baseball ANALYST

APRIL 1983 ISSUE # 6

IN THIS ISSUE!!!

YOUR SUBSCRIPTION IS UP!!

(If you were a charter subscriber)

$12.00 A YEAR (ONLY)

RE-PRINT

BALLPARK EFFECTS ON FIELDING STATISTICS- AMERICAN LEAGUE PARKS

This article summarizes the American League data for the same years as the National League data presented in the earlier article. Baseball Analyst, Issue No.1, June 1982. Again, the purpose of the study was to examine what effects natural as opposed to astroturf surfaces have on the production of errors and double plays.

As in the earlier National League study, the outstanding conlusion is that astroturf significantly lowers the amount of errors committed.

The years for the study are the same as for the National League study. For 1973, 1976, 1979 and 1980, the statistics include full team/opponent performance at the team home park and on the road. For the 1972 and 1978 seasons only "team" home/road performances were compiled.

The raw statistical tables below show, for example, errors by infielders (E_{if}), errors by outfielders (E_{of}) and double plays (DP) in games between Baltimore and their opponents in Baltimore (Games at Home) and at the opponent's home park (Games on Road). This permits a comparison between the percentage of occurrences of double-plays, infield and outfield errors in games between Baltimore and their opponents in games at Baltimore as opposed to games at the opponents' parks.

TABLE I: RAW STATISTICS

Team	Games at Home			(per 1000) plays, excluding strikeouts)	Games on Road		
	E_{if}	E_{of}	DP		E_{if}	E_{of}	DP
BAL	25.50	6.16	45.69		23.31	6.13	47.59
BOS	26.57	4.72	50.72		29.59	6.83	45.51
CLE	26.27	6.14	46.46		26.35	7.04	45.61
DET	28.39	6.97	47.30		27.55	5.79	48.11
MIL	29.65	6.78	42.33		27.08	6.34	42.78
NY	22.73	6.23	43.14		27.97	6.85	45.27
@-TOR	24.26	5.11	43.74		29.78	7.56	50.65
CAL	30.72	7.04	44.80		26.44	7.33	45.26
CHI	29.17	7.67	42.60		27.26	8.37	42.32
@-CHI	27.40	7.44	48.15		31.68	4.55	45.73
@-KC	27.02	6.30	42.04		32.94	7.40	43.80
MINN	32.12	7.90	49.13		29.59	6.97	45.75
OAK	31.54	7.06	41.46		29.65	5.77	38.19
@-SEA	26.36	5.18	48.08		28.40	7.85	47.08
TEX	34.45	6.04	46.36		29.15	6.57	43.36
n-KC	28.00	5.25	46.09		19.03	7.98	52.18

Before proceeding further, a few remarks on the fields. The Seattle and Toronto data is limited to 2 and one-half seasons, the partial 1978 and full 1979 and 1980 data, for the obvious reason that the teams came into existence in 1977. I set apart the Kansas City old Municipal stadium data (1972) apart from the other data because I consider the one-half season (Royals home and away) data insufficient for a meaningful analysis. I include it as it shows the general trend but I believe that it is probably unfair to rank Municipal Stadium relative to the others on the basis of such limited data. As for the CHI and CHI-@, if memory serves me correctly, Bill Veeck removed the carpet from Comiskey Park before 1976. I believe that the rug was there from 1971-75. Therefore, the 1972-73 data is for Comiskey as an astroturf park. (If anyone knows the exact dates the Sox played at home on astroturf, please correct me if I'm wrong.)

On to the comparative data. For obvious reasons it is the comparative data which indicates the differences, if any, caused by the parks. If the Yankees, for instance (as Yankee home games show the smallest occurrence of infield errors per 1000 plays, excluding strikeouts), were a good fielding team during this period, there will likely have been fewer errors committed in games involving the Yankees. What we are concerned with is whether more errors were committed in games at Yankee Stadium than on the road. ("Good fielding," in this context refers only to the error-production component of overall fielding ability.)

In the comparative data tables I have simply divided the Games at Home by Games on Road components in order to obtain a ratio of occurrences home/away. The categories are ranked in order of ballpark.

TABLE II: COMPARATIVE DATA

Rank	Park	E_{if}	Park	E_{of}	Park	DP
1	NY	.813	@-SEA	.660	BOS	1.114
2	@-TOR	.815	@-TOR	.676	OAK	1.086
3	@-KC	.820	BOS	.691	MIN	1.074
4	@-CHI	.865	@-KC	.851	TEX	1.069
5	BOS	.898	CLE	.872	@-CHI	1.053
6	@-SEA	.928	NY	.909	@-SEA	1.021
7	CLE	.997	CHI	.916	CLE	1.019
8	DET	1.030	TEX	.919	CHI	1.007
9	OAK	1.064	CAL	.960	CAL	.990
10	CHI	1.070	BAL	1.005	MIL	.989
11	BAL	1.085	MIL	1.069	DET	.983
12	MIN	1.086	MIN	1.133	@-KC	.960
13	MIL	1.095	DET	1.204	BAL	.960
14	CAL	1.162	OAK	1.224	NY	.953
15	TEX	1.182	@-CHI	1.635	@-TOR	.864
	KC	1.471	KC	.658	KC	.883
Ave.:	NAT	1.044	NAT	.991	NAT	1.022
	AST	.857	AST	.955	AST	.975

First, a few remarks on the tables. The averages referred to are simple numerical averages of the comparative ratios. Because of the unbalanced numbers of years for certain teams the astroturf and natural turf averages together will not equal one.

As in the National League study, the evidence is overwhelming that astroturf reduces infield errors by a significant margin. The difference in numerical average in NL parks was about 14%; the difference in American League parks is about 19%. If anything the American League data is more overwhelming, the four astroparks are among the top six fielding parks, only Fenway and Yankee stadiums intrude into the astroturf domain. Not surprisingly, the top natural park field in the NL, Los Angeles' Dodger Stadium is also an exclusively baseball park.

Again, there is again a consistency with the NL results in showing fewer outfield errors on astroturf. In an interesting critique of the NL study, Craig Wright, Texas Rangers Sabermetrician, was surprised at my reluctance to draw a conclusion about the astroturf effect on outfield errors. I surrender! The consistency of the figures is unmistakeable. I was reluctant to draw a conclusion as to outfield errors mainly because of the smaller amount of data, there are far fewer outfield errors committed and it is, therefore, more difficult to establish trends. Wright's comments and further reflection have persuaded me that the smaller numbers may affect the ability to make comparative conclusions as to individual ballparks but that the overall astroturf/natural turf trend exhibited in the numbers is probably valid. It is, as Wright points out, consistent with common sense.

I also appreciate Wright's supplemental information concerning double plays based on double play opportunities. Wright mentions two factors which could affect the fequency of double plays per double play opportunities on astroturf and natural turf: (1) prevalence of ground ball on different surfaces; (2) (by implication) speed of the ball's movement on different surfaces.

When Wright discussed "prevalence" of ground balls on surfaces, I believe he refers to the prevalence of ground ball outs rather than just ground balls, because the slower movement of a ground ball on most natural surfaces means that an infielder of average range will have a better chance of getting to a ball laterally and can play closer to the plate, resulting on shorter throws (most important at third base and shortstop). He suggests that if ground ball outs occur more frequently on natural turf, and that if overall the frequency of double plays on the different surfaces are about equal, as they appear to be, that the double play/opportunity percentage may be higher on astroturf. This suggestion again has some consistency with common sense, although better data would be necessary to discern such a trend. Maybe with the advent of videotapes and home computers such data may soon come available, at least for those teams which televise a lot of games.

Finally, some remarks about the AL data. Relevant factors in the awarding of Gold Gloves were discussed in the NL article. Those factors include fielding percentage, range, hitting, and overall subjective impressions. Certainly another factor is habit. Examining the fielding difficulty of American League parks relative to fielding percentage, the achievements of Brooks Robinson and Mark Belanger are, in retrospect, more impressive.

Based on the number of Gold Gloves awarded in Baltimore over the past 15 or 20 years I had assumed that Baltimore would at least be one the better natural surfaces. At best, it is an average natural surface. Another achievement that stands out is Buddy Bell's recent popularity at the Gold Glove ballot box, playing on a surface that, statistically at least, looks abysmal. I had always thought that Toby Harrah's horrible fielding statistics in Texas had to be in part caused by the playing surface. While Harrah's fielding statistics have improved somewhat in Cleveland, Bell has continued to run up impressive numbers in Texas.

As for future prognostication, players such as Damaso Garcia and/or Griffin and Fernandez (whoever they settle on) in Toronto, the Cruz boys in Seattle, U. L. Washington are astroturf-type infielders playing on astroturf who stand a good chance of winning Gold Gloves. Alan Trammell is a throwback to the Lou Boudreau-type infielder with adequate range on grass, but certainly no Ozzie Smith when it comes to covering ground. Gold Gloves have been rare in the Detroit infield. Robin Yount worked hard to become a better than average shortstop but mainly won the award on the strength of an MVP hitting season, as is often the case when the choice is a close one. Glenn Hoffman and Willie Randolph are good infielders playing on good surfaces who should win a Gold Glove at some time. Playing in New York certainly helped Graig Nettles and Bucky Dent achieve notoreity (deserved or not) for fielding prowess.

In conclusion, ballpark fielding effects should be put in their proper perspective. The outstanding fielder will probably manage to amass, more often than not, impressive fielding statistics even on the worst of fields. It is important, though, to remember that if your favorite shortstop committed 30 or more errors last season, that does not necessarily mean he's a bad fielder, or even that he necessarily has bad hands. That difference of five to ten errors a year that makes an infielder look like a bum may be more due to the conditions than the quality of his play or his ability. Owners and management should likewise be aware of ballpark effects in evaluating the fielding abilities of their players and should be aware of the necessity of selecting players who can thrive on the home playing surface, where, after all, half of the games are played.

Quality versus Quantity

Often asking the correct question is the hardest part of doing a study. One of the questions involving baseball that I like to think about is 'What is 'better' "? This is as in "Ted Williams was the best hitter of all time" or "Pete Rose is a better hitter than Richie Ashburn." What in the world can this mean, if one is trying to be objective?

As a mathematician, I was taught that trying to prove things without axioms to prove them from is futile; what is a proof in one axiom set is garbage in another. Unfortunately, baseball has no such sets of axioms; fans and baseball people make them up and then prove things which others, who have a different set of axioms, will naturally disagree.

One of the axioms in discussions of 'better' is quantity versus quality. This can be a touchy and difficult subject. For instance, it is obvious that in Hall of Fame ballotting, quantity is very important. Players such as Koufax and Dizzy Dean are the exceptions rather than the rule. Compare Richie Ashburn's statistics at retirement to Pete Rose's at the same age (both were born near the start of the season, so Rose's stats go through the 1976 season):

	G	AB	R	H	2B	3B	HR	RBI	BA	SL%	BB	OB%	SB
Ashburn	2189	8365	1322	2574	317	109	29	586	.308	.382	1198	.394	234
Rose	2184	8886	1459	2762	483	101	134	838	.311	.433	957	.378	106

What leaps out at me is that Rose has more power, Ashburn more speed. On items related to how good the team is (At Bats per game, runs and RBI's) Rose far exceeds Ashburn. But in unrelated items, the power versus speed is more noticable. And, of course, Ashburn gets on base more often due to walks, not unimportant for a leadoff hitter. Yet Ashburn was recently passed by for 15 years by the writer's committee, while Rose was a first ballot shoe-in long ago. The reasons are many, but one seems to stand out above all the rest: Ashburn didn't play long enough to establish himself as an 'institution', or one of the 'grand old men' of the game.

So, in trying to answer my question, lets lay down an axiom (although a subjective one at best:) Suppose you were a general manager and you had a crop of rookies to chose from in a pool. IN ADDITION, you had in front of you the ENTIRE CAREER RECORD of each player BEFORE THEY WERE TO PLAY THEIR FIRST MAJOR LEAGUE GAME. Further suppose that you have no other factors, such as team need to use as a selection factor. Then AXIOM #1 is "The better ballplayer is the one a genius general manager would choose in such a situation."

Now this is an intriguing axiom; it pits quantity versus quality in some perspective. For instance, it is clear that one would want to pick Rose over Ashburn, for why not get someone who will produce similarly over 20+ years like Rose rather than only 15. Here, as in Hall of Fame selections, quantity becomes a legitimate comparison factor.

Would you have not have rated Stargell or Phil Niekro or Gaylord Perry lower if they had retired after the 1978 season-at ages that would not have been embarressing? Yet Ashburn never even saw those ages in his career! Clearly quantity is a large factor in baseball comparisons. As that general manager, you want to get the most out of each player, and longevity is a factor. Ralph Kiner hit more home runs per at bat than Hank Aaron, therefore, one could say he was the greater home run hitter; but surely there is value in hitting 755 home runs over 369; many of those extra home runs by Aaron will win ball games that Kiner's replacement couldn't.

There are other problems to consider even with our simple axiom; Rose's team oriented statistics are an example. There is a problem with the axiom; the statistics are inherently biased. If one leaves in the team oriented values, then one knows how the rest of his 'to be picked' fictitious team will do. For example, let's say one of your rookies is Joe Outfielder, whose lifetime stats are to be a .300 On Base percentage with all singles and walks (no isolated power, no steals) but his final statistics show him averaging 200 runs a year! (This is an example of an 'ad absurdum argument, of course). Should you pick him?? Similarly, but on a lesser scale, Rose's extra At Bats per game, even including Ashburn's walks indicates that your team will be better. So, maybe axiom #1 ought to be modified to not include these factors, such as runs and RBI's and not both games and at bats.

Anyway, now that one has established some axiom, there is much more enjoyment in comparing ballplayers. Of course, the next problem is that players of different eras have different absolute statistics, so maybe we shall limit the information to output per year divided by the league average that year in that statistic. For example, a home run champion in the dead ball era would be given his relative worth, rather than his absolute worth. Now we are getting into an area for which I am going to right another paper about another axiom, which involves comparison of players over eras, but that is another full subject. For now, let us assume the statistics were compiled in the same rough time eras.

Just for fun, lets take pairs of post World War II ballplayers and I'll let you be the general manager. Who would you pick:

Brooks Robinson or Ron Santo; Luis Aparicio or Phil Rizzuto; Carl Yastrzemski or Mickey Mantle
Whitey Ford or Gaylord Perry; Dick Radatz or Ron Reed; Dick Groat or Larry Bowa;
Phil Niekro or Sandy Koufax; Harmon Killebrew or Ralph Kiner; Mark Fidrych or Roger Craig
Don Drysdale or Luis Tiant; Bill Freehan or Roy Campanella; Joe Black or Enrique Romo
Roger Maris or John Mayberry; Duke Snider or Al Kaline; Vernon Law or Doc Medich; Ernie Broglio or Jon Matlack; Catfish Hunter or Don Sutton; Bill Robinson or Wes Covington.

The argument goes on, but at least there are some criteria.

Continuing the question of quality, is it quality per At Bat that is most important, how about per game, per year or per career (sic)? Obviously as one lengthens the time in the denominator, [illegible] oes more towards quantity and less towards quality. Just for fun I took some intermediate time (five years) and compiled an interesting set of statistics of a player's best 5 years, not having to be consecutive. I am sure this list tells one something about quality- and maybe even a little about quality. I hope I have added a small amount of knowledge to this subject. - Dan Heisman

DAN HEISMAN 4/1/8_

BEST 5 YEARS
(NOT CONSECUTIVE)

Name	AB	H	HR	RBI	SB	BA	YRS
Gehrig	2942	1064	232	850	51	.362	34 30 27 36 ~~28~~ 31
Foxx	2870	1000	234	806	30	.348	38 36 32 33 30
DiMaggio	2726	944	176	706	12	.346	48 39 40 37 41
Delahanty	2644	1048	56	646	171	.396	93 96 97 95 99
Cobb	2868	1124	34	550	371	.392	11 17 12 15 09
Clemente	2829	968	101	465	28	.342	66 64 67 61 70
Carew	2935	1052	46	387	186	.358	73-77
Ashburn	2876	975	13	235	117	.339	51 58 53 55 48
Aaron	2962	975	219	631	56	.329	59 71 62 63 57
Wagner	2712	979	31	537	243	.361	08 05 01 00 03
Waner	2998	1112	42	478	44	.371	27 34 28 36 30
Williams	2473	916	179	630	10	.370	41 42 49 57 48
Wilson	2759	914	177	707	33	.331	26-30
YASTRZEMSKI	2876	877	166	504	67	.305	67 63 70 69 77
LAJOIE	2840	1078	32	523	122	.380	01 97 10 04 06
KLEIN	3114	1118	180	693	51	.359	29-33
KEELER	2808	1107	11	339	251	.394	95-99
J. JACKSON	2833	1076	32	445	135	.380	11 16 12 13 20
HORNSBY	2857	1144	167	664	42	.400	22 29 24 25 21
HEILMANN	2774	1071	74	620	40	.386	21 27 23 25 24
GREENBERG	2935	974	201	788	34	.332	35 40 37 38 34
SPEAKER	2681	1032	48	505	112	.385	23 20 16 12 25
SISLER	3062	1154	55	533	166	.377	20 25 21 22 27

Name	AB	H	HR	RBI	SB	BA	YRS
Simmons	2712	1035	131	687	33	.382	25 30 27 31 29
RUTH	2594	972	260	766	60	.375	20 31 21 23 27
ROSE	3226	1081	45	317	49	.335	68 73 69 79 76
F. ROBINSON	2809	900	203	607	75	.320	66 61 62 57 69
OTT	2748	906	173	673	27	.330	32 29 34 36 30
MUSIAL	3042	1090	136	546	30	.358	48 49 46 51 43
MAYS	2924	954	222	586	122	.326	54 65 55 58 62
MANTLE	2505	823	217	560	62	.329	56 57 61 58 64

TEAM WON/LOST PERCENTAGE AS A FUNCTION OF RUNS AND OPPONENTS' RUNS

by Dallas Adams

In a recent paper* this author presented data on the probability of scoring an exact number of runs in a game. One application of this data is to find a formula expressing team Won/Lost percentage as a function of the number of runs the team scores on offense and the number of runs it allows on defense.

The referenced data was compiled from actual major league scores for the 1967 through 1976 seasons, a total of 232 team-seasons, and is reproduced below in Table 1.

To review briefly the data reduction procedure: working season by season, each team was grouped with others which showed a similar runs scored per game average. There were 11 groups in all; the characteristics of each group are included in Table 1. For all the teams in a given group, the number of times each team scored each specific number of runs was tabulated. Once this was completed for all teams, the numbers within each group were converted to percentages. That is, for each group the percentage of times it was shutout was computed, likewise the percentage of games in which it scored exactly one run, etc.

The percentages of Table 1 can be converted to probabilities simply by dividing by 100. For a given group (call it Group A) then, the probabilities of that group making a game score of exactly no runs, exactly one run, etc. are known. Similarly, for any second group, Group B, which is considered to be the opposing team, the probabilities of this opposing team being shutout, held to exactly one run, etc. are also known. Thus it is possible to compute the theoretical winning percentage when any given group plays against any other group.

If P_k is the probability of Group A scoring exactly k runs in a game and Q_j is the probability of Group B scoring exactly j runs in a game, then the probability of Group A scoring more runs in the game than Group B is:

$$W = \sum_{k=0}^{18} \left[P_{k+1} \sum_{j=0}^{k} Q_j \right]$$

Similarly, the probability of Group A scoring fewer runs in the game than Group B is:

$$L = \sum_{j=0}^{18} \left[Q_{j+1} \sum_{k=0}^{j} P_k \right]$$

There remains a certain portion of games in which both groups score the same number of runs. It is assumed that each group goes on to win, in extra innings, the same proportion of those "tie games" as it won of the non-tied games. Thus, ties can be ignored and the won/lost percentage of Group A when playing against Group B is W/(W+L).

Since the average number of runs scored per game by each group is known from the original data, each group can then be"played" against every other group with the result that a table, see Table 2, of theoretical won/lost percentage (PCT), runs per game (R/G), and opponents' runs per game (OR/G) can be constructed.

By means of least squares regression analysis, some equations giving PCT as a function of R/G and OR/G have been obtained. Three of these

* The Distribution of Runs Scored

TABLE 1

GROUP CHARACTERISTICS AND THE PERCENTAGE OF GAMES IN WHICH EXACTLY THE GIVEN NUMBER OF RUNS ARE SCORED

EXACT NUMBER OF RUNS	GROUP 1	GROUP 2	GROUP 3	GROUP 4	GROUP 5	GROUP 6	GROUP 7	GROUP 8	GROUP 9	GROUP 10	GROUP 11
0	14.52	11.42	9.79	9.98	8.11	7.21	6.49	5.93	4.75	4.01	4.94
1	18.91	16.90	15.35	13.72	12.90	11.54	10.32	10.00	7.60	7.41	8.02
2	17.03	17.64	17.79	16.28	14.53	14.29	13.37	12.10	12.11	11.42	6.17
3	16.41	15.81	16.79	15.13	15.73	14.69	14.58	13.34	14.68	13.58	11.11
4	10.97	12.22	12.99	12.28	13.78	13.93	13.39	13.11	13.21	11.11	17.90
5	9.09	9.41	9.22	10.61	11.05	11.09	11.26	11.66	11.78	12.04	12.35
6	5.02	6.13	5.72	8.00	7.82	8.40	9.15	9.66	9.64	10.80	9.26
7	2.72	4.95	4.72	5.12	5.76	6.34	7.00	7.69	7.55	8.95	6.17
8	1.78	2.15	3.53	3.85	3.72	4.02	5.04	5.49	5.65	5.86	4.94
9	1.36	1.53	1.68	1.94	2.54	3.14	3.34	3.68	3.52	3.40	7.41
10	1.25	0.83	0.92	1.27	1.58	2.21	2.21	2.36	3.52	2.78	2.47
11	0.31	0.53	0.81	0.86	0.95	1.25	1.43	1.81	2.19	3.70	3.70
12	0.31	0.22	0.33	0.36	0.58	0.68	1.15	1.32	1.24	1.85	1.85
13	0.21	0.18	0.16	0.24	0.35	0.48	0.53	0.73	0.71	1.54	2.47
14	0.00	0.00	0.08	0.17	0.25	0.23	0.31	0.54	0.95	0.31	1.24
15	0.00	0.04	0.08	0.05	0.11	0.20	0.18	0.21	0.43	0.31	0.00
16	0.10	0.04	0.03	0.02	0.13	0.09	0.05	0.18	0.05	0.62	0.00
17	0.00	0.00	0.00	0.02	0.05	0.11	0.04	0.08	0.10	0.00	0.00
18	0.00	0.00	0.00	0.05	0.01	0.05	0.05	0.08	0.19	0.31	0.00
19 & more	0.00	0.00	0.00	0.05	0.04	0.05	0.09	0.05	0.14	0.00	0.00
RANGE OF R/G	2.75-2.99	3.00-3.24	3.25-3.49	3.50-3.74	3.75-3.99	4.00-4.24	4.25-4.49	4.50-4.74	4.75-4.99	5.00-5.24	5.25-5.49
NO. OF TEAMS	6	14	23	26	49	40	34	24	13	2	1
AVERAGE R/G	2.912	3.183	3.371	3.612	3.864	4.122	4.348	4.599	4.867	5.157	5.290

equations are listed:

(1) $\mathrm{PCT} = A + B\left[\log_e(R/G)\right]$

where $A = .034116(OR/G)^2 - .3854(OR/G) + .8984$

$B = -.01809(OR/G)^2 + .1497(OR/G) + .1233$

(2) $\mathrm{PCT} = .5 + (R/G - OR/G)\left[.1465 - .005389(R/G) - .006996(OR/G)\right]$

(3) $\mathrm{PCT} = .5 + .4917\left[\log_e(R/OR)\right]$

where R/OR is R/G divided by OR/G

As a check of the accuracy of these equations, eight seasons (one season from each decade since 1901) were selected at random and the formulas were used to calculate the theoretical won/lost percentage for each team. These theoretical percentages were then compared with actual percentages* and standard errors of .025, .026, and .027 were computed for equations 1, 2, and 3, respectively. Thus, equations 1 and 2 are very nearly of equivalent accuracy. Equation 3 is less accurate but compensates by having the advantage of being simpler and faster to use.

TABLE 2

THEORETICAL WON/LOST PERCENTAGES

RUNS PER GAME ↓	OPPONENTS RUNS PER GAME										
	2.912	3.183	3.371	3.612	3.864	4.122	4.348	4.599	4.867	5.157	5.290
2.912	.5000	.4584	.4356	.4099	.3810	.3578	.3334	.3131	.2884	.2678	.2569
3.183	.5416	.5000	.4764	.4494	.4200	.3938	.3696	.3477	.3222	.2997	.2875
3.371	.5644	.5236	.5000	.4719	.4419	.4152	.3905	.3680	.3420	.3189	.3050
3.612	.5901	.5506	.5281	.5000	.4711	.4443	.4194	.3963	.3708	.3466	.3334
3.864	.6190	.5800	.5581	.5289	.5000	.4726	.4472	.4232	.3977	.3724	.3578
4.122	.6422	.6062	.5848	.5557	.5274	.5000	.4747	.4503	.4250	.3990	.3842
4.348	.6666	.6304	.6095	.5806	.5528	.5253	.5000	.4755	.4502	.4237	.4083
4.599	.6869	.6523	.6320	.6037	.5768	.5497	.5245	.5000	.4753	.4485	.4334
4.867	.7116	.6778	.6580	.6292	.6023	.5750	.5498	.5247	.5000	.4726	.4564
5.157	.7322	.7003	.6811	.6534	.6276	.6010	.5763	.5515	.5274	.5000	.4845
5.290	.7431	.7125	.6950	.6666	.6422	.6158	.5917	.5666	.5436	.5155	.5000

*The seasons used were 1904 NL, 1916 AL, 1927 AL, 1931 NL, 1948 NL, 1955 AL, 1966 AL, 1976 NL

ADJUSTED HOME PARK FACTOR

- Pete Palmer

Apparent home park factor (HFA) can be measured as follows:

$$HFA = \frac{\text{(Runs scored and allowed)/Game at home}}{\text{(Runs scored and allowed)/Game on road}}$$

However, there are four additional factors which modify this result. First, it is runs per inning that is really desired, not runs per game. Innings can be reliably approximated using available information. Total innings pitched (IP) can be obtained from club pitching records. Total innings batted (IB) can be calculated from club batting statistics. Take total batters faced pitcher (at bats plus sacrifices plus walks, hit batsmen and interfences) less runs scored and runners left on base, all divided by three. As a sidelight, the number of team wins can be found accurately by taking decisions over two plus innings pitched minus innings batted. This is simply because the home team does not bat in the last inning unless the winning run is scored in that inning. This basic fact must be considered in apportioning the innings batted and pitched between home and road. Two more figures are needed, home wins (HW) and road losses (RL). With this, innings batted at home (IBH) and on road (IBR) can be found, along with innings pitched at home (IPH) and on road (IPR). Partial last innings are negligible.

$$IPH = (IP + RL)\text{Games-home/Games-total} \qquad IPR = IP - IPH$$

$$IBR = (IP + HW)\text{Games-road/Games-total} \qquad IBH = IB - IBR$$

A team with home wins equal to road losses will have no correction here, but a typical correction will be on the order of one percent. So the improved home factor (HFI) is calculated as:

$$HFI = \frac{\text{(Runs scored and allowed)/Inning at home}}{\text{(Runs scored and allowed)/Inning on road}}$$

However, true home factor (HFT) includes another correction, touched on by Jim Reuter in the February 1983 Analyst. Since the road park average rating for any team does not include the teams's own home park, it is actually slightly to the other side of average from the home factor. This is defined as true road factor (RFT). Jim's equation, which is correct, relates to the number of teams (NT) in the league.

$$HFT/NT + RFT(NT-1)/NT = 1.00$$

This means that the average home park factors for all parks is equal to one. What Jim missed was that initial input was HFI, not HFT. By definition:

$$HFI = HFT/RFT$$

These last two equations can be solved for HFT as a function of HFI, and then RFT.

$$HFT = \frac{NT \cdot HFI}{NT - 1 + HFI} \qquad RFT = \frac{HFT}{HFI}$$

This correction increases as HFI departs from average, and can be over two percent. Since the actual proportion of appearances at home and on the road by individuals is also affected by the fact that the home team doesn't bat in the last inning in games that it won, the individual batting factor (BF) and pitching factor (PF) must be modified.

$$BF = HFT \cdot IBH/IB + RFT \cdot IBR/IB$$

$$PF = HFT \cdot IPH/IP + RFT \cdot IPR/IP$$

This correction is usually less than one percent. To be perfectly accurate, individual day-by-day records could be broken down into home and away categories, rather than assuming that the player breakdown was the same as the team.

A final correction, which is not related to home park but is related to individual performance is that batters on a particular team do not have to face pitchers on that team. To begin this calculation, team bat rating (TBR) and team pitch rating (TPR) are found, neglecting the fact that they don't have to face each other.

$$TBR = \frac{\text{Runs scored-team/IB-team}}{BF \cdot \text{Runs-league/innings-league}}$$

$$TPR = \frac{\text{Runs allowed-team/IP-team}}{PF \cdot \text{Runs-league/innings-league}}$$

Next, this fact is included. Since each rating depends on the other, these equations can be repeated, using as input the output of the previous iteration. However, two or three iterations should be sufficient.

$$TBR' = TBR \cdot \frac{(NT - 1)}{(NT - TPR')} \quad \text{set TPR' to TPR in first iteration}$$

$$TPR' = TPR \cdot \frac{(NT - 1)}{(NT - TBR')} \quad \text{set TBR' to TBR in first iteration}$$

This factor also becomes larger for team with ratings removed from average, and can be in the one to two percent range. These factors can be combined with the batting and pitching factors to give and overall factor.

$$OBF = BF \cdot \frac{(NT - TPR')}{(NT - 1)}$$

$$OPF = PF \cdot \frac{(NT - TBR')}{(NT - 1)}$$

In my original research, I calculated home park factors by using runs scored and allowed. However, I have since decided that it is fairer to use just runs allowed. This is because the home team may be made up of batters who can take advantage of the particular park and thus make it appear to be a better hitting park than it should be. This could also be true of pitchers, but I don't believe that pitchers are anywhere near as apt to be chosen in that manner. The equations in this article still apply, except that in the HFA and HFI equations, these factors are divided by average home/road figures.. Looking at the results for all teams from 1977 through 1982, there are only two cases in which the home park factor using only runs allowed differs from the overall park factor by more than four percent. Kansas City shows .97 with pitchers only, 1.02 counting both, while Houston shows .80 with runs allowed, .86 with both. In 1981, Houston allowed only 106 runs in 51 home games, the lowest in major league history. Strangely enough, the best hitting park in each league shows about the same park factor either way, indicating that neither team has succeeded in tailoring its lineup. Fenway Park was 1.17 either way, while Atlanta was 1.16 for the opponents, 1.17 overall.

The standard deviation for a six-year park factor can be found using a reasonable approximation for baseball scoring. If runs were scored singly like goals in a hockey game, the standard deviation would be equal to the square root of the number of runs involved. However, since in baseball, runs can come in groups, it turns out that every actual distribution of runs I have seen has a standard deviation equal to the square root of twice the number of runs involved. These six-year averages involve approximately 4000 runs at home and 4000 on the road, so the standard deviation is SQRT(2·8000) or 126 over 4000 or about 3 percent. Using just runs allowed, you get SQRT(2·4000) or 89 over 2000 or about 4.5 percent. Thus the error is increased by using a smaller sample, but the mean value should be more accurate. Even the large differences at Kansas City and Houston are not unexpected considered the size of the standard deviation. The fact that runs have a standard deviation equal to twice the square root of the number of runs can be substantiated using table two from Adams' Distribution of Runs Scored in issue number one of the Analyst.

For an example, take Atlanta in 1977. The figures were:

at home - 81 games, 40 wins, 416 runs scored, 488 runs allowed
on road - 81 games, 60 losses, 262 runs scored, 407 runs allowed
batters faced pitcher - (5534 ab + 83 sh + 34 sf + 537 bb + 17 hb + 1 int)
innings batted - (6206 bfp - 678 r - 1139 lob)/3 = 1463
innings pitched - 1445

$HFA = \dfrac{904/81}{669/81} = 1.351$ This would lead to an apparent bat factor of (1.135 + 1)/2 or 1.18

$IPH = (1445 + 60)/2 = 752$ $IPR = 1445 - 752 = 693$
$IBR = (1463 + 40)/2 = 751$ $IBH = 1463 - 752 = 712$

$HFI = \dfrac{904/(752 + 712)}{669/(693 + 751)} = 1.333$

$HFT = \dfrac{12 \cdot 1.333}{11 + 1.333} = 1.297$ $RFT = 1.297/1.333 = .973$

$BF = 1.297 \cdot 712/1463 + .973 \cdot 751/1463 = 1.131$
$PF = 1.297 \cdot 752/1445 + .973 \cdot 693/1445 = 1.142$

Runs-league/innings-league = 8556/17515 = .488

$TBR = \dfrac{678/1463}{1.131 \cdot .488} = .840$ $TPR = \dfrac{895/1445}{1.142 \cdot .488} = 1.111$

$TBR' = .840 \cdot 11/(12-1.111) = .849$
$TPR' = 1.111 \cdot 11/(12- .840) = 1.095$

$TBR' = .840 \cdot 11/(12-1.095) = .847$
$TPR' = 1.111 \cdot 11/(12- .849) = 1.096$

$OBF = 1.131 \cdot (12-1.096)/11 = 1.121$
$OPF = 1.142 \cdot (12- .847)/11 = 1.158$

Thus the overall bat factor is 1.12, quite a bit lower than the apparent 1.18

In Jim's example, an HFI of 1.20 would lead to an HFT of 1.180, RFT of .983 and a BF of 1.082, assuming innings batted at home and away were the same. Assuming a typical home win percentage of .535, the figure is reduced to 1.079.

Using just runs allowed, the overall league data is needed.
4135 runs allowed at home, 4421 runs allowed on the road
9031 innings pitched at home (17515 innings + 547 road losses)/2
8484 innings pitched on the road (17515-9031)

league average $\dfrac{\text{runs allowed per inning at home}}{\text{runs allowed per inning on road}} = \dfrac{4135/9031}{4421/8484} = .879$

$HFI = \dfrac{488/752}{407/693 \cdot .879} = 1.257$

HFT = 1.231 RFT = .979 BF = 1.102 PF = 1.110
TBR = .862 TPR = 1.143 TBR' = .873 TPR' = 1.130

OBF = 1.087 OPF = 1.124

Thus the fact that the Braves' batters were better able to take advantage of their home park than most teams does not penalize them. The final park and staff adjustment is only half the apparent park factor.

Here is data for all major league parks covering the 1977-82 period.

	RUNS										HOME RUNS					
	RUNS ALLOWED ONLY					RUNS SCORED+ALLOWED					ALLOWED			HIT+ALLOWED		
team	HFT	TBR	TPR	OBF	OPF	HFT	TBR	TPR	OBF	OPF	HFT	TBR	TPR	HFT	TBR	TPR
Balt	.88	1.09	.96	.95	.94	.92	1.06	.94	.97	.96	1.06	1.16	.94	.98	1.20	.98
Bos	1.17	1.03	.93	1.08	1.08	1.17	1.03	.93	1.08	1.08	1.15	1.18	1.00	1.19	1.16	.98
Cal	.96	1.06	1.02	.98	.98	.95	1.06	1.02	.98	.97	.97	1.05	.98	.97	1.05	.98
Chi	1.03	.99	1.00	1.01	1.01	.99	1.01	1.02	.99	.99	.82	1.04	.94	.82	1.04	.95
Clev	.97	1.00	1.07	.98	.98	.99	.99	1.06	.99	1.00	.93	.79	.97	1.03	.75	.90
Det	1.04	1.00	.96	1.02	1.02	1.06	1.00	.96	1.03	1.03	1.27	.98	1.01	1.28	.97	1.00
KC	.97	1.10	.99	.99	.98	1.02	1.08	.97	1.01	1.00	.77	.99	1.10	.76	.99	1.11
Milw	.98	1.10	1.00	.99	.98	.95	1.12	1.01	.98	.97	.84	1.42	1.12	.84	1.42	1.12
Minn	1.04	.96	1.01	1.02	1.02	1.06	.96	1.00	1.03	1.03	1.00	.80	1.01	1.05	.78	.99
NY	.94	1.07	.92	.98	.97	.94	1.07	.92	.98	.97	.91	1.23	.90	.94	1.21	.89
Oak	.89	.93	1.07	.94	.95	.90	.92	1.07	.95	.96	.87	1.00	1.11	.89	.99	1.10
Sea	1.05	.88	1.08	1.02	1.03	1.05	.88	1.08	1.02	1.03	1.38	.75	1.00	1.44	.74	.98
Tex	.96	1.00	.97	.98	.98	.95	1.01	.98	.98	.97	.95	.93	.92	.89	.96	.95
Tor	1.11	.80	1.02	1.05	1.07	1.08	.81	1.03	1.03	1.05	1.08	.74	1.02	1.05	.75	1.03
Atl	1.16	.92	1.03	1.07	1.08	1.17	.92	1.03	1.07	1.09	1.69	.89	.90	1.69	.89	.90
Chi	1.14	.92	1.01	1.06	1.07	1.18	.91	1.00	1.08	1.10	1.29	.81	.92	1.30	.81	.92
Cin	1.04	1.03	1.00	1.02	1.02	1.02	1.04	1.01	1.01	1.00	1.12	1.07	1.01	1.07	1.09	1.03
Hou	.80	1.02	.99	.91	.90	.86	.98	.95	.95	.94	.47	.93	.99	.55	.86	.92
LA	.94	1.09	.94	.98	.96	.94	1.09	.93	.98	.97	1.14	1.31	.86	1.09	1.34	.88
Mon	.99	1.02	.95	1.00	.99	.99	1.02	.95	1.00	.99	.85	1.25	1.11	.85	1.25	1.11
NY	.99	.90	1.03	.99	1.00	.95	.92	1.05	.97	.98	1.04	.72	1.04	1.01	.73	1.05
Phi	1.00	1.09	1.00	1.00	.99	1.03	1.07	.98	1.02	1.01	1.09	1.11	.97	1.12	1.09	.95
Pit	1.09	1.01	.95	1.04	1.04	1.08	1.02	.95	1.04	1.04	1.16	1.04	.99	1.12	1.05	1.00
StL	1.04	1.03	.99	1.02	1.02	1.01	1.05	1.00	1.00	1.00	.80	.86	1.03	.77	.87	1.04
SD	.86	1.01	1.09	.93	.93	.85	1.01	1.10	.93	.93	.75	.83	1.12	.74	.84	1.13
SF	.95	.98	1.02	.98	.98	.93	.99	1.03	.97	.97	.78	1.14	1.06	.76	1.16	1.08

A number greater than one in the TBR column indicates good batting, one less than one in the TPR column indicates good pitching. It is interesting to note that with the park adjustment, the Red Sox have the second best pitching in the majors over the past six years, while their batting ranks only ninth. Houston's batting is slightly better than average, and almost as good as Boston's, while Atlanta and Chicago NL rank close to the bottom of the league in batting. The home run data was calculated in the same manner.

My data file includes games, wins, losses, ties, runs scored and allowed and homers hit and allowed at home, away and total for every major league team from 1900 to the present. In addition, I have individual data on almost every prominent American League player, about 150 in all. The individual data and American League park data was reported in the 1978 Baseball Research Journal.

EVALUATING PITCHERS PERFORMANCES

Dear James Bill,

I been readin this here Baseball Analyst for almost a year now and I aint ever read such pure twaddle, sheer foolishness, absolute falderal and utter garbage. Its time all these rank amateur Savvymetricians cleared the floor and let a real pro take over.

I do kinda take to your method of Runs Produced. Trouble is, you got no cimilar method for the pitchers. I'll pull your fat out of the fire, boy. Just you harken up to the next few paragrafs and I'll smarten up your readers --- if you got any left.

Lets start with the basics. All pitchers is either starters or relievers. Now, I grant you some few (12-15%?) is both during any given season. Usually these starter-reliever-part-timers are young unestablished pitchers so they wont interfer none with what Im proposin. All of us knowledjubul students of the game know that the most important measurement of a starting pitcher is how many men he puts on base (BRA). Walk to strikeout ratios are all well and good but a 2:1 SO to W ratio dont guarantee success --- Duggy Bird, Frank Di Pino, Fowlkes, Hammaker, Tugo McGraw, Donnie Moore, Bert Roberge, Ruhle, Tom Brennan, Moose Haas, Jon Matlack, Gaylord Old Man, Shane Rally, El Tiante all had poor to indifferent 1982s while lookin slick in this area. Hits to innings pitched falls into the same misery --- lots look good here but walk so many fellers that they get zonked when they can least aford to.

Relief pitchers on the other hand often stroll on the scene when the cards is stacked agin em. If they gotta pitch around somebody they dont care if they gotta walk em. BRA dont mean jack shit in their case. The one thing thats gonna burn their derriere is base hits so this is the best measurement for these good ol boys. All you gotta do is check out the stats of Don Stanhouse and Butch Metzger when these guys was in Hog Heaven to get my point.

By figurin the BRA for all AL starters we come up with a figure of 1.391 --- any pitcher who has more than 1/3 of his game appearances as starts is put in this category. The BRA for the NL is 1.341. The H:IP ratio is .972 for AL relievers and.934 for NL boys.

Everbody follow me so far? So good. Once a pitcher ratio is added or subtracted to or from the league ratio were hot to trot. Then we multiply this figure by his percentage of his teams total inning pitched in his respectable category

Example follows on next page for those who aint got too many smarts.

E*X*A*M*P*L*E

	IP		BRA			
Toronto starters	1039		1.391			
Dave Stieb	288		1.219			
	.2772	X	172	47.67	or	48
Toronto relievers	405		.972			
Joey McLaughlin	70		.771			
	.1728		2o1	34.73	or	35

What really grabs you about this system is that it determines the quality element by comparing the pitchers performance with the league average and then extends its value by comparing the quality factor with members of his own team to show the participation factor. BRA performance alone wont cut it. If a guy works in 45 innings and has a BRA of plus .412 it sure dont measure up to the other guy who works in 250 innings and gas a BRA of .400. So % of a fellers team appearances is more important than the number of games in which he shows up. Why? Well, number of games appearing in is great for a shorty reliever but plays hob with the long boys. After all, its easier to pitch 1 inning in 3 of 5 days than it is to pitch 3 innings in 3 of five days and yet under the game appearance criteryon the shorty gets more credit.

Well, that's it. Just check out the Top 5 for 1982 below. Then Ill show you how the teams stack up.

National League

Starters	
Joe Niekro	64
Joaquin Andujar	61
Mario Soto	60
Steve Wonder	55
Reuss	51
Rogers	49
Tortilla Fats	44
Sutton	42
Lea	33
Christensen	30

American League

Starters	
Dave Stieb	48
Jim Palmer	46
Clancy	40
Barker	40
Eckersley	40
Bannister	38
Hoyt	32
John	31
Sutcliffe	30
Caldwell	30

Pete Vuckovich - the Cy Young winner - finished up with a sub-mediocre - 23. Selecting him was like naming my 85-year-old maiden aunt, Tilly Grossbeak, Miss County Seat of Turnup Green, Mississippi. Which actually happened. Everbody knows either Mr. Jockey Shorts or Stieb shoulda been the winner.

Befor we get to the teams let me silence all you well meanin but untutored critiks. Your gonna say "Your system dont allow for those guys who pitch in both relief and the beginun . So what. How many of those dudes are there? And if you care that much go back to the box scores and figure it out. Do you want me to do all the work?

THE CLUBS

NATIONAL LEAGUE

	Starters	Relievers	Total
Atlanta	- 36	25	- 9
Chicago	- 60	- 16	- 76
Cincinnati	- 37	46	9
Houston	109	9	118
Los Angelus	118	13	131
Montreal	98	- 63	35
The Mutts	- 138	- 28	- 166
Philadelphia	- 25	- 75	- 100
Pittsburgh	- 33	28	- 5
St. louis	16	40	56
San Diego	42	73	115
San Francisco	- 98	- 9	- 107

AMERICAN LEAGUE

	Starters	Relevers	Total
Baltimore	79	25	104
Boston	- 92	18	- 74
California	75	38	113
Chicago	- 12	51	39
Cleveland	- 16	12	- 4
Detroit	48	15	63
Kansas City	9	14	23
Milwaukee	- 18	3	- 15
Minnesota	- 67	- 2	- 61
New York	6	58	64
Oakland	- 79	- 94	- 173
Seattle	19	48	67
Texas	- 36	-144	- 180
Toronto	58	- 3	55

Well, there it is. Prove positive of the results. Now, all you smarty-pants Savvy-metricians can go back to the drawing boards and beat your brains out to come up with somethin better. I challenge you.

Scuse me. I forgot the relievers.

National League

Bedrosian	66
DeLeon	63
Bair	46
Reardon	43
Scurry	43
Orosco	36
Forster	34
Kern	32
Garber	28
Lesley	27

American League

The Goose	74
Caudill	60
Armstrong	46
R L Jackson	43
Van de Berg	42
T Martinez	40
Felton	39
Fingers	38
J McLaughlin	35
Andy Hassle	31

Dont this all look purty? Eat you hearts out - you great logishuns.

Your pal,

Cuthbert Magnolia

Just a couple of quick points from the editor:

1) I strongly disagree with Dan Heisman's comment that "It is obvious that in Hall of Fame balloting, quantity is very important. Player's such as Koufax and Dizzy Dean are the exception rather than the rule." On the contrary, it is only in the last fifteen years that career stats have come to mean anything at all in Hall of Fame selections, and to this day it is seasonal, not career, numbers that define Hall of Fame membership.

The Hall of Fame is rife with players, like Koufax, Dean, Jackie Robinson, Hack Wilson, Klein, Campanella, Bill Terry, Lindstrom, Waddell, Chesbro, Chance, Baker, Joss, and many others who were outstanding players for maybe five years, and did not compile impressive career totals. Many players had a longer, flatter career spiral and compiled bulkier totals, like Bunning, Pinson, Doc Cramer, Aparicio, Fox, Wilhelm, and Quinn. And in other cases, players like Goose Goslin, Roger Conner, Sam Crawford and Zack Wheat had to wait for Hall of Fame selection until after contemporaries who had shorter, more meteoric careers.

In few cases is there really a conflict, because in every generation those players who reach the highest peaks and those who have the longest careers are pretty much the same players. The correlation between peak value and career length is extremely high. But when there is a conflict, it is peak value, not gross value, which is the dominant criterion.

2) I choose a different option than Pete Palmer, I believe, in defining what it is we are looking for when we adjust for ballpark effects. My feeling is that I am adjusting not for the "true ability" of the player, which is a thing that we can never know, but for the value of the player in the most relevant circumstances.

For that reason, I find Pete's individual park adjustments, while certainly interesting, not germane to the question of the player's value. Suppose that you have a player who plays in a pitcher's park, but who, because of his ability to take advantage of some unique feature of the park, hits extremely well there. We might be talking about a player who is a dead-pull power hitter playing in Baltimore. Should that player's run estimates be reduced when adjusting him to a normal park? Absolutely not. Wins are the goal of the team, and therefore wins define real value for the team. If the player creates 100 runs while playing in Baltimore, those 100 will likely result in more wins for his team than 100 runs in Fenway.

Thus I disagree with Pete's decision that "it is fairer to (base the adjustments) on just runs allowed. . .because the home team may be made up of batters who can take advantage of the particular park and this might make it appear to be a better hitting park than it should be." No, I say. I say, if they can take advantage of the park, then it is a hitter's park--indeed, that is exactly what a hitter's park is, a park which allows hitters to take advantage of it. What the park "should be" is something we can never know.

The process of adapting talents to parks is a two-sided one, offensive and defensive. I see no advantage whatever in ignoring one of those.

Consider an economic analogy. Suppose inflation raises the cost of goods by 10%, and raises average salaries by 10%. Your own salary, however, does not go up. Should we say, in adjusting for your personal economic situation, that you have not been affected by the inflation? Of course not. But that is exactly what has happened. If Fenway increases runs by 10%, it raises the cost of a win by 10%. When we adjust for the "worth" or "value" of the player, we must adjust not for the impact of the park upon any individual, nor for the impact of the park upon one side of the equation. We must adjust for its impact upon the entire on-field economy.