Average	4.66						162	755	81.0
Replacement(.350)	3.42						162	554	56.7

As you can see, "Iteration+1" is now close to 3 wins more valuable than "Start", and "Iteration-1" is now close to 3 wins less valuable than "Start."

Clem Comly, 308 Colonial Drive, Wallingford, PA, 19086-6004; ccomly@erols.com ♦

Receive BTN by E-mail

You can help save SABR some money, and me some time, by receiving your copy of *By the Numbers* by e-mail. BTN is sent in Microsoft Word 97 format; if you don't have Word 97, a free viewer is available at the Microsoft web site (www.microsoft.com).

To get on the electronic subscription list, send me (Phil Birnbaum) an e-mail at birnbaum@magi.com. If you're not sure if you can read Word 97 format, just let me know and I'll send you this issue so you can try

If you don't have e-mail, don't worry–you will always be entitled to receive BTN by mail, as usual. The electronic copy is sent out two business days after the hard copy, to help ensure everyone receives it at about the same time.

New Statistic

A Leadoff-Hitter Rating System

Bill Felber

Perhaps alone among the nine batting-order slots, there are specific, measurable characteristics we look for in a leadoff hitter. The author enumerates those skills, develops a rating system to evaluate leadoff hitters, and lists the 1997 number-one hitters and the fifty best of all time.

One

Attached is a rating of leadoff hitters, one per team, for 1997. I developed an interest in rating leadoff hitters out of my judgment that–perhaps alone among batting order positions–it could be done. That is, that there are specific and measurable qualities we value in a leadoff hitter. My deduction was that the qualities essentially are three-fold:

1. Ability to reach base, which can be measured by on-base average;

2. Ability to advance unaided into scoring position, which can be measured by the player's ratio of extra base hits and stolen bases per plate appearance, reduced, when available, by times caught stealing (I multiply this by two because of its impact), and the result compared to the league average;

3. Ability to score runs per plate appearance relative to the league average per plate appearance, measured by runs scored.

Beyond that, I concluded that these qualities were not equally important. Ability to reach base is the most important, ability to score runs–since it is situation-dependent–is the least important. Therefore, in devising a formula, I weighted it to give priority to on base average, and to lessen the impact of runs scored. I believe any player can be given a rating, based on how well his abilities in these areas compare with the league norms, with 100 being major league average. Since I weight the three areas in terms of importance, I multiply a player's score in on base percentage by 1.2, multiply a player's score in runs scored by .8, and do not multiply a player's score in advancing unaided to second. This produces a score for an average major league of 300. The precise formula is expressed in three parts as follows.

1. (Player's OBA as a percentage of league average OBA) X 1.2 = Rating.

2. ((Player's extra base hits + (steals - (caught stealing X 2)))/plate appearances) as a percentage of ((League average XBH + (steals - (caught stealing X 2))))/plate appearances = Rating.

3. (Player's runs scored/plate appearances) as a percentage of (league runs scored/plate appearances.)

You will quickly discern that much of this formula is debatable. Why weight runs scored at .8 and why rate on base average at 1.2? Plainly that is an editorial judgment on my part as to the relative importace of the skills being measured. Beyond that, in the formula for advancing to scoring position unaided, should a double and a home run be given equal weight? Should a caught stealing be a double deduction? Should some formula be devised to account for the fact that steals move a runner into scoring position? Finally and perhaps most pertinently, is the formula simply too complicated? I think those are all sensible questions. This reasearch is basically intended to pass this to you in the hope that it gets some discussion going.

1997 NL Leadoff Hitters

	OB	To 2nd	Runs	Score
Biggio	149.55	169.26	134.63	453.44
DeShields	128.65	157.47	95.43	381.55
Womack	117.48	161.00	80.83	359.31
Young	129.37	120.55	99.48	349.40
Grudz	110.63	128.09	73.56	312.28
Everett	110.99	117.26	79.42	307.67
White	121.80	98.00	81.03	300.83
Weiss	135.86	91.01	73.68	300.55
Johnson	133.33	70.01	86.34	289.68
Hamilton	127.50	63.17	97.37	288.14
Cummings	118.92	102.73	65.98	287.63
Veras	128.65	74.58	78.77	282.00
Lofton	147.39	34.68	89.10	271.17
Reese	102.34	88.73	72.94	264.01
NL LEADOFF AVERAGE				317.12

1997 AL Leadoff Hitters

	OB	To 2nd	Runs	Score
Knoblauch	137.65	152.66	103.60	393.91
Garciaparra	120.71	150.95	104.42	376.08
Stewart	129.88	156.51	82.27	368.66
Henderson	139.06	124.39	103.38	366.83
Raines	142.24	93.81	110.45	346.50
Hunter	117.88	131.38	95.20	344.46
Anderson	138.71	104.94	88.56	332.21
Cora	126.71	91.41	103.05	321.17
Roberts	121.76	100.96	84.46	307.18
McDonal	127.41	71.74	106.33	305.48
Goodwin	110.82	102.61	89.62	303.05
Durham	118.94	77.21	93.86	290.01
Damon	119.29	56.94	83.81	260.04
Vina	110.12	43.55	67.77	221.44
AL LEADOFF AVERAGE				324.07

Two

If the formula seems complicated, perhaps illustrating it by example might de-mystify it a bit. By my calculations, the best leadoff man in 1997 was Craig Biggio. Here's how his rating is calculated:

```
1.      (Biggio's on base average) .415
        (League on base average)   .333
                                    Score 124.62    (.415 is 124.62% of .333)
                                  X 1.2 = 149.55

2. A.   (Biggio's extra base hits) = 67
        (Biggio's steals (47) minus twice Biggio's caught stealing (10 X 2 = 20) = 27
        Subtotal = 94.
        (Biggio's plate appearances) = 703
        Biggio's ratio of reaching second unaided = .134

2. B    (League's extra base hits) = 6,555
        (League's steals (1,817) - caught stealing (841) X 2 = 1,682) = 135.
        Subtotal = 6,690
        (league's plate appearances) = 84,907
        (League's ratio of reaching 2nd unaided) = .079

2. C.   Biggio ratio (.134) as a percentage of league ratio (.078) = 169.26
                                                                    Score 169.26
```

```
3. A.    (Biggio's runs scored) = 146
                (Biggio's plate appearances) = 703
                (Biggio's ratio of runs to plate appearances) = .207

     B   (League runs scored) = 10,440
                (League plate appearances) = 84,907
                (League ratio of runs scored to plate appearances) = .123

     C.  Biggio's ratio of runs scored as a percentage of league ratio) = 168.29
                                                                          168.29 X .8 = 134.63
                                                                               Score = 134.63
Biggio's total score:     149.55
                                  169.26
                                  134.63
                                  453.44
```

Three

Because this formula is normalized over time via inclusion of league average figures, it can also be used to measure players from different eras, and in this way develop an all-time ranking for leadoff hitters. There are two inherent problems with this, however. The first is that caught stealing statistics are not available for players prior to the second or third decade of this century. To deal with this, in making all-time calculations, I drop the caught stealing portion of the formula because I have no better solution. Second, many players–notably guys like Bobby Bonds and Tim Raines–are leadoff hitters for parts of their careers, and yet hit somewhere else at other times. That being the case, I utilize only a peak rating system that measures a player's three best concurrent seasons, and I include players who it would make sense to me to have been used as leadoff hitters (Ty Cobb, Wade Boggs) irrespective of whether they actually batted leadoff the preponderance of that period. I have also excluded players who do very well on this system, but who would make illogical leadoff candidates (Babe Ruth being the prototype). Again, these are debatable judgments. Based on this, I have also included a list of the top 50 leadoff hitters of all time with their score (average per year for the three-year period indicated).

Top 50 Leadoff Hitters of All Time

	Name	Years	OB	To 2nd	Runs	Total
1	Henderson	81-83	149.54	264.37	120.69	534.60
2.	Raines	85-87	154.66	196.81	128.80	480.27
3	Brock	72-74	136.50	234.67	106.17	477.34
4	Hamiton	94-96	165.25	179.82	124.49	469.56
5	Bonds	71-73	132.53	205.41	128.60	466.54
6	Lopes	74-76	128.52	217.95	113.02	459.49
7	Coleman	85-87	123.36	218.09	112.28	453.73
8	Lofton	93-95	137.98	191.84	121.79	451.61
9	McGraw	98-00	177.82	142.05	131.24	451.11
10	Flick	04-06	151.86	183.75	112.16	447.77
11	Stirnweiss	44-46	137.03	191.18	116.63	444.84
12	Wilson	79-81	125.47	205	111.6	442.07
13	Campaneris	68-70	131.66	195.95	109.13	436.74
14	Carew	73-75	156.92	161.54	114.91	433.37
15	Martin	33-35	131.48	174.32	124.74	430.54
16	Knoblach	94-96	143.3	173.68	113.33	430.31
17	Burns	17-19	144.54	167.95	114.29	426.78
18	Dykstra	92-94	151.6	151.65	118.62	421.87
19	Case	41-43	125.42	182.19	110.44	418.05
20	LeFlore	76-78	134.77	173.25	109.91	417.93
21	Beaumont	01-03	147.81	132.05	137.35	417.21
22	Moreno	78-80	121.5	188.64	102.54	412.68
23	Milan	11-13	138.03	170.97	102.67	411.67
24	Carey	21-23	124.94	177.11	108.75	410.8
25	Wills	61-63	132.11	164.94	111.3	408.35

	Name	Years	OB	To 2nd	Runs	Total
26	Nixon	91-93	135.88	157.3	114.69	407.87
27	Alomar	92-94	143.35	158.06	106.45	407.86
28	Ryan	88-90	145.96	154.13	106.85	406.94
29	Buford	69-71	151.32	128.37	121.83	401.52
30	Garr	72-74	131.63	157.14	111.49	400.26
31	Biggio	95-97	139.76	138	121.94	399.7
32	Rivers	74-76	123.72	180	95.14	398.86
33	Bush	09-11	140.2	129.07	129.46	398.73
34	Bescher	11-13	137.93	151.06	106.45	395.44
35	Agee	69-71	129.74	160.53	104.29	394.56
36	Hack	40-42	151.03	130	108.82	389.85
37	Shotton	13-15	145.28	127.91	105.93	379.12
38	Devore	10-12	134.59	143.01	108.66	386.26
39	Anderson	94-96	126.26	157.89	101.82	385.97
40	Werber	34-36	105.95	160.24	96.47	362.66
41	McAuliffe	66-68	143.28	121.92	118.02	383.22
42	Rice	19-21	138.88	159.04	83.31	381.23
43	Holmes	45-47	138.91	139.44	102.96	381.31
44	Molitor	92-94	144.43	135.48	100	379.91
45	Aparicio	59-61	117.06	168.42	94.24	379.72
46	Hooper	19-20	140.66	139.24	123.48	403.38
47	Combs	27-29	150.1	117.05	110.54	377.69
48	Reese	46-48	138.74	125.71	113.82	378.27
49	Boggs	86-88	170.48	97.85	108.62	376.95
50	Dark	51-53	125.5	134.21	116.49	376.2

If you think this is worth feedback, I would be interested in receiving it.

Bill Felber, 3110 Gary, Manhattan, KS, 66502; 785-537-1302; felber@flinthills.com ♦

Runs, Earned Runs, and Adjusted Runs

Doug Pappas

Though laudably attempting to distinguish runs that are the pitcher's reponsibility from those that more properly should be attributed to defense, ERA contains its own biases. Treating unearned runs as earned tends to overcompensate. The author therefore suggests a new statistic, Adjusted Runs, which holds the pitcher responsible for half the unearned runs scoring under his watch.

Many baseball analysts consider Earned Run Average the best "traditional" statistic. Unlike offensive statistics, ERA directly measures the pitcher's most fundamental duty: to prevent the opposition from scoring.

However, Earned Runs contain their own bias. In a laudable effort to separate the pitcher's responsibility from that of his defense, a pitcher is not charged with runs that would have been prevented by errorless fielding.

Unfortunately, Earned Runs' all-or-nothing approach also exonerates pitchers who follow a two-out error with six straight hits, and helps hide the mistakes of those "helped" by butterfingered teammates or a strict official scorer.
Noting these problems, some recommend abolishing the distinction between earned and unearned runs, ranking hurlers based on Run Average. A switch from ERA to RA would affect almost one-third of all pitching crowns since 1901–28 in the American League and 35 in the National–and cost Christy Mathewson four of his five titles. (See Table 2 below.)

Some single-season changes merit further comment. In 1907, Cub hurler Jack Pfiester led the league with a 1.15 ERA...but with 36 of his 61 runs unearned, his RA bloats to 2.82, worst on a rotation whose other four members finished 1-4 in the NL in RA that year! Four years later, Nap Rucker finished 12th in the ERA race, 0.75 behind leader Christy Mathewson, yet led the league in RA, 0.08 ahead of Big Six. In 1944, luckless Card hurler Ted Wilks led the league in RA at 2.64...but every run against him was earned, causing him to finish fourth in the ERA race. And Greg Maddux's 1995 ERA of 1.63 looks even more impressive since he allowed only one unearned run all year–a RA of 1.67, down from 1.93 in 1994. Can Maddux possibly keep improving?

When comparing pitchers, variances in unearned run percentage can be as significant as park effects. Extreme strikeout and fly-ball pitchers yield only half as many unearned runs as their ground-balling colleagues, with the modern extremes represented by fly-ball/HR hurler Jim Deshaies at 5.6% and sinkerballer Randy Jones at 16%. Not surprisingly, knuckleballers also surrender more than their share of unearned runs: 18.2% of the runs scored against Hoyt Wilhelm were unearned, with Wilbur Wood at 14.6% and Phil Niekro 13.9%.

The effect on cross-era comparisons is even more pronounced because the percentage of unearned runs has been declining throughout baseball history, as seen in Table 1.

The harsh light of RA does not flatter dead-ball pitchers. Rube Waddell's career 2.16 ERA was 0.70 below Tom Seaver's, yet Seaver posted a better RA, 3.15 to 3.23. Sandy Koufax's RA is lower than Eddie Plank's, Jim Palmer's below Grover Cleveland Alexander's. Pitchers whose careers began in the 19th century fare even worse–Cy Young's 2.63 ERA inflates to a RA of 3.87, higher than Fritz Peterson or Jim Perry.

So what should an analyst do? Ignoring unearned runs gives the pitcher an undeserved break, while ignoring the defensive lapses blames the pitcher for the sins of his defense. The fairest result lies somewhere between–accordingly, I propose holding pitchers 50% responsible for the unearned runs against them through a new measurement, Adjusted Run Average (ARA).

ARA is easy to compute: (Earned Runs + ½ Unearned Runs) x 9 / IP. Usually the seasonal ARA champ will also win the ERA or RA title, but in four seasons the ARA crown goes to a pitcher who led the leader in neither conventional category. Dead-ballers dominate the list of 25 lowest career ARAs (Table 3), but Hoyt Wilhelm moves far up the list and Whitey Ford and Sandy Koufax crack the top 25. Rube Waddell, whose personality must have infected his fielders, falls from fourth in ERA to eighth in ARA, while his near-contemporary Nap Rucker climbs from 15th to 10th and fireballer Walter Johnson passes Waddell and Christy Mathewson for fourth on the all-time ARA list.

Table 1: Unearned Run Percentage by Decade

1876-80	53.9%
1881-90	40.7%
1891-1900	32.2%
1901-10	29.8%
1911-20	24.2%
1921-30	16.2%
1931-40	14.2%
1941-50	13.4%
1951-60	11.4%
1961-70	11.9%
1971-80	11.3%
1981-90	10.4%
1991-94	9.3%

ARA corrects for the deficiencies of Earned Runs without overcompensating. It can be calculated retroactively from "box score" statistics, and yields a result which can be substituted for ERA in all existing analytical frameworks. Most important, I believe that ARA produces a more accurate measurement of pitching quality than either RA or ERA–and, therefore, that analysts should consider switching to ARA.

Table 2: American League Season Leaders since 1950

1950	Wynn	3.70(2)	3.20	3.45
	Houtteman	3.67	3.53(5)	3.60
1951	Rogovin	3.28	2.78	3.03
1952	Reynolds	2.58	2.07	2.32
1953	Lopat	2.93	2.43	2.68
1954	Garcia	2.95	2.64	2.80
1955	Pierce	2.18	1.97	2.08
1956	Ford	2.79	2.47	2.63
1957	Sturdivant	2.90(2)	2.54(2)	2.72
	Shantz	3.02(4)	2.45	2.73
	Sullivan	2.84	2.73(5)	2.78
1958	Ford	2.55	2.01	2.28
1959	Wilhelm	2.55	2.19	2.37
1960	Baumann	3.26	2.68	2.97
1961	Donovan	3.20(2)	2.40	2.80
	Stafford	3.00	2.68(2)	2.84
1962	Aguirre	2.79	2.21	2.50
1963	Peters	2.56	2.33	2.44
1964	Chance	1.81	1.65	1.73
1965	McDowell	2.64	2.18	2.42
1966	Peters	2.37	1.98	2.17
1967	Horlen	2.30	2.06	2.18
1968	Tiant	1.85	1.60	1.73
1969	Palmer	2.39	2.34(2)	2.36
	Bosman	2.75(2)	2.19	2.47
1970	Segui	3.00(2)	2.56	2.78
	Palmer	2.89	2.71(2)	2.80
1971	Blue	2.11	1.82	1.96
1972	Perry	2.07	1.92(2)	1.99
	Tiant	2.26(3)	1.91	2.09
1973	Palmer	2.61	2.40	2.51
1974	Perry	2.74	2.52	2.63
1975	Palmer	2.42	2.09	2.26
1976	Blue	2.72	2.36(2)	2.537
	Fidrych	2.74(2)	2.34	2.538
1977	Tanana	2.69	2.54	2.61
1978	Guidry	2.00	1.74	1.87
1979	Guidry	3.17	2.78	2.98
1980	Norris	2.79	2.54(2)	2.66
	May	2.88(2)	2.47	2.67
1981	McCatty	2.42	2.32	2.37
1982	Stanley	3.21	3.10(2)	3.15
	Sutcliffe	3.38(3)	2.96	3.17
1983	Boddicker	3.27	2.77	3.02
1984	Stieb	2.93	2.83	2.88
	Boddicker	3.27(3)	2.79	3.03
1985	Stieb	3.02(2)	2.48	2.75
	Saberhagen	3.02	2.87(3)	2.94
1986	Clemens	2.73	2.48	2.60
1987	Key	3.21(2)	2.76	2.98
	Clemens	3.20	2.97(2)	3.08
1988	Higuera	2.61	2.45(2)	2.53
	Anderson	3.11	2.45	2.78
1989	Saberhagen	2.54	2.16	2.35
1990	Clemens	2.33	1.93	2.13
1991	Clemens	3.08(2)	2.62	2.85
	Ryan	3.02	2.91(5)	2.97
1992	Appier	2.55	2.46	2.51
1993	Appier	2.79	2.56	2.68
1994	Ontiveros	3.04	2.65	2.85
1995	Johnson	2.73	2.48	2.61

National League Season Leaders since 1950

1950	Maglie	3.10	2.71	2.91
1951	Spahn	3.21(2)	2.98(3)	3.096
	Roe	3.17	3.03(6)	3.105
	Nichols	3.52(7)	2.88	3.20
1952	Hacker	2.72	2.58(2)	2.65
	Wilhelm	3.40(11)	2.43	2.92
1953	Spahn	2.54	2.10	2.32
1954	Antonelli	2.71	2.29	2.50
1955	Friend	3.60	2.83	3.22
1956	Spahn	2.95	2.79(2)	2.87
	Burdette	3.23(3)	2.71	2.97
1957	Podres	2.94	2.66	2.80
1958	Miller	2.97	2.47	2.72
1959	Law	3.08	2.98(5)	3.03
	Jones	3.29(3)	2.82	3.06
1960	Broglio	3.03	2.75(2)	2.89
	McCormick	3.09(2)	2.70	2.90
1961	Spahn	3.29	3.01	3.15
1962	Koufax	2.98	2.54	2.76
1963	Koufax	1.97	1.88	1.92
1964	Koufax	1.98	1.74	1.86
1965	Koufax	2.41(2)	2.04	2.22
	Marichal	2.38	2.14(2)	2.26
1966	Koufax	2.06	1.73	1.89
1967	Niekro	2.78(2)	1.87	2.33
	Short	2.49	2.40(3)	2.44
1968	Gibson	1.45	1.12	1.28
1969	Gibson	2.47	2.18(4)	2.29
	Marichal	2.70(5)	2.10	2.45
1970	Seaver	3.19(2)	2.81	3.00
	Walker	3.09	3.04(3)	3.06
1971	Seaver	1.92	1.76	1.84
1972	Carlton	2.18	1.98	2.08
1973	Seaver	2.30	2.08	2.19
1974	Capra	2.78(2)	2.28	2.53
	Niekro	2.71	2.38(2)	2.55
1975	Messersmith	2.57	2.29(2)	2.43

	Jones	2.97(4)	2.24	2.61
1976	Rau	2.77(2)	2.57(2)	2.668
	Seaver	2.76	2.59(3)	2.673
	Denny	3.09(4)	2.52	2.80
1977	Candelaria	2.49	2.34	2.42
1978	Rogers	2.63	2.47(2)	2.55
	Swan	2.70(2)	2.43	2.57
1979	Richard	3.02(2)	2.71	2.866
	Hume	2.98	2.76(2)	2.871
1980	Sutton	2.38	2.21	2.29
1981	Ryan	2.05	1.69	1.87
1982	Niekro	2.63	2.47(2)	2.55
	Rogers	2.73(2)	2.40	2.57
1983	Hammaker	2.98(2)	2.25	2.611
	Denny	2.86	2.37(2)	2.615
1984	Pena	3.03(2)	2.48	2.75

	Gooden	2.97	2.60(2)	2.79
1985	Gooden	1.66	1.53	1.59
1986	Scott	2.39	2.22	2.30
1987	Ryan	3.19	2.76	2.98
1988	Hershiser	2.46	2.26(3)	2.36
	Magrane	3.10(8)	2.18	2.64
1989	Hershiser	2.63	2.31(2)	2.47
	Garrelts	2.70(2)	2.28	2.49
1990	Darwin	2.32	2.21	2.27
1991	Martinez	2.84	2.39	2.61
1992	Swift	2.24	2.08	2.16
1993	Rijo	2.66	2.48(2)	2.57
	Maddux	2.87(2)	2.36	2.61
1994	Maddux	1.96	1.56	1.76
1995	Maddux	1.67	1.63	1.65

Table 3: Top 25 RA/ERA/ARA, min. 1500 IP during 1901-94

Player	Years	RA	Rank	ERA	Rank	ARA	Rank
E.Walsh	1904-17	2.66	1	1.82	1	2.24	1
A.Joss	1902-10	2.82	2	1.89	2	2.35	2
T.Brown	1903-16	2.96	4	2.06	3	2.51	3
W.Johnson	1907-27	2.89	3	2.16	6	2.53	4
C.Mathewson	1900-16	3.04	5	2.13	4	2.59	5
O.Overall	1905-13	3.05	7	2.23	7	2.64	6
E.Reulbach	1905-17	3.03	6	2.28	8	2.66	7
R.Waddell	1897-10	3.23	18	2.16	5	2.70	8
E.Plank	1901-17	3.14	11	2.35	10	2.75	9
N.Rucker	1907-16	3.10	9	2.42	15	2.76	10
E.Cicotte	1905-20	3.21	16	2.37	11	2.79	11
J.Tesreau	1912-18	3.17	14	2.43	16	2.80	12
H.Wilhelm	1952-72	3.09	8	2.52	24	2.81	13
J.Scott	1909-17	3.30		2.32	9	2.81	14
G.McQuillan	1907-18	3.29	23	2.38	12	2.84	15
E.Killian	1903-10	3.29	24	2.38	13	2.84	16
D.White	1901-13	3.30	25	2.39	14	2.85	17
C.Bender	1903-25	3.31		2.45	17	2.88	18
G.Alexander	1911-30	3.21	17	2.56	25	2.89	19
H.Wiltse	1904-15	3.35		2.47	21	2.91	20

All statistics from *The Bill James Baseball Encyclopedia For Windows*.

Doug Pappas, 100 E. Hartsdale Ave. #6E, Hartsdale, NY, 10530-3244, dougp001@aol.com. ♦

Speed and Opposition Errors

Dan Levitt

Does batter speed force opposition errors? The author investigates this question, as well as whether speedy teams tend to overshoot their runs created estimates.

In the 1997 Statistical Analysis Committee meeting some of the discussion revolved around the effect of speed as a factor in forcing opposition errors. I have completed some preliminary research in this area, although analyzed for teams rather than individuals.

My study compares team Speed Scores with opposition errors and opposition unearned runs to check if any correlation exists between team speed and forcing errors. Bill James' Speed Score is used as a proxy for team speed; the calculation of which is outlined below. Data for opposition errors and unearned runs is from *The Sporting News*.

For comparison purposes I also checked the correlation with the Runs - Runs Created differential. The theory being that a fast team might be more likely to outperform its runs created projection by causing additional errors, taking the extra base or other things that runs created does not count.

The analysis uses 1996 and 1997 data; I have not found data for opposition errors available from earlier publicly available sources. The Runs - Runs Created calculation could obviously be further looked at by using previous years.

The results can be seen below:

Correlation of Speed score with:	
Opposition Errors	.42
Unearned Runs	.26
(R-RC)%	.27

Based on taking statistic courses many years ago as an engineering major, I would suggest that the data indicates a positive but not overwhelming relationship between team speed and opposition errors.

Specifically, the top five fastest teams averaged 134 opposition errors, the top ten 127. The slowest five teams averaged 109, and the slowest ten teams 113 opposition errors.

Clearly only two years of data may not be enough to fully clarify the relationships between speed and opposition errors. Additionally, speed scores may not be the ideal proxy for team speed. For example, it may cause more opposition errors to have several very fast players and several very slow players than a team of all players with average speed--both of which may have similar team speed scores.

The top ten speed scores from 1996 and 1997:

Team	Year	SpS	OE	UR	(R-RC)%
Colorado	1996	6.42	140	72	0.011
Pittsburgh	1997	6.36	137	78	-0.040
Kansas City	1996	6.08	133	75	-0.007
Houston	1996	6.02	121	90	-0.006
Houston	1997	5.94	137	85	-0.023
Minnesota	1997	5.92	100	61	-0.009
Cincinnati	1997	5.88	120	74	-0.077
Pittsburgh	1996	5.86	135	81	0.006
Minnesota	1996	5.82	125	86	0.017
Cincinnati	1996	5.80	123	69	-0.002

The bottom ten speed scores from 1996 and 1997:

Team	Year	SpS	OE	UR	(R-RC)%
California	1996	4.06	112	58	-0.070
Boston	1997	4.37	110	66	-0.085
Oakland	1996	4.54	95	49	-0.022
Oakland	1997	4.62	93	55	-0.063
Atlanta	1996	4.72	135	75	-0.033
Detroit	1996	4.78	103	62	0.048
Florida	1996	4.81	115	54	-0.036
New York	1997	4.87	116	83	0.045
Atlanta	1997	4.89	132	83	-0.023
Cleveland	1997	4.91	120	67	-0.064

Where:

Calculation	**Description**	**Formula**
SpS1	Speed score based on SB%	((SB+3)/(SB+CS+7)-0.4)*20
SpS2	Speed score based on SB attempts	SQRT((SB+CS)/((H-2B-3B-HR)+BB+HP))/0.07
SpS3	Speed score based on triples	3B/(AB-HR-K)/0.02*10
SpS4	Speed score based on runs per time on base	((R-HR)/(H+BB-HR-HP)-0.1)/0.04
SpS5	Speed score based on GDP's	(0.055-GDP/(AB-HR-K))/0.005
RC	Runs created	(H+BB+HP-CS-GDP)* ((H+_2B+2*_3B+3*HR)+0.52*(SB+SH+SF)+0.26 *(BB-IBB+HP))/(AB+BB+HP+SH+SF)
SpS	Net Speed Score	Average of highest-scoring four of the five Speed Scores above
OE	Opponents Errors	
UR	Unearned runs scored	
(R-RC)%	Actual runs less runs created as a % of Runs	

Any feedback from committee members would be appreciated.

Dan Levitt, 4401 Morningside Road, Minneapolis, MN, 55416, danrl@ibm.net ♦

Offensive Replacement Levels

Clifford Blau

What is the required level of offensive performance required for a player to keep his full-time job? Is there a fixed "offensive replacement level" below which a player almost always loses his job? The author researches historical data to answer these questions.

Bill James, in his 1985 Baseball Abstract, suggested that if one studied the rate of return for regulars at different levels of offensive production, one could determine the amount of hitting needed for players to keep their jobs. This required performance Mr. James calls the sustenance level, and I call here the offensive replacement level. Determining the offensive replacement level at each position is important for several reasons. First, it is valuable to use in estimating the value of a player to his team, which is simply how much better he is than whomever his team could replace him with without giving up another player. Since regulars who produce at less than the offensive replacement level lose their jobs, it is obvious that teams believe that there are players readily available who perform at least at that level. Thus, the replacement level can be used for evaluating players in general, rather than considering what specific person would take over for a given player if he were replaced. Second, it is useful for comparing the relative importance of fielding at each position, at least as perceived by managers.

Theoretically, one could determine a player's defensive value by examining at what point he loses playing time due to weak hitting (see Eddie Brinkman for an example.)

The complete results of my study can be found in Table 2. Both the replacement level and average production for each position have the expected relationship, in that first basemen and left fielders have the highest replacement levels and averages, and shortstops and second basemen the lowest. For those not familiar with runs created per game, Table 1 lists some representative seasons showing conventional statistics along with the corresponding runs created per game.

RC/G	Name, Year	AB	H	2B	3B	HR	BB	SB	BA	SA
2.0	Larry Bowa, 1973	446	94	11	3	0	24	10	.211	.249
2.5	Craig Robinson,1974	452	104	4	6	0	30	11	.230	.265
3.0	Mark Belanger, 1970	459	100	6	5	1	52	13	.218	.259
3.5	Mike Andrews, 1972	505	111	18	0	7	70	2	.220	.297
4.0	Jerry Remy, 1978	583	162	24	6	2	40	30	.278	.350
4.5	Bill Melton, 1974	495	120	17	0	21	59	3	.242	.404
5.0	Brooks Robinson, 1970	608	168	31	4	18	53	1	.276	.429
5.5	Rico Petrocelli, 1970	583	152	31	3	29	67	1	.261	.473
6.0	Bobby Murcer, 1973	616	187	29	2	22	50	5	.304	.464
6.5	Paul Molitor, 1979	584	188	27	16	9	48	33	.322	.469
7.0	George Brett, 1977	564	176	32	13	22	55	14	.312	.532

Table 1: Typical batting lines for each level of offensive production in Runs Created per Game.

Methodology

The first step was to determine the regular at each position (except pitcher) for each major league team from 1969 to 1989. I did this using the Baseball Encyclopedia, generally choosing the player listed there as the regular as long as he had made at least 243 outs (about 9 games (9 x 27) worth, a full season being 18 games worth (18 x 9 players = 162 games,)) although sometimes, especially with catchers, I might go as low as 189 outs (7 games.) If there was no regular, then that position was not included in the study for that team and year. Next, using the statistics in Total Baseball, I determined the runs created per 27 outs (RC/G) for each regular for the years 1969 to 1988. I then noted if the player was a regular the next season at any position (or DH), even if he played several positions and could not be considered a regular at one position, such as Dick Allen in 1970. If he was traded and played regularly for his new team, he was considered to have kept his job. If he did not play regularly the following season due to injury, but returned to regular status the year after that, I counted him as keeping his job. However, if he never played regularly again, his last season was not included in the study. After I had all these data, I set up ranges of one-half or one runs created per 27 outs for each position. Then it was a simple matter of counting up the number of players in each category and the number who returned the following season.

Results

As you can see in Table 2, the rate of return varies greatly from position to position. The rates did not generally vary much between leagues. I had supposed that the replacement level would be slightly higher in the National League, since the designated hitter allows American League teams to put a weak fielder at DH and put a good field/no hit player in his stead, but that wasn't the case. Bill James had theorized that there would be a range of one-half runs created per 27 outs where the replacement level would change sharply. However, that did not prove to be true. The rates tend to change slowly; in some cases, it was lower at a slightly higher level of production, e.g. at second base, players hitting between 3.5 and 3.9 RC/G returned 78% of the time while at 4.0-4.4 RC/G, the rate of return was 71%. Note that although I studied a total of 504 team/seasons, the sample size in each range is fairly small.

RC/G	1B		2B		SS		3B	
< 2.0			0.0%	3	33.3%	15	0.0%	1
2.0-2.5	42.9%	7	50.0%	18	59.1%	44	0.0%	4
2.5-3.0			57.1%	49	68.1%	91	40.0%	20
3.0-3.5	40.0%	10	61.7%	81	78.7%	89	66.7%	39
3.5-4.0	48.4%	31	80.2%	86	78.4%	97	67.2%	61
4.0-4.5	63.0%	54	86.1%	72	88.9%	63	83.3%	84
4.5-5.0	78.6%	70	89.2%	65	100%	35	90.9%	66
5.0-5.5	87.3%	142	94.7%	76	100%	36	88.5%	104
5.5-6.0								
6.0-6.9	90.3%	93	100%	39	100%	16	100%	59
7.0+	98.7%	75					100%	44

The replacement level for each position is about 1.5 to 2 runs below average production. I expected it to be one run below, since using the Pythagorean method of predicting winning average, at the average level of 4.2 RC/G, 1 run below gives a winning average of .367, 1.5 below yields an average of .292, and 2 runs below average gives a result of .215. In the past I have used one run above average for pitchers, which seems superficially valid. Thus, the level for hitters shouldn't be so low for an average fielder. It may be that it is artificially lowered by only good fielders being allowed to play regularly at low levels of offense (e.g. the young Ozzie Smith), so the offensive replacement level for average fielders is actually higher than is shown on the chart.

RC/G	LF		CF		RF		C	
< 2.0			25.0%	4	0%	1	0%	1
2.0-2.5	40.0%	10	33.3%	3	63.6%	11	45.0%	20
2.5-3.0			55.6%	18			46.9%	49
3.0-3.5	36.7%	30	55.9%	34	55.0%	20	62.2%	74
3.5-4.0	40.0%	30	66.2%	68	55.6%	36	70.3%	64
4.0-4.5	69.8%	63	67.7%	65	58.0%	69	67.7%	65
4.5-5.0	81.7%	71	89.5%	76	79.2%	77	79.2%	72
5.0-5.5	78.8%	66	93.9%	115	86.6%	67	91.4%	81
5.5-6.0	89.8%	59			86.8%	76		
6.0-6.9	94.7%	95	98.7%	76	100%	61	95.7%	46
7.0+	96.4%	56	96.8%	31	98.5%	67		

Table 2 – The number under each position is the percentage of regulars who returned as regulars the following season. The number to the right is the number of regulars in each category. Boxes around two ranges indicate that the ranges were combined in the table (eg., at first base there were 7 regulars between 2.0 and 3.0 runs created per game.

In considering the validity of this study, one must consider several factors. One is that a single season's statistics may not accurately represent a player's ability- indeed, several players who kept their jobs after a poor season were established players having an off year and their teams obviously believed that they could do better the next year. Another is that other factors than offense are more important than I have hypothesized. Further, the overall level of offense varied during the period. Also, a manager may overestimate a player's offense based on an incomplete understanding of statistics. Finally, a team may not have an adequate replacement available and/or doesn't want to take a chance on an unproven player. Any of these factors could result in a "true" offensive replacement level being higher or lower than I have calculated.

Other Study

Another study of this question was done by Phil Birnbaum and published in *By The Numbers* in September, 1994. Mr. Birnbaum used the entire 20th Century as the basis for his study. Some differences in methodology included counting a player as returning if he ever played regularly again and counting all outfield positions together. He also tested the results when a player's Total Baseball Fielding Runs were taken into account. Both times the results were generally consistent with my findings: no clear replacement level but rather a gradual drop in return rate as hitting decreased.

Clifford Blau, 16 Lake St. #5D, White Plains, NY, 10603, proboy@ix.netcom.com ♦

Peer Reviewers Wanted

Occasionally, I might need some help reviewing articles submitted to *By the Numbers*.

First, I might eventually need some help from any statisticians out there. Although I have a degree in statistics, I've forgotten most of what I knew, and that wasn't much to begin with. Should someone submit some complicated significance test, it would be nice if I had somewhere to verify it, if it seems necessary.

Second, it might be sometimes nice to have an article reviewed by someone who's an expert in the particular topic. If you have a specific field of expertise – run statistics, for instance, or Hall of Fame voting – you'd be a good candidate to look at a study on the same subject.

And, finally, we might eventually want to review *every* submission to BTN. I don't foresee this happening right away, but, hey, if this thing really takes off, and I get twenty manuscripts a month, I may need some serious help.

If you're interested in participating in any of these three review processes, drop me a quick e-mail. I'll print everyone's expertise and e-mail address in the next issue.

Summary

August on SABR-L

Compiled by Phil Birnbaum, but hopefully by someone else next time

A summary of the sabermetric research and discussion posted in August on SABR-L, the internet mailing list.

Aug. 13: As part of a discussion of the suitability of the term "scoring position," **Tom Ruane** posts the percentage of runners on each base scoring on a hit:

Start	Hits	Scored	Percentage
Batter	42320	4985	8.5%
First	13897	3121	22.5%
Second	8297	6209	74.9%
Third	4571	4551	99.6%

Ruane concludes that the term "looks like a reasonable expression to me."

Aug. 17: **Cyril Morong** reports that in championship series between 1969 and 1983, the better hitting team defeated the better pitching team by a record of 14 series to 13. This breaks down into 8-3 in the AL, and 6-10 in the NL.

Aug. 23: Computing the percentage of innings scoring runs after an error, to the percentage of innings scoring runs after a hit, **Tom Ruane** finds that slightly more innings score runs after errors. Mr. Ruane gives data for both leagues, 1980-1996.

Aug. 24: Interestingly, **Tom Ruane** finds that the home team's winning percentage drops from the usual .540 all the way to .471 in h igh-scoring games (18 or more runs for the two teams combined.) The winning percentage drops consistently with

scoring, to .423 in 24+ run games, and to .314 in 28+ run games. The next day, **Stephen Grant** suggested two reasons: first, home field advantage should be higher in tight games (because the home team bats last, and can tailor its strategy); and, second, when the home team wins, fewer runs are likely to be scored because the bottom of the ninth is not played. (The first reason was also independently posted by **Jeff Angus**.)

Aug. 24: A busy **Tom Ruane** posted a lengthy analysis of the percentage of games in which players got a hit. Since 1980, the lowest percentage belonged to Kiko Garcia (43 of 111 games in 1980), and the highest to Kenny Lofton (94 of 112 in 1994). **Keith Karcher** weighed in the next day with an analysis for the 1998 Anaheim Angels.

Aug. 29: **Jeff Smith** asks whether 1.2*OBP+SLUG works better as an offensive formula then simply OBP+SLUG. **Michael Schell** answers on September 1, finding that a coefficient of 1.4 brings the formula closest to Linear Weights.

Aug. 30: In response to **Michael Schell**'s previous posting about variation in batting average, **Dan Levitt** tells us that, surprisingly, there is no more variation in BA between players in an expansion year than in any other year. Both standard deviations are exactly the same: .029. Mr. Schell responds, suggesting that there is likely at least a little bit of increased variation, but that perhaps it's lost in the noise. ♦

Is this a useful feature?

If most of us already receive SABR-L, seeing it again ain't gonna be much help. If, on the other hand, most of us don't get SABR-L, or don't look at it much, this might be an important "for the record" summary of what kind of issues are being looked at. What do you think?

Let me know, at the usual e-mail address – birnbaum@magi.com. If you like it, please consider volunteering to compile it for a month. Warning–it's more time-consuming than it looks.

Book Reviews Wanted

Every year, a number of books and magazines are published with a Sabermetric slant. Many of our members have never heard of them. It's only within the last few weeks, for instance, that I heard about the "Big Bad Baseball Annual," and I still don't know who writes it or what's in it. If there's me, there's probably other committee members who'd like very much to hear when this kind of stuff comes out.

If you own a copy of this, or any other baseball book of interest, we'd welcome a summary or a full-length review. Since we've hardly published for the last couple of years, even reviews of older books – say, 1997 or later – would be welcome. The only restriction, please: the book should have, or claim to have, some Sabermetric content.

Send reviews to the usual place (see "Submissions" elsewhere in this issue). Drop me a line if you want to make sure no other member is reviewing the same publication, although multiple reviews of the same book are welcome, particularly for major works – the Bill James Book of Baseball Managers, say (hint, hint).

And if you're an author, and you'd like to offer a review copy, let me know – I'll find you a willing reviewer.

Submissions

Submissions to *By the Numbers* are, of course, encouraged. Articles should be concise (though not necessarily short), and pertain to statistical analysis of baseball. Letters to the Editor, original research, opinions, summaries of existing research, criticism, and reviews of other work are all welcome.

Articles should be submitted in electronic form, either by e-mail or on PC-readable floppy disk. I can read most word processor formats. If you send charts, please send them in word processor form rather than in spreadsheet form; this will make it easier for me to format your charts for publication. Unless you specify otherwise, I may send your work to others for comment (ie, informal peer review).

I will acknowledge all articles within three days of receipt, and will try, within a reasonable time, to let you know if your submission is accepted.