

# By the Numbers

Vol. 7 No. 1

June 1997

Neal Traven, Editor

Newsletter of the SABR Statistical Analysis Committee

© 1997 Society for American Baseball Research

## INSIDE

Returning to Action .....	1
Committee Leadership .....	1
Notes, Near and Far .....	1
SABR News .....	2
Current Research .....	2
And Furthermore.....	4
Editor's Comment .....	5
Final Thoughts .....	5

## Returning to Action

After far too long between newsletters, the revitalized Statistical Analysis Committee is back in print. I won't try to make any excuses about our tardiness. Instead, I'll start off by bringing you up to speed on how your current committee Co-chairs ended up in this position and then move some meatier stuff.

## Committee Leadership

Heading into the 1996 SABR National Convention, the Statistical Analysis Committee Co-Chairs both informed SABR brass that they would be unable to continue in their leadership roles. Before and during the Convention, a number of Committee members expressed some

degree of interest in assuming the position of Chair. I think it's no accident that, at their concluding business meeting in Kansas City, the SABR Board of Directors selected the two applicants who regularly attend national conventions -- Clem Comly and myself -- as the new Co-Chairs of the Committee.

Although our appointment may not have been a surprise, from SABR's standpoint it surely *is* a coincidence that Clem and I are both lifelong Phillies fans. As you can see from the addresses shown in the box at the bottom of the page, Clem still lives in the Delaware Valley while I've moved into Red Sox territory. For my first foray into Fenway Park immediately before the Convention, I'll observe the incongruous sight of my favorites visiting the Bosox.

Clem and I would be happy to hear from any of you regarding Committee matters. My preferred mode of communications is e-mail, though of course you can also use snail-mail or voice. To the best of my knowledge, Clem has somehow avoided obtaining an e-mail address so far, so you'll have to use those archaic media to get in touch with him.

This issue of *BTN* is being distributed based on a hodgepodge of mailing lists and other sources. Prior Co-chair Dave Raglin sent me a list that is, by his reckoning, about two years old. From John Zajc at the SABR office in Cleveland, I received a list of all SABR members who'd checked off "05" (statistical analysis) when joining or paying their dues. For a fleeting moment, I actually entertained fantasies of printing, collating, addressing, stamping, and mailing newsletters to over 1800 members. Once it dawned on me how monumental a task that would have been, I fell back to Plan B -- adding to Dave's old list the names of a number of SABRites who I'm sure will be pleased to find themselves

here on the committee mailing list.

After my struggle to ascertain who we are, I'd like to get a better handle on Statistical Analysis Committee membership. Thus, I ask for feedback of some sort from **every** recipient who wishes to remain on the committee. If you want to stay on the committee (and its mailing list), please let me or Clem know. In addition, if you know a SABR member who would like to join our merry little band, by all means tell us that as well!

At the moment, I can't guarantee that *By the Numbers* will appear in your mailbox on a fixed schedule. In large measure, how often committee news and views are disseminated to the membership depends on *you, the membership*. I can't distribute information that I don't know about, so the newsletter won't contain much of anything unless and until you send along materials for potential inclusion. Right now, there is very little backlog ... I was thrilled to find a manuscript in my mail just as this issue went to press, but aside from that all I have is some very sketchy material, my own work (which desperately needs to be reanalyzed before it sees the light of day), and the research paper you'll see later in this issue of *BTN*.

## Notes, Near and Far

With the explosion of information availability on the Internet, many of you are probably far more attuned than I am to what's going on here, there, and everywhere. Even so, this newsletter could serve to facilitate the exchange of research ideas, questions about previous inquiries on subjects you find interesting, data sources you've uncovered, information you wish to share with this audience, and so forth.

*By the Numbers*  
SABR Statistical Analysis Committee

Co-Chairs:

Clem Comly  
308 Colonial Drive  
Wallingford PA 19086  
610-563-7948

Neal Traven  
124 Tri-City Road  
Somersworth NH 03878  
603-749-6177  
baseball@worldpath.net

Just in case you might not have heard all there is to know, then, here are some items of possible interest to committee members:

- ◆ SABR members who've expressed interest in statistical analysis were very well represented among the research papers presented at last year's National Convention. By my count, 14 of the 25 speakers had research interest "05" listed in their membership records. That's 56% of the presentations, a far greater proportion than our approximate 30% representation in the SABR membership database. Eight of those people are on the *BTN* mailing list.
- ◆ **Mike Emeigh** ([mwe@nomos.com](mailto:mwe@nomos.com)) is working on situational simulations. He's already inquired on SABR-L and other Internet sources, but some of you may not subscribe to the list. His approach sounds to me like a combination of Markov chains and Monte Carlo methods, but don't take my word for it.

- ◆ **Jim Box** ([box@stat.duke.edu](mailto:box@stat.duke.edu)) is webmaster for the Statistics in Sports Section of the American Statistical Association. The ASA publishes several journals including *Journal of the American Statistical Association*, *American Statistician*, and *Chance*. All have published research on baseball, but *Chance* has published the most and its technical level is directed to the widest audience. The ASA-SIS website address is <http://www.stat.duke.edu/~box/sis/>.

- ◆ several months ago, **Jim Cochran** ([cochraj@ucbeh.san.uc.edu](mailto:cochraj@ucbeh.san.uc.edu)) sent me a letter about a cluster of sessions on "Sports and Gaming in Management Science" being planned for the spring 1999 meeting of INFORMS (the Institute for Operations Research and Management Science) in Cincinnati. As Chair of that cluster, he wants to encourage analytically-oriented committee members to present their research in that cluster. Though I didn't get back to Jim (until now, that is), this is definitely worth looking into, particularly for those with an academic bent.

- ◆ at the recent Lajoie-Start Southern New England regional meeting, new committee member **David J. Smith** ([dsmith@ultranet.com](mailto:dsmith@ultranet.com)) asked me whether I knew of anyone doing analytic work on catching -- passed balls and wild pitches, as well as other defensive attributes. I referred him to his near-namesake for historical data, but I'm sure that someone else on the committee has more information on the topic than I do.
- ◆ speaking of which, the latest issue of *The Retro Sheet* notes that **David W. Smith** ([dwsmith@strauss.udel.edu](mailto:dwsmith@strauss.udel.edu)), a.k.a. the "real" Dave Smith, has moved his seven filing cabinets and 95 cartons of Retrosheet materials to a new home -- 20 Sunset Road, Newark DE 19711. No change in phone number. The Retrosheet webpage has also relocated ... it can now be found at <http://www.retrosheet.org>. As the old saying goes, please change the bookmark in your browser.

## SABR News

Publications Director **Mark Alvarez** seeks ways to improve the quality and quantity of statistically sound articles in SABR publications, particularly the *Baseball Research Journal*.

Acknowledging that his own expertise in the statistical arena is low, he's looking for a panel of interested parties who would be willing to rotate annually through an appointment as Statistical Editor. The position requires reading and critiquing statistical submissions (perhaps as many as 30 a year), and making recommendations about their publication-worthiness to Mark.

Mark outlines three requirements for inclusion on this rotating panel:

- a) have the respect of their statistically-minded peers,
- b) be good SABR team players,
- c) be willing and able to meet deadlines

Each year's Statistical Editor would be listed on the masthead of *BRJ*. It is, of course, a volunteer position.

In response, I informed Mark that I would be willing to serve as the first Statistical Editor. We hadn't gotten the reborn *By The Numbers* off the ground at that point, and it seemed prudent to put someone in place right away. On Clem's suggestion, I also informed Mark that we would be most interested in publishing articles deemed too technical or too esoteric for *Baseball Research Journal* in future issues of our own committee newsletter. Mark was quite amenable to that concept.

If you wish to volunteer to join the rotation of SABR Statistical Editors-to-be, please let me or Clem know. Include a brief statement describing how you meet the requirements presented earlier. If you really, really, really want to be Statistical Editor right now, I won't stand in your way. Mark Alvarez, though, has the final say on the composition of this panel.

---

After too many years of bleary-eyed meetings in the very early mornings (often following very late nights) of SABR National Conventions, we finally lucked out this time around! The annual Statistical Analysis Committee meeting in Louisville is scheduled for **Saturday, June 21 at 9:00AM**. The only conflicting meeting is that of the Biographical Committee.

## Current Research

Our feature presentation in this issue of *By the Numbers* is a paper sent in by **Bill Gilbert** ([73623.2146@compuserve.com](mailto:73623.2146@compuserve.com)). Continuing a long-term interest, Bill examines the astonishing outburst of offensive performances last summer.

---

*Who Were the Best Offensive Players in a Record Offensive Year?*

**By Bill Gilbert**

A strong case can be made that the 1996 season was the greatest offensive year ever. The major league batting average was up three points (.267 to .270), on-base average was up two points (.338 to .340), and slugging average was up by ten points (.417 to .427, second only to .434 in 1930).

Since none of these averages provide a complete picture by itself, a more comprehensive measure of total offensive performance is useful. Such a measure would include the following elements:

1. The ability to hit safely.
2. The ability to get on base.
3. The ability to hit with power.
4. The ability to add value through baserunning.

The first three elements are measured by batting average, on-base average, and slugging average, respectively. A measure of offensive performance which encompasses all three as well as baserunning achievements is *bases per plate appearance* (BPA). This measure accounts for the net bases accumulated by a player per plate appearance. It is calculated as follows:

$$BPA = \frac{TB+BB+HB+SB-CS-GIDP}{AB+BB+HB+SF}$$

where:

- BPA = Bases per Plate Appearance
- TB = Total Bases
- BB = Bases on Balls
- HB = Hit by Pitch
- SB = Stolen Bases
- CS = Caught Stealing
- GIDP = Grounded into Double Plays
- AB = At Bats
- SF = Sacrifice Flies

The numerator accounts for all of the bases accumulated by a player, reduced by the number of times he is caught stealing or erases another runner by grounding into a double play. The denominator accounts for plate appearances when the player is trying to generate bases. Sacrifice hits are not included as plate appearances, since they represent the successful execution of the batter's attempts to advance another runner rather than himself.

The major league average BPA in 1996 was .471, up from .463 in 1995. The major league BPA was .467 in 1994, .446 in 1993 and .423 in 1992. A player with a BPA of .500 or above is making a significant offensive contribution. A BPA of .550 represents an outstanding season and .600 is reserved for superstars. Among players with enough plate

appearances to qualify for the batting title in 1996, 41 recorded a BPA of .550 or better compared to 37 in 1995, 38 in 1994, 29 in 1993, and 13 in 1992.

In 1996, 21 players were over .600 versus 15 in 1995, 16 in 1994, eight in 1993, and four in 1992. No players had a BPA over .700 in 1995 but in 1996 two players (neither a strong MVP candidate) were

well over .700 -- Mark McGwire (.765) and Barry Bonds (.730). Bonds was also over .700 in 1994 (.738) as were Jeff Bagwell (.768), Frank Thomas (.747), Albert Belle (.741), and Ken Griffey, Jr. (.708). Bonds was the only player over .700 in 1992 (.734) and 1993 (.740).

Table 1 is a list of players with enough

1996			1995				
1996	BPA	Lg	1995	BPA	Lg		
1	Mark McGwire	.765	A	1	Albert Belle	.692	A
2	Barry Bonds	.730	N	2	Edgar Martinez	.681	A
3	Gary Sheffield	.690	N	3	Barry Bonds	.671	N
4	Brady Anderson	.689	A	4	Frank Thomas	.662	A
5	Ellis Burks	.683	N	5	Reggie Sanders	.656	N
6	Ken Griffey Jr	.683	A	6	Larry Walker	.646	N
7	Jim Thome	.668	A	7	Tim Salmon	.638	A
8	Albert Belle	.660	A	8	Ron Gant	.626	N
9	Juan Gonzalez	.655	A	9	Jim Thome	.623	A
10	Alex Rodriguez	.653	A	10	Mike Piazza	.619	N
11	Jeff Bagwell	.651	N	11	Barry Larkin	.617	N
12	Edgar Martinez	.651	A	12	Mo Vaughn	.610	A
13	Ken Caminiti	.647	N	13	Dante Bichette	.609	N
14	Frank Thomas	.646	A	14	John Valentin	.609	A
15	Barry Larkin	.641	N	15	Rafael Palmeiro	.606	A
16	Andres Galarraga	.632	N	16	Mickey Tettleton	.595	A
17	Chuck Knoblauch	.625	A	17	Manny Ramirez	.592	A
18	Mo Vaughn	.618	A	18	Jose Canseco	.587	A
19	Bob Higginson	.617	A	19	Craig Biggio	.586	N
20	Manny Ramirez	.607	A	20	Jay Buhner	.579	A
21	Rafael Palmeiro	.607	A	21	Sammy Sosa	.579	N
22	Greg Vaughn	.599	A/N	22	Tino Martinez	.579	A
23	Jay Buhner	.592	A	23	Chuck Knoblauch	.577	A
24	Bernard Gilkey	.591	N	24	Ray Lankford	.577	N
25	Todd Hundley	.590	N	25	Bobby Bonilla	.567	N/A
26	Sammy Sosa	.590	N	26	Jeff Bagwell	.566	N
27	Bernie Williams	.586	A	27	Chili Davis	.565	A
28	John Jaha	.583	A	28	Harold Baines	.564	A
29	Chipper Jones	.582	N	29	Rickey Henderson	.560	A
30	Mike Piazza	.580	N	31	Eric Karros	.558	N
31	Roberto Alomar	.576	A	32	Kenny Lofton	.556	A
32	Henry Rodriguez	.573	N	33	Tony Phillips	.553	A
33	Ray Lankford	.573	N	34	Mark Grace	.553	N
34	Rusty Greer	.571	A	35	Jim Edmonds	.553	A
35	Ryan Klesko	.570	N	36	Raul Mondesi	.550	N
36	Dave Nilsson	.568	A	37	Brady Anderson	.550	A
37	Tim Salmon	.561	A				
38	Dante Bichette	.559	N				
39	Kenny Lofton	.558	A				
40	Steve Finley	.558	N				
41	Vinny Castilla	.550	N				

plate appearances to qualify for the batting title who achieved a BPA of .550 in one of the last two seasons.

The 1996 list includes 20 repeaters from 1995. It also contains six players with a BPA over .550 in both years who didn't have enough plate appearances in 1995 to qualify (McGwire, Sheffield, Burks, Griffey, Gonzalez, and Klesko). Five players who made the 1995 list were over .550 in 1996 but were short of plate appearances in 1996 (Sanders, Walker, Gant, Canseco, and Edmonds). The list of 12 players who made the 1995 list but slipped below .550 in 1996 is of interest. Since most of these players are over 30 years old, this may be evidence that they are on the way down (John Valentin, Tettleton, Biggio, Tino Martinez, Bonilla, Chili Davis, Baines, Henderson, Karros, Phillips, Grace, and Mondesi).

Finally, the 1996 list includes some players who were relatively unnoticed in putting together outstanding seasons (Higginson, Gilkey, Jaha, Greer, Nilsson, and Finley). For some, it may turn out to be a career year but for others it may represent a breakthrough to a new level.

Table 2 displays all players who compiled a batting average over .300, an on-base average over .400, a slugging percentage over .500, and bases per plate appearance over .600 in 1996:

Table 2

Player	BA	OBA	SLG	BPA
Mark McGwire	.312	.467	.730	.765
Barry Bonds	.308	.461	.615	.730
Gary Sheffield	.314	.465	.624	.690
Ellis Burks	.344	.408	.639	.683
Jim Thome	.311	.450	.612	.668
Albert Belle	.311	.410	.623	.660
Alex Rodriguez	.358	.414	.631	.653
Jeff Bagwell	.315	.451	.570	.651
Edgar Martinez	.327	.464	.595	.651
Ken Caminiti	.326	.408	.621	.647
Frank Thomas	.349	.459	.626	.646
C Knoblauch	.341	.448	.517	.625
Mo Vaughn	.326	.420	.583	.618
Bob Higginson	.320	.404	.577	.617

Thomas has made this list for the past four years and Belle has been on it for three

years. Thome and Martinez are the only other repeaters from 1995.

Also of interest is a list of players with BPA of over .550 in 1996 who, for one reason or another, did not have enough plate appearances to qualify for the batting title. Following is a list of players with 100 or more plate appearances who fell short of having enough playing time to qualify for the batting title:

- Jose Canseco (.641) -- a sample of what could have been.
- Brian Giles (.636) -- hits well enough to play regularly.
- Larry Walker (.635) -- one of the best when healthy.
- Kevin Mitchell (.617) -- produces when he plays.
- Jim Edmonds (.605) -- held back by injury.
- Eric Davis (.604) -- sensational comeback year.
- Matt Stairs (.589) -- could be a sleeper in 1997.
- Reggie Jefferson (.589) -- is he really this good?
- Rex Hudler (.578) -- career year at age 35.
- Ron Gant (.575) -- big list candidate when healthy.
- Curtis Pride (.574) -- productive in a limited role.
- Mike Stanley (.569) -- fourth straight productive season.
- Scott Brosius (.563) -- moved game to a new level in 1996.
- Phil Nevin (.562) -- may finally be maturing.
- Reggie Sanders (.560) -- still has chance to be a superstar.
- Tim Lincecum (.550) -- still produces when he plays.

This list may be of particular interest to fantasy league players. It includes a number of players who may be undervalued in 1997 because of injury-plagued seasons in 1996 (Canseco, Walker, Edmonds, Gant and Sanders). It also contains relatively unknown players who may be on the verge of stardom (Giles, Stairs, Pride, Brosius and Nevin). Some of the names on this list in 1995 who had big years in 1996 were McGwire, Sheffield, Palmer, Klesko, Jaha, Eric Young, Griffey Jr., and Burks.

### And Furthermore...

In an e-mail reply to Bill after reading his interesting analysis of 1996, I said in part:

*One thing that bothered me just a bit about your paper is the long-discussed problem of adjustment for park, year, and*

*so forth. Simply noting that (for example) Coors Field inflates offense is, in my opinion, insufficient. The same goes for, say, comparing a 1996 BPA to one from 1968.*

*I acknowledge that "too much" adjustment can suck some of the thrill out of the numbers, but I still wonder whether you'd be interested in saying just a bit about what we might call 'relative BPA' -- the ratio of some of the leading BPAs to the league-average BPA for the year in question.*

While reiterating his own preference for examining the numbers *in situ*, Bill did send me a short addendum that analyzes offensive performance adjusted for year effects. Following that amplifying essay on BPA, I'll be back with some additional thoughts.

### Relative BPA -- 1992-1996

By Bill Gilbert

As shown in Table 3, in the five year period from 1992 through 1996, there were nine occasions when a player was more than 50% better offensively than the major league average as measured by *bases per plate appearance* (BPA). The relative BPA is the ratio of a player's BPA to the average major league BPA. Thus, a relative BPA of 1.50 indicates that a player is 50% more productive offensively than the average major league player that year.

Table 3

Player	Lg	Yr	BPA	Rel BPA
1 Barry Bonds	N	92	.734	1.74
2 Barry Bonds	N	93	.740	1.66
3 Jeff Bagwell	N	94	.768	1.64
4 Mark McGwire	A	96	.765	1.62
5 Frank Thomas	A	94	.747	1.60
6 Albert Belle	A	94	.741	1.59
7 Barry Bonds	N	94	.738	1.58
8 Barry Bonds	N	96	.730	1.55
9 Ken Griffey Jr	A	94	.708	1.52

The major league average BPA was .423 in 1992, .446 in 1993, .467 in 1994, .463 in 1995, and .471 in 1996.

Barry Bonds is the only player with more home runs in one such season and he did it in four of the past five years. In the only year he didn't make this list, he was 45% better than the major league average in 1995. If there was any doubt, this confirms that Bonds is the best offensive player in an era of strong offense.

Only four of the nine seasons were in the offensive-minded American League and none were accomplished in extreme hitters' parks. Jeff Bagwell did it in a notorious pitchers' park in 1994.

## Editor's Comment

The advent of Major League baseball played a full mile above sea level, simultaneous with the oft-repeated increase in offense following expansion, has reinvigorated the arguments in support of park and year adjustments. While marveling at the prodigious home run totals and other gaudy numbers being rung up by the likes of Bichette, Galarraga, and Stilla, even the most innumerate commentator seems to understand (however dimly) that there's something unusual about Colorado.

On the other hand, striking changes in patterns of offense have occurred in the past and will occur again. Undoubtedly, at some point in the future media pundits will wail about the disappearance of hitting and the pre-eminence of pitching in baseball. The more things change, ...

The issue of whether or not to adjust for quantifiable factors such as year and park effects has been argued at great length, often with (too) great heat. In many cases, proponents of each side end up talking past each other, repeating and repeating the same points as if somehow the Nth time around will convince the adversary of the error of his ways.

Despite the dangers of inflamed passions, I believe that our newsletter, and our committee, can serve as a forum for real, honest, well-mannered, thoughtful

discussion of the issue. Therefore, I invite any and all subscribers interested in presenting their thoughts on the matter to do so by contributing their reasoning and evidence to *BTN*. If there's enough response, perhaps we could establish a working group charged with clarifying and tightening the salient points on all aspects of the issue.

In addition to submissions to *BTN*, perhaps members attending our committee meeting in Louisville can begin this process.

Through it all, I'll try to remain as much of an agnostic on the subject as I can. While my own thoughts tend to favor adjustment, I am far from a zealot on the matter.

## Final Thoughts

As I said earlier, the content and even the publishing schedule of *By the Numbers* depends on you. I can't write about data needs, research plans, address changes, or committee news and notes unless you tell me about them. Similarly, I can't publish research reports unless I have those papers in hand.

The current inventory of research reports is meager. As mentioned on page 1, I received a manuscript literally the day before this writing these final paragraphs of this issue of *BTN*; it's been given only the most fleeting of glances so far.

Which brings me to the question of media for submissions. While Clem has no net-access that I'm aware of, my own preference is very strongly toward e-mail as a mode of communications.

I'd greatly appreciate it -- and am also far more likely to reply in short order -- if you send me short items such as requests, address changes (snail or e-mail), and the like in the form of e-mail messages.

For manuscripts and other such materials, formatting of tables, graphs, and equations is a vital factor. I'm as concerned about mistakes and misstatements due to errors

such as misinterpreted notation as you, the author, would be. Therefore, I'd like those sorts of submissions sent as both hard copy -- showing the way you mean the material to be read -- and on electronic media. The latter form is essential, as I have no intention to ever retype something from hard copy.

I use OS/2 Warp and Windows 3.1 (not Windows95) on a Pentium/133 with 32 MB. My wordprocessor of choice is WordPerfect for Windows 6.1 and I also have slight access to Microsoft Word (version unknown). I can easily read Quattro Pro/Excel/Lotus spreadsheets, Paradox/Access databases, Power Point/Presentations graphics, and a good variety of standard statistical analysis software formats. I cannot make any use of Macintosh diskettes or Macintosh software.

I can receive formatted files as mail attachments from some mailers, but I'm not convinced that attaching, say, a WPWin file to an e-mail message will always work. In general, then, you could begin by sending such materials as an attachment. I'll let you know whether I can successfully open the file(s) you send. If I can't, I'll ask for a diskette in a mailer. And of course, I'd also like to see the hard copy so that I can be certain that the little problems that always crop up to prevent true WYSIWYG haven't conspired to create a terminally-garbled document.

Submissions will be reviewed by me and/or Clem, or another reviewer of our choosing (any volunteers?). We may ask you to rework the report in some manner -- updating to a more recent year, perhaps, or maybe expanding/contracting some sections of the paper.

All in all, I hope to hear interesting and valuable information from committee members in the near future. Especially since those from whom I hear nothing may be dropped from the Statistical Analysis Committee mailing list in the future.