By the Numbers

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

Volume 6, Number 1

March, 1994

Greetings

Another baseball season is upon us. I hope you share my love for this time of year when all teams fervently believe that they will win the pennant "if we just catch some breaks" (e.g., if their unproven rookie pitcher can only win 20 games, if their broken-down first baseman can only recapture his youth by belting 40 home runs, etc.). True fans all too gladly surrender to this annual dose of optimism. I for one am sure that this is the breakthrough year for Royce Clayton and that he'll finally play up to expectations (nothing short of Honus Wagner incarnate will do).

While I'm on a roll, let me comment on the results of this year's Hall of Fame voting. Orlando Cepeda's near miss on his last year of BBWAA balloting points out that the "second-tier" of 1960's hitting stars have had a tough time. Ron Santo, Dick Allen, Jimmy Wynn, Ken Boyer, Joe Torre, Frank Howard, Vada Pinson, Willie Davis, Boog Powell, Bill Freehan, Rocky Colavito, Tony Oliva, Bill Mazeroski, Curt Flood, et al., will likely never be enshrined in Cooperstown (though Brooks Robinson and Billy Williams made it). I am not making a case for any or all of the above, only pointing out that many comparables to these players can be found in the Hall of Fame. The difference may lie in the negative effect the pitching-dominated 1960's had on these hitters' career stats.

SABR's national convention is coming up sooner than you may think. Make your plans to be in Arlington, Texas from Thursday June 16 thru Sunday June 19. The convention is always a lot of fun and provides a golden opportunity to meet many knowledgeable baseball fans from around the country. In addition, a good many of the research presentations will be based upon statistical analysis and provide a cross-section of work being done in this area. The time and date of the Statistical Analysis Committee meeting has not yet been finalized. Hope to see you in Arlington!

Although we are maintaining a moderate backlog of material, we always encourage members to submit their research for publication in the newsletter. Short articles work best in this format, with appendices or more detailed material made available to interested readers. If possible, please send an electronic version (I use Word for Windows, though can handle other word processors or Mac) of material to accompany the hard copy.

This issue of the newsletter contains five interesting articles. **Tony Blengino** kicks us off with an article investigating whether good pitching really stops good hitting. Note that Tony's research is independent of Don Coffin's work which was published in the last issue of the newsletter. Though they look at different data, and use slightly different methods, Tony and Don reach roughly the same conclusions.

Charlie Pavitt looks into the issue of resting players during the long season. Specifically, Charlie asks the question "Does Cal Ripken Tire?".

Bill Gilbert follows-up a previous article on the prevalence of "triple milestone" hitters in the offensive-minded season of 1993.

Willie Runquist writes on the difficulty of using statistical methods to "prove" or "disprove" the existence of clutch hitters.

Lawrence Hadley and Elizabeth Gustafson use economic and statistical methods to investigate the relationship between baseball's free agency and competitive balance.

Please send material, comments, etc., to my address: **Rob Wood**, 2101 California St. #224, Mountain View, CA 94040. My home number is (415) 961-6574, and my daytime number is (415) 854-7101.

****

Notes from committee members

In this section of the newsletter, I will pass along news/ideas/information I receive from committee members.

Mike Savage inquires into the existence of an index or bibliography of past issues of the committee newsletter. If you are aware of one, please let me know and I will make it available to interested members.

Steve Constantelos hopes that SABR continues its practice of surveying members to get their views of (usually long-ago) teams or players. Steve's particular interest lies in the awarding of retroactive Gold Gloves. If you have any views on this matter, please contact me.

Alex Woo wonders if any computer baseball games use minor league projections or any "aging" effects (ala Bill James's Brock-system) on major leaguers. If you have any knowledge of such efforts, please contact me.

Dan Rappoport has developed a data base of Cy Young and MVP voting. He has written up various findings (such as total cumulative votes, consecutive year streaks, etc.). If you would like a copy of the report, contact Dan at Apt. 5M, Holly House, Princeton, NJ 08540.

****

Notes from SABR headquarters

David Pietrusza and Morris Eckhouse have passed along to me information regarding SABR's proposed participation with Public Broadcasting Service's "Learning Link Online Service". PBS is considering adding a Baseball Forum to its Learning Link online service. SABR and PBS are presently discussing the arrangements and SABR's possible role. An announcement may be made as early as April.

Most users of the Learning Link would be schoolchildren (K-12). At this time, David and Morris are asking committees if we would be interested in assisting in this effort. Though details are not yet finalized, at a minimum one member of each committee would assist in answering online questions from Learning Link users.

Reasons for SABR's participation include: (i) online services are growing and SABR desires to not miss out on the opportunity to link up to the electronic super-highway; (ii) to counteract the fact that baseball is losing ground to other sports among younger age groups; (iii) to tie into the Ken Burns's PBS film on baseball; (iv) to demonstrate that SABR provides benefits which extend beyond its own membership.

If you are interested in helping out or learning more about the proposed project, please contact me.

****

CONVENTIONAL WISDOM DESTRUCTION 101: GOOD HITTING BEATS GOOD PITCHING

by Anthony P. Blengino

It usually happens sometime in April -- inevitably a baseball announcer (of the talking head variety) will comment on the relative importance of pitching in baseball. "Well, Joe, as you know, pitching is 'x' percentage of the game", he will utter. The percentage will vary, usually between 65-75% (tracking that, in itself, could be a good topic for a future paper), but it's never less than 50%.

My research shows that their conclusions, like much other conventional baseball wisdom, are simply not true. In fact, since the beginning of divisional play in the National League, scoring runs has been more directly correlated with winning than preventing runs. Further, using the same population of teams, my research shows that it has been more likely for an average defensive (meaning pitching and fielding) team with above average offense to win division titles than it has been for an average offensive team with above average defense.

These and other factors combine to show that, while it is still impossible to put an exact percentage on the relative contribution of defense to winning, that percentage is much lower than estimates routinely derived from conventional baseball wisdom, and is most likely LESS than 50%. In addition, my research offers some insights on the relative importance of specific offensive and defensive statistics, and the identification of some of the strongest (and weakest) offensive, defensive, and overall NL teams since the beginning of

divisional play -- including some very strong squads that didn't even win divisional titles.

Research Methods: I set out to measure all-around offensive and defensive performance of all National League teams since the beginning of divisional play in 1969. I identified 10 core offensive and defensive statistics to be used as the database for my analysis. The ten offensive categories are:

- Runs scored
- Doubles
- Triples
- Home runs
- Walks
- Strikeouts
- Batting average
- On-base percentage
- Slugging percentage
- Stolen bases.

The ten defensive categories are:

- Complete games
- Shutouts
- Home runs allowed
- Walks
- Strikeouts
- ERA
- Opponents' batting average
- Opponents' on-base percentage
- Errors
- Double plays.

I then ranked all NL teams (from 1969 to 1993), awarding each 1st place team in each category x points, each 2nd place team x-1 points, etc. I then compared each team's total offensive, defensive, and overall points accumulated to the maximum points attainable, and calculated offensive, defensive, and overall Statistical Winning Percentages (SWP).

Like the year-end standings, all offensive, defensive, and overall SWP's average out to .500 leaguewide for each season, making comparisons across eras possible, even when the raw offensive and defensive statistics vary widely (e.g., 1972 vs. 1987). For instance, in 1992 the Pittsburgh Pirates ran up an offensive SWP of .791 by accumulating 87 of a possible 110 offensive points, and had a defensive SWP of .673, accumulating 74 of a possible 110 defensive points, their overall SWP was .732, which as we shall see, is quite good, even for a division winner.

The Division Winners: After calculating offensive, defensive, and overall SWP's for *all* NL teams since 1969, I organized the data for the 50 division winners in the study. (I arbitrarily decided to include the 1981 Cards and Reds rather than the Expos and Dodgers; the Redlegs and Redbirds had the best records, but didn't win either half in that sorry season.) Their SWP's appear in Table A.

The offensive SWP's of the NL division champs range from a high of .900 (1976 Reds) to a low of an incredible .227 (1973 Mets). The defensive SWP's ranged from a high of .905 (1977 Dodgers) to .373 (1982 Braves). The overall SWP's ranged from a high of .850 (1988 Mets -- the Kirk Gibson-ized ones) to a low of .461 (again, those 1982 Braves). The average offensive SWP of these 50 teams was .668 and the average defensive SWP was .646, combining for an overall SWP of .657.

The offensive SWP is higher, but not by a statistically significant margin. The very fact that the offensive SWP is higher, however, makes it quite unlikely that the combination of pitching and fielding contribute significantly more than 50% to the winning of ballgames, as conventional baseball wisdom would suggest. A look at the offensive and defensive SWP's of all NL teams since 1969 reveals more specific information with respect to the relative contributions of offense and defense to the winning of division titles.

High Offense vs. High Defense: Since 1969, 46 NL teams have had an offensive SWP of .700 or higher (see Table B). Of these 46 teams, 23 (50.0%) won division titles. Since 1969, 47 NL teams have had a defensive SWP of .700 or higher (see Table C). Of these 47 teams, only 16 (34.0%) won division titles. This difference, obviously, is statistically significant, and favors offense's relative contribution to the winning of division titles.

There is a somewhat mitigating statistic in favor of defense. Since 1969, there have been only 3 NL division champs with extremely low offensive SWP's (1988 Dodgers= .332, 1969 Mets= .300, and 1973 Mets= .227), while the lowest defensive SWP was achieved by the 1982 Braves (.373). This indicates that a base level of defense is necessary for a team to win its division. While no noticeable base level of offense appears to be necessary, incremental offensive performance above very low levels is much more important than incremental defensive performance above the base level. In other words, you'd better have at least average pitching if you

STATISTICAL WINNING PERCENTAGES
NL DIVISION WINNERS (1969–1993)

Table A

EAST

		OFF.	DEF.	OVERALL
1993	PHL	0.762	0.577	0.670
1992	PIT	0.791	0.673	0.732
1991	PIT	0.859	0.673	0.766
1990	PIT	0.795	0.545	0.670
1989	CHI	0.655	0.432	0.544
1988	NYM	0.800	0.900	0.850
1987	STL	0.659	0.527	0.593
1986	NYM	0.745	0.809	0.777
1985	STL	0.800	0.841	0.821
1984	CHI	0.818	0.495	0.657
1983	PHL	0.577	0.595	0.586
1982	STL	0.736	0.573	0.655
1981	STL	0.709	0.609	0.659
1980	PHL	0.800	0.541	0.671
1979	PIT	0.745	0.614	0.680
1978	PHL	0.655	0.655	0.655
1977	PHL	0.868	0.627	0.748
1976	PHL	0.805	0.627	0.716
1975	PIT	0.609	0.718	0.664
1974	PIT	0.664	0.518	0.591
1973	NYM	0.227	0.764	0.496
1972	PIT	0.686	0.714	0.700
1971	PIT	0.764	0.545	0.655
1970	PIT	0.545	0.673	0.609
1969	NYM	0.300	0.773	0.537
	AVG.	0.668	0.616	0.642

WEST

	OFF.	DEF.	OVERALL
ATL	0.588	0.873	0.731
ATL	0.655	0.745	0.700
ATL	0.768	0.595	0.682
CIN	0.673	0.686	0.680
SF	0.614	0.627	0.621
LA	0.332	0.727	0.530
SF	0.459	0.745	0.602
HOU	0.555	0.795	0.675
LA	0.609	0.750	0.680
SD	0.559	0.486	0.523
LA	0.459	0.741	0.600
ATL	0.550	0.373	0.462
CIN	0.641	0.605	0.623
HOU	0.618	0.836	0.727
CIN	0.564	0.586	0.575
LA	0.777	0.700	0.739
LA	0.564	0.905	0.735
CIN	0.900	0.545	0.723
CIN	0.809	0.527	0.668
LA	0.814	0.700	0.757
CIN	0.723	0.632	0.678
CIN	0.759	0.545	0.652
SF	0.691	0.523	0.607
CIN	0.773	0.582	0.678
ATL	0.555	0.464	0.510
AVG.	0.616	0.627	0.621

HIGH OFFENSE TEAMS
OFF. SWP > .700
NL (1969–1993)

Table B

		OFF.	DEF.	OVERALL
1976	***CIN***	***0.900***	***0.545***	***0.723***
1977	***PHL***	***0.868***	***0.627***	***0.748***
1991	***PIT***	***0.859***	***0.673***	***0.766***
1984	PHL	0.850	0.409	0.630
1971	STL	0.841	0.423	0.632
1981	PHL	0.832	0.355	0.594
1984	***CHI***	***0.818***	***0.495***	***0.657***
1974	***LA***	***0.814***	***0.700***	***0.757***
1975	***CIN***	***0.809***	***0.527***	***0.668***
1976	***PHL***	***0.805***	***0.627***	***0.716***
1980	***PHL***	***0.800***	***0.541***	***0.671***
1988	***NYM***	***0.800***	***0.900***	***0.850***
1985	***STL***	***0.800***	***0.841***	***0.821***
1983	ATL	0.795	0.368	0.582
1990	***PIT***	***0.795***	***0.545***	***0.670***
1973	ATL	0.795	0.205	0.500
1992	***PIT***	***0.791***	***0.673***	***0.732***
1978	***LA***	***0.777***	***0.700***	***0.739***
1987	NYM	0.773	0.641	0.707
1970	***CIN***	***0.773***	***0.582***	***0.678***
1991	***ATL***	***0.768***	***0.595***	***0.682***
1969	CIN	0.768	0.250	0.509
1982	PIT	0.768	0.382	0.575
1973	SF	0.764	0.286	0.525
1971	***PIT***	***0.764***	***0.545***	***0.655***
1993	SF	0.762	0.654	0.708
1993	***PHL***	***0.762***	***0.577***	***0.670***
1975	PHL	0.759	0.541	0.650
1972	***CIN***	***0.759***	***0.545***	***0.652***
1972	HOU	0.750	0.491	0.621
1992	CIN	0.750	0.582	0.666
1979	***PIT***	***0.745***	***0.614***	***0.680***
1977	CIN	0.745	0.491	0.618
1986	***NYM***	***0.745***	***0.809***	***0.777***
1974	CIN	0.736	0.586	0.661
1982	***STL***	***0.736***	***0.573***	***0.655***
1982	MTL	0.727	0.705	0.716
1973	***CIN***	***0.723***	***0.632***	***0.678***
1969	PIT	0.718	0.532	0.625
1980	STL	0.714	0.514	0.614
1972	CHI	0.709	0.686	0.698
1986	PHL	0.709	0.491	0.600
1982	LA	0.709	0.777	0.743
1981	***STL***	***0.709***	***0.609***	***0.659***
1983	STL	0.709	0.382	0.546
1970	SF	0.705	0.273	0.489

BOLD = Won Division Title (23/46 = 50%)

HIGH DEFENSE TEAMS
DEF. SWP > .700
NL (1969–1993)

Table C

		OFF.	DEF.	OVERALL
1977	***LA***	***0.564***	***0.905***	***0.735***
1989	LA	0.223	0.900	0.562
1988	***NYM***	***0.800***	***0.900***	***0.850***
1993	***ATL***	***0.588***	***0.873***	***0.731***
1973	LA	0.586	0.864	0.725
1985	***STL***	***0.800***	***0.841***	***0.821***
1980	***HOU***	***0.618***	***0.836***	***0.727***
1974	ATL	0.418	0.818	0.618
1975	LA	0.505	0.809	0.657
1981	HOU	0.477	0.809	0.643
1986	***NYM***	***0.745***	***0.809***	***0.777***
1981	LA	0.473	0.800	0.637
1979	HOU	0.377	0.795	0.586
1986	***HOU***	***0.555***	***0.795***	***0.675***
1976	NYM	0.423	0.791	0.607
1993	HOU	0.596	0.781	0.689
1982	LA	0.709	0.777	0.743
1982	PHL	0.459	0.773	0.616
1972	LA	0.509	0.773	0.641
1991	LA	0.500	0.773	0.637
1969	***NYM***	***0.300***	***0.773***	***0.537***
1970	NYM	0.414	0.768	0.591
1985	NYM	0.645	0.768	0.707
1973	***NYM***	***0.227***	***0.764***	***0.496***

		OFF.	DEF.	OVERALL
1984	PIT	0.318	0.750	0.534
1985	***LA***	***0.609***	***0.750***	***0.680***
1992	***ATL***	***0.655***	***0.745***	***0.700***
1987	***SF***	***0.459***	***0.745***	***0.602***
1979	MTL	0.527	0.745	0.636
1983	***LA***	***0.459***	***0.741***	***0.600***
1984	LA	0.173	0.741	0.457
1990	NYM	0.664	0.736	0.700
1987	HOU	0.255	0.732	0.494
1988	***LA***	***0.332***	***0.727***	***0.530***
1982	HOU	0.273	0.727	0.500
1984	HOU	0.550	0.723	0.637
1990	MTL	0.450	0.718	0.584
1971	LA	0.641	0.718	0.680
1975	***PIT***	***0.609***	***0.718***	***0.664***
1969	STL	0.564	0.718	0.641
1972	***PIT***	***0.686***	***0.714***	***0.700***
1980	LA	0.486	0.714	0.600
1983	PIT	0.455	0.714	0.585
1989	STL	0.659	0.709	0.684
1976	LA	0.432	0.705	0.569
1983	HOU	0.523	0.705	0.614
1982	MTL	0.727	0.705	0.716

BOLD = Won Division Title (16/47 = 34%)

OVERALL SWP > .700
NL (1969–1993)

Table D

		OFF.	DEF.	OVERALL
1988	***NYM***	***0.800***	***0.900***	***0.850***
1985	***STL***	***0.800***	***0.841***	***0.821***
1986	***NYM***	***0.745***	***0.809***	***0.777***
1991	***PIT***	***0.859***	***0.673***	***0.766***
1974	***LA***	***0.814***	***0.700***	***0.757***
1977	***PHL***	***0.868***	***0.627***	***0.748***
1982	LA	0.709	0.777	0.743
1978	***LA***	***0.777***	***0.700***	***0.739***
1977	***LA***	***0.564***	***0.905***	***0.735***
1992	***PIT***	***0.791***	***0.673***	***0.732***
1993	***ATL***	***0.588***	***0.873***	***0.731***
1980	***HOU***	***0.618***	***0.836***	***0.727***
1973	LA	0.586	0.864	0.725
1976	***CIN***	***0.900***	***0.545***	***0.723***
1982	MTL	0.727	0.705	0.716
1976	***PHL***	***0.805***	***0.627***	***0.716***
1993	SF	0.762	0.654	0.708
1985	NYM	0.645	0.768	0.707
1987	NYM	0.773	0.641	0.707
1990	NYM	0.664	0.736	0.700
1992	***ATL***	***0.655***	***0.745***	***0.700***
1972	***PIT***	***0.686***	***0.714***	***0.700***

BOLD = Won Division Title (15/22 = 68%)

OFFENSIVE & DEFENSIVE CATEGORY SWP (NL DIVISION WINNERS – 1969–1993)

Table E

OFFENSE	
RUNS	0.808
OBP	0.769
SLP	0.748
AVG	0.733
HR	0.701
BB	0.642
2B	0.627
3B	0.604
SB	0.557
K	0.488
Average	0.668

DEFENSE	
ERA	0.780
OBP	0.750
AVG	0.703
BB	0.685
SHO	0.654
E	0.645
HR	0.619
CG	0.593
DP	0.531
K	0.507
Average	0.647

want to win a division title -- but, given that, extreme levels of offensive performance are much more contributory to winning than extreme levels of defensive performance.

I'm one of those odd people who love dynasties and juggernauts -- I never root for the underdog, unless that team is from Philadelphia. That said, I had a particular interest in identifying the teams with overall SWP's above .700. There were 22 such teams (see Table D), and 15 of them (68.2%) won division titles.

Relative Contributions of Individual Offensive and Defensive Statistics: The nuts and bolts that assemble the aforementioned offensive, defensive and overall SWP's are the 20 core statistical categories. Their relative importance can be measured by calculating the average rank in each category achieved by the 50 division winners, in essence calculating a winning percentage for each statistic.

The winning percentages for all 20 categories can be found in Table E. As expected, the highest ranking offensive category was runs scored (.808), while the highest ranking defensive category was ERA (.780). The next three highest winning percentages are quite interesting -- offensive on-base percentage (.769), its defensive counterpart, opposing OBP (.750), and offensive slugging percentage (.748), all outpace the much more heralded batting average statistics, which finished 6th and 7th.

The strength of the OBP and SLP stats speaks to the importance of the walk and the extra-base hit (and to their prevention), which is not measured by the batting average statistics. This also helps explain the success of the 1993 Phillies, who drew an astronomical number of walks and had a large number of extra-base hits to lead the NL in runs scored. Interestingly, of the 6 stats which are represented on both the offensive and defensive category lists, there are only 2 stats (walks, K's) for which the defensive stat has a higher winning percentage than its offensive counterpart. This amplifies the earlier conclusions reached about the relative contributions of offense and defense.

Just as interesting is the relative unimportance of certain statistics. For instance, offensive strikeouts (.488) has a negative (<.500) correlation with winning. The fielding statistics, errors (.645) and double plays (.531), also rank surprisingly low, though it could be argued that more double plays would tend to be turned by teams that allow more baserunners.

Summary: Analysis of offensive and defensive statistics of NL division winners since 1969 indicates that offensive performance is more contributory to winning than defensive performance. However, it appears that a division winner must have at least adequate pitching to do so, while similar offensive levels are not a prerequisite for winning division titles. The concepts of offensive, defensive, and overall statistical winning percentages (SWP's) are fundamental to this research. Further examination showed that the individual statistics most closely correlated with winning (besides runs and ERA) are offensive and defensive OBP and offensive SLP -- not the more heralded batting average statistics.

Tony Blengino, 413 Brooke Ave., Magnolia, NJ, 08049; (609) 346-2548. An appendix to this article containing detailed information on teams' offensive, defensive, and overall statistical winning percentages is available from the author.

****

DOES CAL RIPKEN TIRE? ANOTHER WAY TO FIND OUT

by Charlie Pavitt

Back in the October 1992 issue of this journal (Vol. 4, No. 3), Harold Brooks reported a study suggesting that infielders who take few off-days during the season may suffer from performance decrements during September/ October; in other words, may tire out. Brooks particularly singled out Cal Ripken as an example, as Ripken never misses a game and showed particularly large performance decrements in the study.

Reading Brooks's article inspired me to dust off some old data I compiled on Ripken relevant to the same general issue. Now it is of course the case that Ripken does get off days, during the approximately 20 days during each season that the Orioles do not play. But between these rests, he can play up to 20 consecutive days, the contract-mandated maximum days a team may be scheduled to play without a rest. My research question is whether Ripken's performance nosedives toward the end of these long stretches. I view this as another way of treating Brooks's central issue of whether playing every day tires position players out.

To answer the research question, I used the six available years of Project Scoresheet data (1984-1989) to compile Ripken's BA, SA, and OBA the day after an off day in the Orioles' schedule, the second day after an off day, and so on up to the longest allowable stretch (20 days after an off day). For example, imagine the Orioles have an off day on August 1. Ripken's performance on August 2 would count as one day after an off day, on August 3 as two days after an off day, and (if the Orioles had played up to the contractual max) on August 21 as 20 days after an off day. I started counting again after every Oriole off day. Thus if August 22 were an off day, then Ripken's performance for August 23 would count as one day after an off day, starting the cycle again. Both games of a doubleheader were included in the same category, so if the Orioles played two on August 5, both would count as four days after an off day.

As you can see from the subsequent table, the longer the stretch since an off day, the less likely the stretch occurs. Thus to keep the sample size reasonably large, I combined some "days since off day" categories. The following shows the results:

Days Since Off Day	G	AB	BA	SA	OBA
1	145	553	.259	.477	.346
2	121	477	.266	.436	.354
3	113	438	.267	.438	.340
4	93	348	.264	.440	.355
5	81	309	.256	.430	.343
6	78	308	.269	425	.336
7-8	111	451	.284	.432	.342
9-10	79	317	.296	.467	.370
11-13	77	301	.269	.445	.345
14-20	73	295	.264	.424	.338

It looks like consecutive days do not faze Cal. However, before combining categories, I noted a possible performance decrement at the end of the last category.

Days Since Off Day	G	AB	BA	SA	OBA
14-16	44	179	.291	.475	.355
17-20	29	116	.224	.345	.313

It appears that Ripken may be tired when it's been 17 or more days since a rest. Again dividing the last category:

Days Since Off Day	G	AB	BA	SA	OBA
17-18	18	74	.270	.419	.341
19-20	11	42	.143	.214	.265

Here the collapse appears specific for the 19th and 20th day, although the sample sizes are far too small to be confident in this conclusion.

If Ripken really is tired after playing so many days in a row, we might ask if an off day at this point rejuvenates him. To answer, I looked at his performance from one to three days after an off day following a stretch of at least 14 days played in a row:

G	AB	BA	SA	OBA
35	137	.234	.401	.291

These numbers look to be considerably worse than Ripken's normal performance from one to three days after an off day (see the top three rows in the first table), more evidence that he is tired, although again the sample size is not very large.

The intent behind my analysis was to be able to add my two cents to the "should-Cal-get-more-rest" battle (I presented this data at a SABR regional meeting inside Memorial Stadium with some Oriole brass in the audience). It suggests that a rest after two weeks straight of playing might do Ripken some good. The method could also be used to examine other players, and see if Brooks's conclusions were also supported using this alternative method.

Charlie Pavitt, 404 Stamford Drive, #2-D, Newark DE, 19711; (302) 733-0468

****

1993: A NEW WAVE OF TRIPLE MILESTONE HITTERS

by Bill Gilbert

As was expected in an expansion year, 1993 was a great season for hitters. However, a comparison with other notable hitting years suggests that a major reason

for the offensive explosion this season is the arrival of a new wave of exceptional young players who can hit for both average and power. Entering the last month of the season, nine players had a good shot at hitting the "triple milestones" of a .300 batting average, 30 home runs and 100 runs batted in. Five of them made it, three missed on batting average and one fell short on home runs. Another player, Mo Vaughn of the Red Sox, put on a September surge and just missed with .297-29-101. (See the table on the next page for the numbers.)

The five players who achieved triple milestones in 1993 were Barry Bonds (3rd time), Frank Thomas (2nd time) and Juan Gonzalez, Ken Griffey, Jr. and Mike Piazza, each for the first time. Piazza was the first rookie to do it since Walt Dropo in 1950. (Interestingly, Dropo never hit any of the three milestones again.) The only other rookies with triple milestones were Wally Berger in 1930 and 20-year old Ted Williams in 1939.

The last time five players achieved triple milestones in a season was 23 years ago in 1970. Four of the five players were in their 30s and the other, Tony Perez, was 28. In 1993, the five players who made it and the five who came close were all in their 20s, and six of them were 25 or less. The last time there were more than five players with triple milestones was the first expansion year, 1961, when there were eight, including Hall-of Famers Willie Mays, Hank Aaron, Mickey Mantle and Frank Robinson, who between them achieved it 22 times.

Eight players also achieved triple milestones in 1953, including three from the same team, Duke Snider, Roy Campanella and Gil Hodges of the Brooklyn Dodgers. However, the real heyday of the triple milestone hitter occurred in 1929-1930 when 10 players did it each year. Babe Ruth, Lou Gehrig, Jimmie Foxx, Al Simmons, Hack Wilson and Chuck Klein, all Hall-of Famers, did it in both years. These six players accomplished it a total of 42 times.

Whether the new wave of hitters will match the accomplishments of their predecessors will not be known for many years. History would suggest that offensive production will fall off a little in 1994 as the leagues adjust to expansion. However, the achievements of the current crop of hitters at such a young age are virtually unprecedented. One question yet to be answered is whether today's young multi-millionaires will be motivated to play long enough to strive for the career achievements of their predecessors.

Bill Gilbert, a member of the Society for American Baseball Research, is a free-lance baseball analyst and writer residing in Baytown, Texas.

****

CLUTCH HITTERS AND OTHER MYTHOLOGICAL ANIMALS

by Willie Runquist

This little ditty is from the "You can't get there from here!" department. The general logic that has been employed in attempts to identify a clutch hitter, or any other special player, is to show that the player's "clutch" performance differs from "ordinary" performance more than would be expected by chance.

The problem with such analyses is usually said to lie in the fact that the power of standard statistical tests is not sufficient to pick up outliers given the samples of plate appearances that we have to deal with. It is also true, however, that the basic logic of these analyses is flawed. Statistical tests of significance are not appropriate vehicles to convey information about specific players regardless of the power of such tests. The following example, will illustrate the problem.

For the sake of argument, a mythical player Hannibal Xerxes has 10,000 career at bats. 1,500 of those at bats have come in "late inning pressure" situations. His career batting average in those situations is .325 while his overall career batting average is .300.

The standard error for a sample of 1,500 at bats with a true average of .300 is .012. If the 1,500 clutch at bats are simply a random sample of his 10,000 at bats, we would expect an average of .325 to occur less than one time in 40. By conventional statistical criteria, we might then assume that Xerxes has met the criteria for being a clutch hitter since his clutch performance differs significantly from what we would expect by chance. His very existence thus justifies the fact of clutch hitters.

Note that there are really two questions involved here: (1) Do clutch hitters exist? (2) Is Xerxes one of them? With appropriate analyses, it may be possible to answer the first question. The second is unanswerable.

What have we learned from Xerxes? Actually very little. Suppose we have clutch hitting data for 1,000 players. Even if clutch hitters do not exist, we would expect to find 25 players whose clutch hitting averages exceed their career average by more than two standard

Triple Milestones - 1993

Player	AGE	August 31 BA	August 31 HR	August 31 RBI	Final 1993 BA	Final 1993 HR	Final 1993 RBI
Did It in 1993.							
Barry Bonds	29	.344	39	101	.336	46	123
Frank Thomas	25	.319	36	109	.317	41	128
Ken Griffey, Jr.	23	.318	39	91	.309	45	109
Mike Piazza	25	.315	28	85	.318	35	112
Juan Gonzalez	23	.313	40	102	.310	46	118
Had a Shot.							
John Olerud	25	.382	23	97	.363	24	107
Rafael Palmeiro	29	.303	33	89	.295	37	105
Albert Belle	27	.298	34	110	.290	38	129
Matt Williams	27	.295	27	86	.294	38	110
Mo Vaughn	25	.305	21	80	.297	29	101
Future Possibilities.							
Carlos Baerga	24				.321	21	114
Jeff Bagwell	25				.320	20	88
Dante Bichette	29				.310	21	89
Travis Fryman	24				.300	22	97
Charlie Hayes	28				.305	25	98
Chris Hoiles	28				.310	29	82
Fred McGriff	29				.291	37	101
Tim Salmon	24				.283	31	95
Larry Walker	26				.265	22	86
Rick Wilkins	26				.303	30	73

Did It in the Past.

	1993 AGE	YEARS			1993 STATS		
Dave Winfield	41	1979			.271	21	76
Eddie Murray	37	1980	1982	1983	.285	27	100
Dale Murphy	37	1983	1985		.143	0	7
Don Mattingly	32	1985	1986	1987	.291	17	86
George Brett	40	1985			.266	19	75
George Bell	33	1986	1987		.217	13	64
Danny Tartabull	30	1987	1991		.250	31	102
Jose Canseco	28	1988			.255	10	46
Ryne Sandberg	33	1990			.309	9	45
Cal Ripken., Jr.	32	1991			.257	24	90
Gary Sheffield	24	1992			.294	20	73

errors. Is Xerxes one of these players, or is he truly a clutch hitter? The fact is we will never know (at least by statistical means).

Statistical criteria operate on the logic that certain events are statistically rare. If we select a player at random from our pool of 1,000, the probability is only .025 that the difference in his average will exceed two standard errors. Therefore, if our single choice results in that rare event, we believe that it is unlikely to be due to random variation. However, if we select Xerxes because we noticed that he seemed to hit better in the clutch, the likelihood that the difference will exceed two standard errors increases substantially. It is not hard to find statistically unlikely events if you go looking for them.

The appropriate test for the *existence* of clutch hitters in general is to divide the difference between the clutch and career averages for each player in the sample by the standard error and compare the entire distribution with that expected by chance. Clutch hitting would be indicated by a statistically significant skew toward the upper end (too many large values).

Suppose we find that the overall distribution differs from that expected by chance[1] and that 40 of the 1,000 players produce differences larger than two standard errors. If Xerxes is a member of this group, we still do not know whether he is there because he is a clutch hitter or because he is one of those whose average varied from the norm by chance. There are a fairly substantial number of the latter.

The issue of power is somewhat of a red herring. If your favorite player did not make the "elite 40" you might argue that the power of the test was not sufficient, but since we cannot identify the true clutch hitters that lie above the criterion, neither can we identify the true "non-clutch" hitters among those who do not make it. Increasing power may increase the number of players above the criterion, but we still do not know which ones really belong there.

Willie Runquist, The Union Bay Oyster Chucker, P.O. Box 289, Union Bay, B.C. V0R-3B0 CANADA

[1] A simple Chi- square test for goodness of fit is the appropriate test, given enough cases at each level. It is not appropriate to only test those who exceed some arbitrary criterion, i.e. two standard errors.

****

FREE AGENCY AND COMPETITIVE BALANCE

by Lawrence Hadley & Elizabeth Gustafson

What has been the impact of free agency on the competitive balance of teams in Major League Baseball (MLB)? Three alternative possibilities have been debated.

The first has long been associated with the owners. They argue the reserve clause is necessary for competitive balance. With free agency, they expect the rich teams to bid the star players away from the poor teams. George Steinbrenner's world champion New York Yankees of 1977-78 are offered as supporting evidence.

A second possibility is associated with the Players' Association as well as many economists. They argue that free agency improves competitive balance because mediocre teams are likely to bid the star players away from the better teams. The economic value of a star player on the free agent market may be greater to a mediocre team. Stars may make a greater incremental contribution to team victories and thus to ticket sales on a mediocre team than on one already loaded with star players.[2]

Third, it is plausible to argue that free agency may have no impact on the allocation of star players between teams. Ronald Coase recently won the Nobel Prize in economics for his theorem that a change in property rights effects only the distribution of wealth but has no impact on the allocation of resources.[3] In MLB, the resources are the players. The allocation of players amongst the teams is determined by the relationship between the players' impacts on team revenues and their cost of being hired. Under free agency, the cost of hiring a player is his salary. Under the reserve clause that same cost is measured in terms of the player's value to the team

[2] This point was first made by Simon Rottenberg, "The Baseball Players' Labor Market," Journal of Political Economy, Vol 64 (June, 1956), pp. 242-258.

[3] Coase, R. H., "The Problem of Social Cost," Journal of Law and Economics, Vol. 3 (October, 1960), pp. 1-44.

if sold or traded to another team.[4] The market is expected to allocate each player to the team where he will generate the most profits regardless of property rights (free agency or the reserve clause).

Certainly free agency revised the property rights of MLB players, and certainly wealth has been shifted from the owners to the players in the form of higher salaries. If Coase is correct, this redistribution of wealth occurs without any impact on the distribution of star players between teams. Therefore no impact on the competitive balance between teams is expected.

These three theories represent all logically possible outcomes. But what has been the actual impact of free agency on competitive balance between the teams in MLB? This is an empirical question that must be resolved with data -- not theories.

Competitive Balance: Definitions: There are two dimensions to competitive balance: turnover in championship teams and dispersion in teams' won/loss records. Turnover measures the variety of teams that win championships while dispersion measures the closeness of all teams in the final team standings. Turnover and dispersion may behave independently of each other. It is possible for the variety of championship teams to increase or decrease while dispersion in won/loss records remains unchanged and vice versa. We measure the dispersion of final team standings with the average deviation (AD) of all teams' winning percents in a single season. In terms of arithmetic:

$$AD = \sum_{i=1}^{n} \frac{|(W_i / G_i) - .500|}{n}$$

where n is the number of teams in MLB, Wi is the total games won by the ith team, and Gi is the total games played by the ith team.

Perfect balance occurs when every team has a .500 record. In this case, AD = 0 because Wi/Gi equals .500 for every team (all teams win half of their games). The more teams' records deviate from .500 (both above and below), the greater the value of AD, and the less competitive balance exhibited for the season.

Competitive Balance: Measurements: Catfish Hunter, Andy Messersmith, and Dave McNally were free agents prior to the 1976 season, but the free agent era actually began with the collective bargaining agreement negotiated in July, 1976. This agreement granted free agency to players with six or more years of experience in MLB who "played out their option year." Free agency had its first significant impacts on MLB's labor market for the 1977 season when 281 players were signed to multiyear contracts, and salaries began their rapid upward path.[5]

Our empirical analysis below examines two eleven-year periods: 1966-76 and 1982-92. The first eleven years are the pre-free agent era and the second eleven are the free agent era. The 1977-1981 seasons are viewed as a transition period, and they are not included in our analysis. In 1966, there were twenty teams in MLB--ten in each league. These included the Orioles, Red Sox, Angels, White Sox, Indians, Tigers, Twins, Yankees, A's, and Senators (to become the Rangers) in the AL, and the Braves, Cubs, Reds, Astros, Dodgers, Mets, Phillies, Pirates, Cardinals, and Giants in the NL. In 1982, MLB had the same 26 teams that finished the 1992 season.

The evidence most often cited to support the view of improved competitive balance after the advent of free agency is the increased variety of teams that have won league championships and the World Series. Indeed, there was greater turnover in the free agent period. From 1966-76, ten teams won league championships, and seven teams won the World Series. From 1982-92, 16 teams won league championships and ten teams won the World Series. On a per team basis, 50 percent won league championships before free agency and 62 percent won after free agency. For the World Series, 35 percent of the teams won in the pre-free agent era while 38 percent won in the free agent era.

If free agency causes players to change teams with greater frequency, turnover in championship teams may be expected to increase. It appears that this has indeed happened. However, this may occur independently of changes in the dispersion of the teams' records.

To examine this dispersion, we have calculated AD for the pre-free agent and free agent eras defined above.

---

[4] For an excellent discussion of the cost of a player to a MLB team, see Robert McCormick, Managerial Economics, Prentice Hall, 1993, pp. 547-548.

[5] Zimbalist, Andrew, Baseball and Billions, Basic Books, 1992, pp. 18-27.

Table 1 (on the next page) presents annual ADs for 1966-76 and 1982-92 and the mean AD for both periods.

We have controlled for expansion teams by only including in our 1966-76 computations the 20 teams that existed in 1966. All games involving a 1969 expansion team have been dropped from our computations.[6] Specifically, we exclude all games that involved the Pilots/Brewers, the Royals, the Expos, or the Padres who were added to MLB in 1969. Over the years 1982-92, no expansion teams were added to MLB so our calculations include the 26 teams of those years. At the start of both eleven-year periods, the newest teams included in our computations had 4 or 5 years of playing experience (four for the Astros and Mets who were added in 1962 and five for the Blue Jays and Mariners who were added in 1977).

The results in Table 1 indicate that the average seasonal dispersion of teams in the pre-free agent era was 53 percentage points in the standings (AD=.053039, and we are counting percents as baseball fans in thousands rather than hundreds). In the free agent era, the average seasonal dispersion was approximately 51 points (AD = .050747). The difference between these two means is 0.002292, or approximately 2 percentage points in the standings.

Two statistical tests were performed to test for significant difference between these two means. The first test assumed the sample mean ADs to be normally distributed and the second assumed no particular distribution for the mean ADs. Neither test showed the difference between the two means to be significantly different from zero at the .10 level of significance. Since this observed 2 percent differential is not significantly different from zero, it may be attributed to random chance.

Conclusions: The statistical evidence in Table 1 supports the hypothesis that free agency has had no impact on the dispersion of teams in the final standings.[7] The increased turnover in personnel associated with free agency may be responsible for the greater turnover in championship teams, but this appears to be independent of any change in the spread of the teams in the final standings. In baseball terms, close pennant races are equally likely in the free agent era and the pre-free agent era.

The Coase theorem provides a logical explanation. Assuming that owners are primarily interested in profits, a player will play on the team where he generates the greatest revenue. Teams in large markets are likely to have a greater number of star players because they have greater potential to generate revenue in those large markets. But at some point, diminishing returns will make it profitable for small-market teams to hire some star players. Free agency does not change the revenue-generating ability of star players in large versus small markets. Therefore, free agency should not impact on the distribution of these star players between teams. It should impact only on the distribution of the revenues between the owners and the players.

If the distribution of star players is unchanged by free agency, then it is not surprising that the dispersion of the teams is also unchanged. In the pre-free agent era, players moved between teams via trades. Now they move via trades and free agency. This additional mechanism for player mobility appears to favor greater turnover in championship teams without having any statistically significant impact on the dispersion of teams in the final season standings.

Lawrence Hadley and Elizabeth Gustafson are Associate Professors of Economics at the University of Dayton, Dayton, OH 45469-2240, (513) 229-2403

---

[6] The data required for these computations are the teams' season records versus each other individual team from 1969-1976. These data were obtained from the American League Red Book, 1970-1977 and the National League Green Book, 1970-1977 which were supplied to us by the front office of the Boston Red Sox.

[7] Using different statistical methods and data from different MLB seasons, Christopher Drahozal reached the same conclusion. See Drahozal, "The Impact of Free Agency on the Distribution of Playing Talent in Major League Baseball," Journal of Economics and Business, Vol 38 (1986), pp. 113-121.

TABLE 1

AVERAGE DEVIATIONS FOR MAJOR LEAGUE BASEBALL:

PRE-FREE AGENT AND FREE AGENT ERAS

Pre-Free Agent Era		Free Agent Era	
Year	AD	Year	AD
1966	.051804	1982	.053181
1967	.053506	1983	.049383
1968	.045411	1984	.041611
1969	.055883	1985	.062363
1970	.060777	1986	.051840
1971	.048097	1987	.048932
1972	.054578	1988	.056574
1973	.044393	1989	.049469
1974	.050047	1990	.045599
1975	.059619	1991	.045612
1976	.059310	1992	.053656
Means	.053039		.050747