BASEBALL ANALYST

JOURNAL OF SABERMETRICS
August, 1988 Vol. #37

I have some personal business to clear up here before I get to the articles. First, a couple of issues ago we printed an article by Dan Heisman, to which I wrote a response note. To refresh your memory, Heisman's article argued that critical situations within a ballgame had to be identified based on what the individual could see at the time. In my response I argued that Heisman had confused "importance" with "pressure", and that whether a game situation is or is not important has little to do with what can be seen at the time; many events which seem small later become tremendously important. I wanted to pass along now what Heisman wrote in reply to my comment.

"Perhaps I can simplify and rephrase my position," wrote Dan, "so that there will be no difference on 'philosophical grounds.' All I am trying to say is:

"1) One measure of performance is 'performance under pressure',

"2) There is no difference in pressure on a player who bats in the sixth inning of a 1-1 game and drives in a run that eventually wins the game 2-1 as compared to a player in an identical 1-1 game in the sixth inning driving in a run that eventually leads to a 10-1 game.

"Therefore, measures of their performance with regard to pressure should be equal also. So any good measure of performance under pressure should treat equal situations equally.

"My intention was in no way to 'attack VI-RBI'; I only mentioned it in passing. Your comment on pressure vs. importance is well taken; I meant 'pressure' more than 'importance.'"

OK, thanks, Dan. There will be another article on the general subject in a future issue, an article by Paul R. Pudaite.

Now, a couple of other matters. First, I am still trying to arrange for someone to take over the Baseball Abstract. I thought I had something arranged, but the things I was looking at have fallen through, and I need to get moving quickly on getting something going.

The situation is that

1) The book has to be done on a small scale, self-published; I cannot for reasons of my own allow for a nationally published Baseball Abstract.

2) I have several hundred cards from people who want to buy the book. If you print 1,000 and sell 800, you can make a nice little bit of money on an endeavor like this. If you print 3,000 and sell 600, you lose your shirt. So if you keep your ambitions in check, you can make a couple of month's work worthwhile.

3) I'll help you in any small way I can, so long as it doesn't take too much of my time, and

4) You can do basically whatever you want to do with the book. Other than setting up a few ground rules, I won't interfere.

If you would like to contribute to a 1989 Abstract in some way, please drop me a card--not a letter, a card--which says so, preferably in less than 15 words.

Editor and publisher: Bill James Address: P.O. Box P, Winchester, KS 66097

If you would like to be considered as the editor to carry on the project or as one of two or three co-editors, then let me know. In this case, it might be a good idea to give me something to go on--a resume, some samples of your work, a letter explaining why and how you want to do the book. But ACT QUICKLY; I'm four months behind where I should be already, and next March will be here real quick like. I don't expect, and I don't think anyone else expects, you to meet to the standard of the past Abstracts; I just want the best person I can get to do the best job that he or she can. If you'd like to be that person, please let me know.

Second, I am trying to hire a full-time employee. I'll pay about what a beginning school teacher makes. The person to be employed must be willing and able to relocate in Kansas. The person hired will need to work regular hours in a regular office, plus some. I need someone who is well-organized, literate, knowledgeable about computers, able to type and answer the phone. The person should be prepared to write letters, sort through mail and answer some of it, review articles for the Baseball Analyst, maintain a Rolodex and return phone calls on my behalf. The person should be willing and able to run errands and perform menial duties; basically, my assistant is going to have to do whatever I don't like to do, and there are a lot of things I don't like to do. The person should be able to arrange interviews and perhaps conduct them.

The job doesn't lead anywhere; if you work as my assistant and want to move up in the world, you should expect to move on after a year or two. I don't promise to print anything you write; if you write something, I'll read it and consider it. The person hired will be expected to put up with an author who doesn't explain very clearly what needs to be done, keeps strange hours and often doesn't want to be bothered, is sometimes obstreperous, distant, arrogant and disorganized. And worse yet, a perfectionist; you should expect to be fired very quickly if you don't get done what I think you ought to get done. On the other hand, you will learn more about writing and more about baseball by working for me than you're likely to learn almost any other way. We'll probably wind up going to a few baseball games together.

If you're interested in the job under those conditions, write immediately or give me a call at 913-774-8211. If you know of anyone who you think would be good, tell him or her to call. But act quickly, because I am planning to hire someone before Thanksgiving.

It is also possible that the person I hire as my assistant will be the person who edits the next Abstract; I'm not sure how that will all work out.

Including this introduction, there are six articles in this edition. Those are:

BALL PARK EFFECTS ON HOME RUNS

Robert O. Wood & Robert K. McCleery

Since the rules of the game have not changed much throughout baseball's history, fans love to compare current players with historical counterparts. However sabermetricians have warned that statistics have to be considered in the environment in which they were produced. In an earlier article we demonstrated the need to "factor out" *era* effects so that the hitters of the 1920-1940 period are not proclaimed the greatest of all time based solely on their superficially gaudy batting statistics. In this article we wish to address the other major environmental influence on a player's statistics that is applicable whether or not historical comparisons are drawn: *park* effects.

One of the most robust findings of sabermetric research is the potentially tremendous effect a player's home ballpark can have on his statistics. This tendency was initially noted in the sizable home/road seasonal splits of many players. Going beyond these observations, formal statistical methods have recently been employed. Using data for all players to play in a particular stadium, park effects have been objectively quantified. This issue has gained such notoriety as to warrant including the home/road splits of runs scored and home run totals for every team and every season since 1900 in the latest edition of the Macmillan Baseball Encyclopedia.

Park effects have most often been calculated with reference to runs scored. This focus stems from the methods of the most prominent analysts. Pete Palmer and Bill James both convert a player's season statistics into a "runs created" variable. They then translate runs into wins, taking into account the environment in which the runs were produced. Their method of analysis, while suitable for the purposes to which they employ it, glosses over a few interesting issues.

In this article the focus will be on the effect of a player's home ballpark on his *home run* statistics. In calculating "home run park effects" we have made use of the historical home run data printed in the Baseball Encyclopedia. We needed to undertake this arduous task since home runs explicitly enter our career value formula which we have used to evaluate ballplayers. [For reference see our earlier article "Ranking the All-Time Greats," Baseball Analyst, volume 28, March 1987.]

We do not solely use the *individual's* home/road splits, for we believe that a player's ability to take better than normal advantage of his home park should not be factored out of his statistics. Seasonal park effects are calculated using all home and visiting players' performances. In addition, we have estimated the different effects an asymmetric park has on right and left handed hitters.

For example, even though Joe DiMaggio hit only 41% (148 of 361) of his home runs at Yankee Stadium, we calculate his career home run park effect to be 0.97. This figure is derived from noting that during his career, there were 126 *more* home runs hit in Yankee home games than in road games (1610 vs. 1484). Combining this with the estimate that left-handed home run hitters have about a 15% advantage in Yankee Stadium, and with estimates of right-handed vs. left-handed at bats there, we arrive at the 0.97 home run park effect for the "average" right-handed hitter in Yankee Stadium during DiMaggio's career.

The fact that Joe D. suffered to an extent greater than normal does not persuade us in the least to grant him a park effect of 0.80 or so. All it indicates is that he was able to take less advantage of his situation than others, perhaps due to his reluctance/inability to hit homers to the short porch in right field. If one goes so far as to remove (factor out) all individual home/road splits, one then effectively evaluates players solely based upon their road performance. We do not wish to go that far, since how a player adapts to his home park is a legitimate variable in evaluating him. [Recall that Willie Mays learned to hit to right-center field after he moved to Candlestick Park, effectively turning the prevailing winds from a hindrance to an aid.]

The magnitude of park effects has only recently begun to be recognized. Career runs scored park effects range from 1.10 to 0.90 for the 860 players who have played at least ten years in the majors. This means that one's home park can increase/decrease runs scored by up to 20% (since only half of a player's games are played at home, only 10% appears as the park effect). Of course, even more extreme park effects can be found by looking at individual season data. We are interested in more long-term effects, so have calculated "career park effects", which are simply the seasonal effects weighted by plate appearances over the player's career.

Home run park effects are an order of magnitude *larger* than those pertaining to runs scored. Career home run park effects range from 1.58 to 0.75 (meaning that home parks range from increasing home runs by over 100% to decreasing home runs by 50%) for all 10-year major leaguers. Home run hitters with extreme home run park effects include Gavvy Cravath (1.58) and Mel Ott (1.40) at one end, and Roberto Clemente (0.82) and Jose Cruz (0.79) at the other.

Let us recite a few seasonal home run totals to give the reader a better sense of the immense size of home run park effects. In 1922, there were a total of 154 home runs hit in tiny Baker Bowl by the Phillies and visitors combined, whereas only 50 were hit in Phillies road games. More recently, in 1977 there were 208 home runs hit in Atlanta Stadium, while only 100 were hit in Braves road games. Pitcher's parks can lead to extreme splits as well. In 1922, a grand total of 8 home runs were hit in mammoth Griffith Stadium, while 48 were hit in Senators road games. In 1965 there were only 57 home runs hit in the Astrodome, but 163 hit in Astros road games.

The advent of free agency has allowed owners of teams which play in home run parks to bid for home run hitters, and for players to have more say in where they play. Wrigley Field is obviously not suited to a speed and defense game, but rather to the home run. Andre Dawson recognized that his talents would mesh perfectly with Wrigley Field, and signed a blank contract to play for the Cubs. Of course the home park cannot take all the credit for Dawson's 1987 MVP season, but surely a large part of it.

The Dawson case is one instance of the general rule that the marketplace will act so that home run hitters gradually move to home run parks. [The "devil's theory" of park effects claims that this migration is impeded because teams that play in home run parks greatly over-estimate their true power.] But this should not be construed as only a recent phenomenon. The same incentives were in place throughout history. In trade talks a home run hitter would have more value to a home run park team, and the higher value would have resulted in a better trade offer. Also, a team would do its best to keep a home run hitter happy, even under the reserve clause arrangement. The case of the Pirates changing the configuration of their park to help Hank Greenberg and Ralph Kiner hit home runs is perhaps the ultimate evidence of this. Management effectively converted Forbes Field from one of the most difficult parks in which to hit home runs to one of the easiest. This reasoning can even extend back to a player's initial signing with a ballclub.

As a result of the twin incentives to a team that plays in a home run park to acquire home run hitters, and to home run hitters to want to play in home run parks, we would expect the distribution of home run hitters to be skewed towards home run parks. Table I presents the cross-tabulation of a player's career home run total with his home ballpark. A pitcher's park is defined as one which decreases home runs by over 7%, a hitter's park as one which increases home runs by over 7%, and a neutral park as one within these bounds.

If one looks at the category of 58 sluggers who have hit 300 or more home runs, only 5 played the bulk of their careers in parks that were very difficult to hit home runs in (the reader may want to guess the identity of these five -- they will be listed below), whereas 31 played predominantly in home run parks. The distribution is skewed as we have predicted, but to a degree that should surprise nearly everyone.

TABLE I
Actual Home Runs vs. Ball Park

Career Home Run Total	Pitcher's Park	Neutral Park	Hitter's Park	Total
0 - 99	184	150	145	479
100-199	82	66	87	235
200-299	27	21	40	88
300 +	5	22	31	58
Total	298	259	303	860
Average Home Run Total	97.8	118.9	136.7	117.9

However, the quick-thinking reader may have noticed a flaw in this analysis: "Of course the observed distribution is skewed. First, having home run hitters will make any park look like a home run park. Second, the table is inherently biased since playing in a favorable ballpark inflates a player's home run total (by definition), so that noticing this fact is nothing more than restating that park effects exist, and are sizable."

The first point raised is not valid, since park effects are calculated so that only the deviation from the normal home/road split matters. The fact that Hank Aaron hit 20 home runs in Milwaukee every year does not make County Stadium a home run park, since Aaron hit 25 on the road every year. The normal home/road split would have predicted that Aaron hit 30 at home.

The second point is right on the mark, yet we wanted to present the numbers in their raw form to allow the reader to see the full impact of park effects. It is often via simplistic numbers that deeper points can be more easily understood. The 5 to 31 ratio presented above is much more unbalanced than can be explained by a normal random distribution of these 36 players.

Table II addresses the second point explicitly. It presents the same cross-tabulation except that actual career home run totals are converted to "neutral park" home runs. As an example of this procedure, Jim Rice's total of 364 home runs (all data is through the 1987 season) is divided by his career home run park effect of 1.11 to yield 328 neutral park home runs.

The imbalance of neutral park home runs is not as pronounced as that of actual home runs. Of those sluggers who hit more than 300 *neutral* park home runs, the ratio of pitcher's to hitter's park players becomes 9 to 21, which remains beyond the point that normal randomness could be used to explain it. [The nine include the earlier five, of course, and four others: E. Mathews (512, 554), W. Stargell (475, 533), D. Winfield (332, 376), J. Adcock (336, 375), D. Baylor (331, 345), J. Wynn (291, 326), G. Carter (291, 310), D. Ennis (288, 306), H. Sauer (288, 303); where following a player's name is his actual and neutral park home runs. Note also that Joe DiMaggio (361, 372) and Bobby Bonds (332, 341) fall barely into the "neutral park" category.]

There are several reasons why we should expect the imbalance to remain in the neutral park data among all 10-year veterans throughout history. First, economic incentives would still come into play. The home run hitters will be sought out by teams in home run parks, since that is the most effective way to win games in those parks. Thus free agency, trade value, and initial signing are valid explanations of the skewness of neutral park home runs as well as that of the actual data.

TABLE II
Neutral Park Home Runs vs. Ball Park

Career Home Run Total	Pitcher's Park	Neutral Park	Hitter's Park	Total
0 - 99	173	153	165	491
100-199	83	62	82	227
200-299	33	22	35	90
300 +	9	22	21	52
Total	298	259	303	860
Average Home Run Total	105.9	118.9	123.3	115.9

Another factor is that home run hitters are typically not great defensive players. They are likely to be first basemen or slow outfielders. Since a home run park is sure to have small outfields, a slugger's defensive deficiencies are less important in such a park. [There is a famous quote regarding Duke Snider which pertains to this issue. Comparing his defense to Mays or Mantle, the argument goes that Duke had it tougher playing in cozy Ebbetts Field, since Mays in the Polo Grounds and Mantle in Yankee Stadium "could run forever to catch a flyball without fear of running into a wall". We do not wish to demean Duke, but to simply point out that this argument is ludicrous. Mantle and Mays had to *be able to* (and did) run down many more balls than Snider, which makes us believe that they were better outfielders.] The inverse relationship between defense and home run hitting, along with the need for speedy outfielders in large ballparks can explain almost all of the residual skewness in the neutral park home run data of Table II.

Other factors may include the psychological impact of playing in a cozy ballpark. Virtually every time Andre Dawson made solid contact last year in Wrigley Field, he found himself circling the bases and being feted by the bleacherites. Talk about not being able to wait to go to the ballpark! The converse case, the frustration of crushing a ball in the Astrodome and seeing it come to rest in an outfielder's glove, is not discussed as much. Although we do not put much stock in this hypothesis, ballplayers often mention it when interviewed.

In this article we have come at park effects from a different angle than most other analysts. We saw vividly the tremendous impact of park effects on career home run totals. The imbalance presented in Table I was crystallized in the 5 to 31 ratio of career 300 home run hitters who played in pitcher's vs. hitter's parks. We have focussed upon home runs since they are the most eye-catching of all baseball events. Of course, ball parks can have tremendous effects on a player's other statistics as well.

After we factor out the ball park's effects on home run totals, we still find a slight skewness among veterans toward home run hitters residing in home run parks. We have cited several possible explanations, perhaps the most persuasive being the typical inverse relationship between a player's power and defensive abilities. Since home run parks have small outfields, teams which play there can afford to play a slugging outfielder who might be a defensive liability in a larger ball park. We hope we have raised the reader's level of awareness of park effects and made him better understand the need to consider a player's statistics in the environment in which they were produced.

LATE HOMERUN HITTERS REVISITED

by Stephen Roney

In the August, 1987 Baseball Analyst, I listed a collection of players who hit 65% or more of their homeruns after they turned 30. Recently, I searched the records for all players who have hit 100 homeruns, and I found 405 players who have hit 100 homeruns (through 1987) since 1900. It is possible that I have missed some active players who have never hit 20 homeruns and have reached the 100 level since 1984. The totals of these players have been updated to include 1988 statistics.

The average of all 405 players who have hit 100 homers is 40% after age 30. Removing active players who have not yet hit 40% of their homers after 30 (many are not 30 yet) raises the average to 42%. Players with 400 or more lifetime homers average 49% after age 30. A list of those players is attached. Thirty-five players have hit 70% or more of their homeruns after they turned 30. An additional eight currently active players are over 60%, plus Dwight Evans at 59.9%.

At the low end, only four players who did not play in the majors in 1988 hit 100 homeruns without hitting any after age 30. They are Earl Williams (138HR), Don Hurst(115), Curt Blefary (112), and Fred Linstrom (103). Ken Harrelson(131) hit 5 the year he turned 30 (in September), so is a candidate for this club.

A note on the (30) column. This represents the number of homeruns hit during the season the player turned thirty in mid-season. April/May birthdays were given full credit for after 30, June/July were divided in half, August/September were counted as before 30.

Brian Downing and Frank White are the top active players in this category.

Players who hit 70% of their homers after age 30

	Mo	Yr	life HR	Bef 30	(30)	Aft30	
LO'Doul	3	97	113	0	0	113	100.0%
GCravath	3	81	119	2	0	117	98.3%
KWilliams	6	90	196	6	10	180	94.8%
WMoryn	4	26	101	3	23	75	93.8%
BCerv	5	26	105	6	3	96	93.5%
EJoost	6	16	134	13	0	121	90.3%
SGordon	8	17	202	14	13	175	88.5%
MMinoso	11	22	186	24	0	162	87.1%
GMcQuinn	5	9	135	12	20	103	86.9%
CWilliams	12	87	251	34	0	217	86.5%
DLong	2	26	132	19	0	113	85.6%
HSauer	3	19	288	42	0	246	85.4%
TWalker	9	87	118	16	2	100	85.0%
DLopes	5	46	155	24	4	127	83.8%
WCooper	1	15	173	30	0	143	82.7%
JRobinson	1	19	137	24	0	113	82.5%
MEasler	11	50	118	24	0	94	79.7%
EHoward	2	29	167	34	0	133	79.6%
BRobinson	6	43	166	24	25	117	79.1%
BDowning	10	50	220	47	0	173	78.6%
MVernon	4	18	172	41	3	128	75.9%
GWoodling	8	22	147	27	12	108	75.8%
FWhite	9	50	156	32	7	117	75.6%
PReese	7	18	126	27	9	90	74.5%
RCollins	3	4	135	35	0	100	74.1%
DClendenon	7	35	159	34	14	111	73.6%
JKuhel	6	6	131	28	16	87	73.4%
HBauer	7	22	164	34	17	113	73.3%
RBoone	6	23	151	30	26	95	72.8%
CJohnson	7	47	196	42	22	132	72.2%
CMaxwell	4	27	148	38	24	86	72.0%
ADark	1	22	126	36	0	90	71.4%
HMcRae	7	46	190	51	8	131	70.8%
BElliott	11	16	170	50	0	120	70.6%
JLemon	3	28	164	49	0	115	70.1%

Active players between 60 and 70% after age 30

	Mo	Yr	life HR	Bef 30	(30)	Aft30	
GNettles	8	44	390	105	22	263	69.0%
KGriffey	4	50	139	45	0	94	67.6%
DrEvans	5	47	403	130	17	256	66.5%
CFisk	12	47	323	114	0	209	64.7%
BBoone	11	47	104	37	0	67	64.4%
PGarner	4	49	108	41	0	67	62.0%
BBuckner	12	49	172	68	0	104	60.5%
KMoreland	5	54	115	41	16	58	60.4%
DwEvans	11	51	347	139	0	208	59.9%

Players over 400 homers lifetime

	Mo	Yr	life HR	Bef 30	(30)	Aft30	
DEvans	5	47	403	130	17	256	66.5%
SMusial	11	20	475	174	0	301	63.4%
BRuth	2	95	714	284	0	430	60.2%
TWilliams	8	18	521	197	25	299	58.8%
WStargell	3	41	475	196	0	279	58.7%
CYastrzemski	8	39	452	162	40	250	57.8%
MSchmidt	9	49	542	190	45	307	57.8%
WMays	5	31	660	279	40	341	56.0%
WMcCovey	1	38	521	232	0	289	55.5%
HAaron	2	34	755	342	0	413	54.7%
DKingman	12	48	442	204	0	238	53.8%
RJackson	5	46	563	254	27	282	53.5%
BWilliams	6	38	426	198	30	198	50.5%
EBanks	1	31	512	269	0	243	47.5%
FRobinson	8	35	586	291	33	262	46.3%
HKillebrew	6	36	573	297	39	237	45.3%
LGehrig	6	3	493	267	32	194	43.1%
DSnider	9	26	407	233	43	131	33.7%
MOtt	3	9	511	342	0	169	33.1%
MMantle	10	31	536	374	0	162	30.2%
JFoxx	10	7	534	379	0	155	29.0%
EMathews	10	31	512	370	0	142	27.7%

MORE HOMER NOTES

In the wake of the big homer year of 1987, where a record 25 players hit 20 homers for the first time, only six new players joined the club in 1988, the lowest total (not counting the 1981 strike season) since 1976, when there were only three. The six brought the total number of players with a 20 homer season to 497 since 1920, 499 since 1900 and 505 total.

I wasn't paying enough attention, so I don't know which of Bobby Bonilla, Hubie Brooks, Andres Galarraga, Mike Greenwell, Dave Henderson and Dan Pasqua was the 500th major league player to hit 20. The first newcomer in 1989 will be the 500th since 1900. Fred McGriff was the 198th player to hit 30, and Jose Canseco the 61st to hit 40.

Canseco is the twelfth youngest player to hit 40, three months older than McGwire was last year, and two months older than Cepeda. (See the Feb, 1988 Analyst for the 9 younger than McGwire). He is also the fifth youngest to hit 30 for the third time. Eddie Mathews had hit 40 three times at this age, and Ted Williams, Frank Robinson and Jimmy Foxx had each hit 30 three times (but not 40).

SIMULATOR II - THE LEADOFF MAN AND HIS EFFECT ON THE LINEUP

By Gary Fletcher

Something I really should have addressed in the last article is just why use a computer simulation to answer a baseball question? In most cases we just go to the actual records to see what evidence can be found to illuminate the problem one way or the other.

But the problem with studying the dynamic interaction of the batting order is that it's such a fast changing animal. The most stable lineup you're ever likely to find will, over a season, be chopped up by injury, by pinch hitters, by trades, by hangovers, etc. And if you have some belief concerning lineup construction, then to test it you need yet another lineup to compare it to, one that hopefully has just one difference. Really, it's just impossible.

To conduct a proper experiment you need a control sample, and there just aren't any baseball teams around who are likely to want to co-operate in the effort.

But with a simulation, it's easy. You can make up a lineup with exactly the same batters coming up second to ninth. You then lead them off with a classic no-power, high on-base% type hitter. Then you take the same guys and lead them off with a slugger who maybe gets on only 25% of the time. To make the controls even tighter, you make sure that both leadoff men created the same number of runs in the same number of plate appearances. And then you run each lineup through ten seasons worth of games, which is 1620 games. And then you check to make sure that each leadoff man had essentially the same batting performances coming up behind him.

And that's what I did.

I used the New York Yankees for the bulk of the lineup. Batting second through ninth were Randolph, Mattingly, Winfield, Pagliarulo, Washington, Ward, Cerone and Tolleson. With one lineup I led off with an Ozzie Smith simulation. For the other lineup I constructed an anti-Ozzie whom I called Unsmith. Unsmith was very similar to Ozzie except he hit half as many doubles, no triples and 35 homers a season while drawing only 20% as many walks.

Now, the key thing here wasn't which lineup would score the most runs. As you can see in the enclosed chart, the Unsmith gang did better than Ozzies' gang, scoring 10 runs a season more. Basically, it just happened that the simulated players did better behind Unsmith than they did behind Ozzie. I knew there was a 50/50 chance that that would happen. What I thought was that Ozzies' lineup would be more efficient, that they would do better relative to their runs created estimate than the Unsmiths.

And, by God, they do. The Ozzies scored 7390 runs with only 7240 runs created. The Unsmiths were also efficient, mind you, but not overly. They scored 7506 runs with a runs created estimate of 7503. That's a +150 for Ozzie and a +3 for Unsmith.

Was this over-efficiency consistent from season to season? Sort of. In the ten seasons the Ozzies were +42, +8, -10, +18, +3, -2, +31, +53, -2 and -11. The Unsmiths were -24, +3, +35, +12, -7, +4, -17, -22, +5 and -12. It looks to me that the ability of Ozzie to get on base represents a potential for clutch performance by Randolph, Mattingly and Winfield which was denied them by Unsmith.

Comparing the two lineups in R/RC the totals just steadily climb, the plusses being bigger and the minuses being much smaller. The Ozzies start with a +66 and grow from there to +71, +46, +52, +62, +64, +112, +143, +146 and finally ending up at 147 runs.

Maybe I shouldn't have said clutch performance, even though that's a very different thing from clutch ability. I'll just say that I think the experiment indicates that a high on base% represents a potential for maximizing production. Sometimes that potential is realized, sometimes it is not. A low on base%, on the other hand, severely limits that potential.

And just how large is the effect? Will Rubin Sierra become a superstar? Well, he might. The thing is that an Ozzie Smith is better suited to the leadoff role than a similarly productive hitter with a low on base% simply because he gives the team a greater chance of maximizing its' production. The effect could be as much as 40-60 runs in any one season. Some years, however, that potential might not be realized even slightly.

Which is why it's a good thing that I ran ten seasons worth of games for each lineup. If I had run just one season for each, the conclusion might easily have been reversed, and incorrectly so.

Speculating wildly now, I think that maybe a slightly refined version of the Earnshaw Cook lineup might be pretty effective. If we redefine on base percentage to exclude home runs-let's call it the Actually On Base%-and then just line the hitters up in descending order based on AOBP, that ought to maximize this effect. Maybe I'll run a few simulated seasons based on that premise.

I also kept track of individual game scores for both lineups and this, too, supports the use of an Ozzie as a leadoff hitter. A rough way of calculating how many wins would result, given the number of runs scored by your own team in each game is to count all games as losses in which they score three or less runs. Ozzies lineup had 621 such games as compared to 622 for the Unsmiths. So they're about even there. If you consider runs over 8 to be superfluous, the Unsmiths certainly scored a lot of meaningless runs. They had 146 games of 9 or more runs. The Ozzies had just 116.

So that's two indications of efficiency on the side of high on base% leadoff hitting. You know, OBP and SLG are so often talked about together that you just inevitably start thinking about them as being equal components. Well, maybe they are. But the more I play around with the simulator-with this and other things-the more I've come to feel that OBP is by far the more important of the two.

There are a couple of other things I have to say. First of all, you might be wondering about the lousy performances of the simulated Ozzie Smith and his counterpart. It doesn't affect the conclusion, since both players were still equally productive-averaging about 61 runs created-and just as obviously dissimilar. I was surprised at that too. All of the other simulated players did about what you would expect of them. The simulator considers the "seeded" statistics to be an average performance, so if half of Ozzies seasons had been substandard I would have understood. But they all were. The same thing happened with the Unsmith simulation. I suspect a bug somewhere in the data input program. Again, this has nothing to do with the actual simulation program.

>>>Also, I think you might be interested in the RBI counts for the three players following the leadoff spot in each lineup. You may or may not be surprised at how quickly the effect of any one player on the RBI of subsequent batters diminishes. The effect on the #2 man is huge, the effect on the #3 and #4 batters much less. After those three batters the effect appears to be zero.

PLAYER BY PLAYER COMPARISONS AND GAME SCORE - TOTALS

NAME	PA	S	DB	TR	HR	W	SB	CS	GDP	R	RBI	RC
O.SMITH	7665	1184	346	39	0	778	337	83	98	831	429	612
UNSMITH	7596	1142	176	0	293	158	493	112	87	747	802	604
RANDOLPH	7486	1523	344	22	73	1074	121	10	249	1081	740	973
RANDOLPH	7458	1451	377	19	77	1168	122	11	179	1152	549	1022
MATTINGLY	7323	1353	488	16	342	596	16	43	194	1071	1256	1310
MATTINGLY	7278	1472	423	19	355	571	19	47	205	1040	1185	1350
WINFIELD	7124	1231	230	12	324	805	47	51	232	968	1084	988
WINFIELD	7088	1178	272	13	337	826	42	59	230	997	1045	1029
PAGLIARULO	6905	735	341	36	383	676	12	30	90	950	1097	908
PAGLIARULO	6899	728	341	36	399	637	16	41	122	1013	1079	894
WASHINGTON	6725	1220	360	0	166	569	180	20	57	724	713	896
WASHINGTON	6718	1282	315	0	181	532	183	24	69	761	766	900
WARD	6550	1120	233	20	171	412	129	15	239	618	691	620
WARD	6553	1209	277	12	189	406	148	18	235	653	703	739
CERONE	6398	1051	291	17	86	612	0	9	151	602	519	600
CERONE	6363	1081	231	28	76	672	0	14	178	588	527	571
TOLLESON	6214	1127	57	0	24	630	92	33	45	545	331	389
TOLLESON	6166	1152	67	0	11	685	102	22	50	555	297	407
TOTALS	62390	10544	2690	162	1569	6152	934	294	1355	7390	6860	7240
TOTALS	62119	10695	2479	127	1918	5655	1125	348	1355	7506	6953	7503

O.SMITH LEADS OFF

TIMES SCORING X RUNS

0 RUNS - 100 1 RUN - 123 2 RUNS - 168 3 RUNS - 232 4 RUNS - 220
5 RUNS - 202 6 RUNS - 201 7 RUNS - 151 8 RUNS - 105 9 RUNS - 67
10 OR MORE RUNS - 49

UNSMITH LEADS OFF

TIMES SCORING X RUNS

0 RUNS - 100 1 RUN - 105 2 RUNS - 197 3 RUNS - 220 4 RUNS - 219
5 RUNS - 223 6 RUNS - 169 7 RUNS - 127 8 RUNS - 111 9 RUNS - 90
10 OR MORE RUNS - 56

WHY EVERYONE WANTS LEFTY PITCHING

Dick O'Brien

One of my favorite personalities of all time is that great American sportsman of Irish descent, Lee O'Durocher. (His name is often misspelled). Romping through a passel of well-publicized pecadillos, this one-time player, manager and beau sabreur of the bistros surely deserves a place in the Hall of Fame along with his fellow bacchant, Larry MacPhail, even though his persona might be considered a tad too vibrant with the conservative Establishment.

One of his early shortcomings, philosophically speaking, was his almost pathological aversion to southpaw pitchers. He frequently referred to them as kooks, wackos and assorted fruitcakes, to quote some of his more palatable labels. But this was a folly of youth as he quickly changed his mind when Vito Tamulis and Larry French donned Dodger blue.

Clubs are always looking for good or indifferent left-handed pitchers and with good reason as we can see below. Since 1981 left-handed pitchers have an overall winning percentage of 51% while getting 33% of the decisions. This puts their superiority over right-handers on a level just below that of the home field advantage. Three entries are given for each league: the first shows left-handers W/L record; the second, their winning percentage; and finally their percentage of total decisions.

	National League			American League		
1981	167-147	.532	.32	251-262	.489	.34
1982	269-249	.519	.26	382-368	.509	.36
1983	274-236	.537	.26	390-377	.508	.36
1984	287-276	.510	.29	344-353	.494	.33
1985	298-300	.498	.31	369-360	.506	.35
1986	323-322	.501	.33	341-314	.521	.31
1987	332-321	.508	.34	375-342	.523	.34
	1950-1851	.513	.29	2452-2376	.508	.34

the left-handers achievement shown above is all the more remarkable when we look at the listing below. By and large lefty pitchers are facing right-handed batters 80% of the time while right-handed pitchers are facing lefty batters only a little over 50%. In this listing two entries are given for each pitcher: (1) the percentage of opposite handed batters he has faced in the last four years and (2) the change in opponents batting average against this group when compared to to his overall OBA. (Stats courtesy of GASB).

LHP			RHP		
Bannister	.84	0	Bedrosian	.53	- 15
Browning	.85	- 2	Blyleven	.57	+ 6
Candelaria	.84	- 7	Boddicker	.55	- 6
Dayley *	.71	+ 7	Burke *	.49	- 37
Di Pino *	.71	- 7	Clancy	.53	- 11
Dravecky	.86	- 7	Clemens	.56	- 4
Fernandez	.85	- 1	Darling	.53	+ 2
Franco *	.79	- 9	Dotson	.53	+ 4
Guidry	.82	- 1	Gossage *	.50	- 16
Hammaker	.83	- 17	Gubicza	.55	- 6
Heaton	.82	- 1	Henke *	.52	- 8
Higuera	.84	+ 3	Hershiser	.54	- 18
Hurst	.85	- 1	Mc Dowell *	.48	+ 1
Jackson	.81	+ 2	Morris	.54	- 1
Key	.79	+ 2	Reardon *	.52	- 30
Knepper	.86	- 1	Reuschel	.52	- 15
Langston	.84	- 10	Ryan	.50	0
Nieves	.83	- 5	Saberhagen	.55	+ 2

LHP			RHP		
Ojeda	.82	- 3	Scott	.52	- 7
Orosco *	.77	- 7	Stewart, D.	.53	- 9
Plesac *	.79	- 3	Stieb	.54	- 4
Righetti *	.77	- 2	Sutcliffe	.55	- 5
Smith, Z	.86	- 2	Wegman	.55	+ 7
Swindell	.85	0	Welch	.53	+ 1
Tanana	.83	0	Witt, Bobby	.55	- 18
Tudor	.84	- 4	Witt, Mike	.57	- 9
Valenzuela	.81	+ 2	Worrell *	.44	- 12
Viola	.82	+ 4			

* Relievers see fewer opposite handed batters because they enter the proceedings later.

The pitchers on this listing were selected at random to serve as a representative sampling, but even so the lefties show an overall net loss of 2.5 points while the righties show a loss of 7.7. When one considers that left-handed batters account for 40% of all at-bats, the left-handed pitchers are more than holding their own.

If we look at those batters who have hit 20 or more home runs per season in any one of the last four years and break down their performance against LHP and RHP we can see where the southpaws show their prowess. A total of 36 LHB and 63 RHB make up this group of sluggers.

Lefty batters versus

	AB	Hits	Home Runs		AB	Hits	Home Runs
LHP	14245	2635	500	RHP	33332	9373	1714
	BA .255		AB/HR 28.49		BA .287		AB/HR 19.04

Righty batters versus

	AB	Hits	Home Runs		AB	Hits	Home Runs
LHP	25256	7052	1310	RHP	56974	15174	2708
	BA .279		AB/HR 19.28		BA .266		AB/HR 21.04

All batters versus

	AB	Hits	Home Runs		AB	Hits	Home Runs
LHP	39501	10687	1810	RHP	89606	24547	4422
	BA. 271		AB/HR 21.82		BA . 274		AB/HR 20.26

From this breakdown of slugging performance we can see that left-handed pitchers yield home runs at a lesser rate and have a lower opponents batting average than their right-handed counterparts against the type of batter who outproduces others in RBI. The southpaw success against these heavy hitters goes a long way in demonstrating why they are in demand.

I. PLAYERS WHO HAVE HAD SEASONS SIMILAR TO CANSECO'S

I hope some of you share this fascination. As you know, I like to identify sets of similar players--similar seasons, similar careers, similar profiles of skills. With Jose Canseco having the year that he had in 1988, I became curious about players who have had seasons similar to this. To identify the most-similar players, I compiled a data base of all players who have hit 30 home runs in a season, and then compared those players by means of similarity scores as explained in the 1986 Abstract.

The three seasons most similar to that of Jose Canseco in 1988 were, in order, by Henry Aaron in 1967 (similarity: 968), Wally Post in 1955 (967), and Willie Mays in 1961 (966). These are the stats:

	G	AB	H	2B	3B	HR	RUN	RBI	BB	SO	SB	AVG	SLG
Canseco	158	610	187	34	0	42	120	124	78	128	40	.307	.569
Aaron	155	600	184	37	3	39	113	109	63	97	17	.307	.573
Post	154	601	186	33	3	40	116	109	60	102	7	.309	.574
Mays	154	572	176	32	3	40	129	123	81	77	18	.308	.584

Following these three (in order, and at a safe distance) are Ben Oglivie, 1980, Henry Aaron, 1963, Gil Hodges, 1954, Jim Rice, 1983, Dwight Evans, 1987, Dale Murphy, 1985, and Tony Perez, 1970.

Sometimes you do these studies and then you don't know why you did them, but this list of players crystalized the half-formed thought that was egging me on. The story of Jose Canseco's unique season, as presented by what might be called the World Series media, is the emergence of this exciting young combination of power and speed, a young player who was once regarded as just a slugger but has now developed into a complete superstar, combining a .300 average with the unprecedented combination of 40 homers and stolen bases. What was bothering me on a subconcious level was the recognition that that's not exactly the real story. The real story is not that Canseco combines power and speed, but rather that a player who DOESN'T have much speed, a player who is essentially in the tradition of Jim Rice, Gil Hodges, Dwight Evans and Tony Perez--that is, a formidable slugger with average or slightly above-average speed as a young player--somehow, through determination and intelligence, managed to push 40 stolen bases out of a body that wasn't meant to run. And that is pointed out, I think, by the list of players who did and who DIDN'T show up as most-similar to Canseco--not Bobby Bonds in any season, not Cesar Cedeno or Tommy Harper or Andre Dawson or Eric Davis, but Wally Post and Ben Oglivie and Tony Perez. They are players who hit many doubles but few triples, players who drove in as many or more runs than they scored, and thus players who didn't tend to steal bases.

If you ignore stolen bases, the most-similar season to Canseco was Post's in 1955, followed by Aaron 1967 and Mays in 1961; it doesn't change very much, in other words. If you double the penalty for a stolen base discrepancy, the order switches to Aaron-Mays-Post, but the same three players still occupy the top three spots.

Canseco stole 40 bases but hit NO triples, none; I believe, although I haven't seen final stats, that he also led the league in grounding into double plays. This man really doesn't run all that well, and I think that in five years his speed may be a real problem for him. I also believe that Canseco will have back trouble as he ages--I can't think of a player of that general body type who didn't have back trouble--but he is unusual not only in his tremendous strength, but also in that he has harnassed so much of his ability at such a young age. Gil Hodges and Wally Post and Al Rosen didn't play regularly in the major leagues until they were older than Canseco is, and thus had probably lost a good part of their speed before they were ever exposed at the major league level. But the other side of that is that Canseco, being so young, may not yet have reached his full maturity as a hitter.

II. GENERAL CHARACTERISTICS OF THE 30-HOMER GROUP

Having gone to the trouble of compiling complete data for players hitting 30 homers in a season, I certainly wasn't going to just drop it. The first thing I looked at was characteristics of the group as a whole. There have been, through 1988, 555 seasons in which players hit 30 home runs.

Of the 555 players who hit 30 home runs:

232 (or 42%) hit .300

383 (or 69%) drove in 100 runs.

259 (or 47%) scored 100 runs.

222 players (exactly 40%) both drove in and scored 100 runs; 37 players drove in 100 but didn't score 100, and 161 players drove in 100 but didn't score 100.

420 players (or 76%) reached 100 in one category or the other.

197 players hit .300 with 30 homers and 100 RBI. A year ago I had a count of 195 players having done that, but only one did so in 1988 (Canseco) so apparently I missed one at that time. I have no idea who I missed. The average batting record of a player hitting 30 home runs was (or is) as follows:

G	AB	H	2B	3B	HR	RUN	RBI	BB	SO	SB	AVG	SLG
150	554	164	28	5	36	101	112	77	87	7	.296	.561

The number of players hitting 30 home runs in each year since 1900 is given in the chart below:

	0	1	2	3	4	5	6	7	8	9	Total
190-	0	0	0	0	0	0	0	0	0	0	0
191-	0	0	0	0	0	0	0	0	0	0	0
192-	1	1	4	2	1	2	1	4	3	10	29
193-	10	5	7	3	8	5	4	8	8	4	62
194-	7	5	2	2	1	0	3	5	9	5	39
195-	11	9	6	11	8	10	12	9	11	11	98
196-	10	13	14	8	15	12	15	8	7	17	119
197-	19	11	7	8	4	7	4	19	9	13	101
198-	9	1	16	12	10	13	13	28	5		107

As you can see, we are in a rather extraordinary period. The explosion of 28 players hitting 30 home runs in 1987 was an unprecedented thing--unprecedented as a total, unprecedented per team, unprecedented in that it came out of a period of relative stability, with 10 to 13 players hitting 30 homers in each year 1982-1986. While the number of players hitting 30 homers in a year has gone up and down and up and down since 1929, the drop to only 5 players hitting 30 homers in 1988 is equally unprecedented, not as a raw total but certainly in the relationship it bears to the previous year.

III. SIMILAR SEASONS BY 30-HOMER MEN

Having finished with a fundamental analysis of the data from which we might learn something, I then turned my attention to what I really get into, which is identifying sets of similar players.

The two most-similar seasons in the 555-player sample were turned in by a pair of unlikely MVPs: Boog Powell, 1970, and Mike Schmidt, 1987. Though the players are very different, their seasons were, in fact, exceptionally similar:

	G	AB	H	2B	3B	HR	RUN	RBI	BB	SO	SB	AVG	SLG
Powell	154	526	156	28	0	35	82	114	104	80	1	.297	.549
Schmidt	147	522	153	28	0	35	88	113	83	80	2	.293	.548

Powell was the American League's Most Valuable Player in 1970; Schmidt, though not selected in 1987, did win three MVP Awards with somewhat similar seasons. In addition to their perfect match in the extra base hit categories, Powell and Schmidt were separated by only three hits, by only one RBI, six runs scored and one stolen base. They struck out the same number of times, 80 each. The only significant difference in their seasons is the difference of 21 walks. The similarity score for these two years is 988.3.

Almost as similar as these two years were two seasons turned in by the same player, Dwight Evans of Boston. Evans' 1982 and 1984 seasons are almost a perfect match:

	G	AB	H	2B	3B	HR	RUN	RBI	BB	SO	SB	AVG	SLG
1982	162	609	178	37	7	32	122	98	112	125	3	.292	.534
1984	162	630	186	37	8	32	121	104	96	115	3	.295	.532

The next-best match was that of the two players closest to Jose Canseco, Henry Aaron and Wally Post. While Aaron and Post's similarity to Canseco was in the 960s, their similarity to each other checks in at 987.4. They are followed by two players who had extremely similar seasons in 1962, Ernie Banks and Leon Wagner:

	G	AB	H	2B	3B	HR	RUN	RBI	BB	SO	SB	AVG	SLG
Banks	154	610	164	20	6	37	87	104	30	71	5	.269	.503
Wagner	160	612	164	21	5	37	96	107	50	87	7	.268	.500

The 1962 season was the height of my baseball-card collecting-and-farting-around-with mania, and I noticed this similarity at that time; indeed, noticing that was one of the things which triggered my fascination with these things. I remember concluding that these two were probably the two most similar seasons in my collection, and as I recall there was a third member of the group, another somewhat similar year about the same time. What was that. . .let's see if I can find it. Oh, it was Colavito, the same year, the same yellow Topps card:

	G	AB	H	2B	3B	HR	RUN	RBI	BB	SO	SB	AVG	SLG
Colavito	161	601	164	30	2	37	90	112	96	68	2	.273	.514

Pretty neat, huh? I think if I squinted hard enough I could still tell you what it said in the little cartoon in the upper right-hand quadrant of the card, over the stats. Three players, same season, all had 164 hits and 37 homers in almost the same number of at bats.

The fifth most-similar seasons were by Eddie Mathews, 1965, and Jim Gentile, 1962:

	G	AB	H	2B	3B	HR	RUN	RBI	BB	SO	SB	AVG	SLG
Mathews	156	546	137	23	0	32	77	95	73	110	1	.251	.469
Gentile	152	545	137	21	1	33	80	87	77	100	1	.251	.475

They were followed by Sid Gordon, 1948 (30 homers, 107 RBI, .299) and Yogi Berra, 1956 (30, 105, .298); Berra and Gordon had the same number of at bats, with Berra having one less hit. They were followed by Gil Hodges, 1954, and Roy Sievers, 1957:

	G	AB	H	2B	3B	HR	RUN	RBI	BB	SO	SB	AVG	SLG
Hodges	154	579	176	23	5	42	106	130	74	84	3	.304	.579
Sievers	152	572	172	23	5	42	99	114	76	55	1	.301	.579

Hodges and Sievers, whose records for some reason look even more similar than most of these, were truly similar players, both being right-handed hitting first basemen who were good defensive players despite limited mobility (although Sievers was actually in the outfield most of 1957). In other cases, players who are not really similar at all, like Yogi Berra and Sid Gordon, just happen to post very similar stats for a season. Sometimes the system picks up on similarities you might miss in a casual look but can't ultimately deny, like Willie Stargell and Roy Campanella. Willie Stargell and Roy Campanella? The most-similar season to Campanella in 1950 was by Willie Stargell in 1979. The most-similar season to Campanella's in 1951 was also by Stargell--in 1966. And the most-similar season to Campanella's in 1955 was by Stargell again, again with 1966 season. As a coincidence, that is kind of beyond coincidence; there are real similarities between them as offensive players.

Very often players turn out as most-similar to themselves in another season. Wally Berger's 1934 and 1935 seasons (I'll let you look them up) just miss qualifying for the list of best matches in the study. This is particularly common for great players, for unique players and for consistent players; however, some players who are great, unique and consistent (like Mike Schmidt) still never turn in two seasons which are so similar that no one else can elbow between them. The most-similar season to Babe Ruth, 1920, is Babe Ruth, 1921 (and vice-versa) and most-similar to Ruth in 1924 was Ruth in 1926 (and vice-versa). Most-similar to Ruth in 1930 is, again, Ruth in 1926, and most-similar to Ruth in 1931 is, again, Ruth in 1926. Other players who were most-similar to themselves included Lou Gehrig (1927/30 and 1932/37), Jimmie Foxx (1938/33), Mel Ott (1934/36 and also 1942/37), Earl Averill (1932/34), Hank Greenberg (1937/40), Ted Williams (1946/42 and 1957/41), Charlie Keller (1943/46), Ralph Kiner (1949/47), Stan Musial (1953/54), Hank Sauer (1949/50), Eddie Mathews (1954/55), Willie Mays (1955/65 and 1959/63), Ernie Banks (1955/59), Henry Aaron (1957/62), Rocky Colavito (1962/64), Harmon Killebrew (1964/70 and 1967/70), Boog Powell (1969/70), Jim Ray Hart (1964/66), Dick Allen (1969/74), Jimmy Wynn (1969/74), Bobby Bonds (1969/78 and 1973/71), Dave Kingman (1975/76 and 1982/86), Andre Thornton (1982/84), Eddie Murray (1985/80), Dale Murphy (1983/85), Cecil Cooper (1982/83), Dwight Evans (listed above), and Pedro Guerrero (1982/83).

It is also extremely common for the most-similar season to any season to be turned in by another player in the same year. This happens not only in unique years like 1930, but as a matter of course; because the imprint on the statistics of time and place is so strong, the most-similar season to a year from the late forties is likely to be another year from the late forties. For the same reason, teammates often show up as most-similar. It isn't as common for a player to repeatedly ring into the same player for matches, as Campanella did for Stargell, but it happens; Mickey Mantle in 1956 matched Jimmie Foxx in 1938 and Mantle in 1957 matched Foxx in 1939. The hardest season to match in the group is Roger Maris in 1961, with the best match being another MVP, Harmon Killebrew in 1969; the similarity is only 925.2.

One thing I looked for was "group matches", players who had identical totals in a set of categories such as doubles, triples and homers or runs scored and driven in. There were 31 sets of players with identical totals of doubles, triples and home runs. Mark McGwire's 1988 totals of 22 doubles, 1 triple and 32 home runs, for example, are exactly the same as those posted by Larry Parrish in 1987:

	G	AB	H	2B	3B	HR	RUN	RBI	BB	SO	SB	AVG	SLG
McGwire	155	550	143	22	1	32	87	99	76	117	0	.260	.478
Parrish	152	557	149	22	1	32	79	100	49	154	3	.268	.483

There was one "three-way" set; Al Rosen in 1950, Graig Nettles in 1977 and Eric Davis in 1987 each had 23 doubles, 4 triples and 37 home runs.

There were 15 sets of players who had identical totals of at bats and hits. Darryl Strawberry's 1988 totals of 146 hits in 543 at bats, for example, are exactly the same as those posted by Bill Melton in 1971. There was, again, one three-way set; Dick Stuart in 1964 and Henry Aaron and Norm Cash (both in 1966) each had 168 hits in 603 at bats.

There were 17 sets of players who had identical totals of strikeouts and walks.

There were 34 sets of players with identical runs scored and RBI totals. Glenn Davis' 1988 totals of 78 runs scored and 99 RBI are exactly the same as those posted by Willie Montanez in 1971. There was, again, one three-way set, which again involved Al Rosen in 1950; Rosen, Ralph Kiner in 1953 and Eddie Murray in 1980 all had 100 runs scored and 116 RBI. One of the more interesting sets of matched run and RBI totals were Mel Ott in 1929 and Lou Gehrig in 1932, who matched one another with the rather remarkable totals of 138 runs scored and 151 RBI--both more runs scored and more RBI than for any other set of players.

There were no players whose batting and slugging percentages were literally identical. However, rounded off to three digits there were five sets of players with the same batting and slugging percentages. There were, however, NO "double matches", no players who had identical totals in two of these groups. Which brings me to my final question: is it likely that there are any players, anywhere in history, whose records are exactly identical in a season?

In the major leagues, probably not. In a 555-player group there are 153,735 player-to-player comparisons. If your data base included all regulars since 1900, I estimate that you could make around 100,000,000 (one hundred million) player-to-player comparisons. Even so, I doubt that any two player records are absolutely identical. In this sample, even Schmidt and Powell, the two most-comparable players, are identical in only four of 11 categories. On the other hand, if you included minor league records, the number of comps would go into the hundreds of billions, and then you very probably could find two players with identical records. I'll tell you what--I'll pay a hundred dollars to the first person who finds them. I recommend you use a computer.

IV. SPECIFIC CHARACTERISTICS OF THE 30-HOMER GROUP

Finally, I couldn't leave a good group of players like this without making up lists of players who turned in extreme performance. Who had the lowest batting average for a player hitting 30 home runs? That sort of thing.

The lowest batting average ever for a 30-homer man was .204, by Dave Kingman in 1982; the next-lowest was .210, by the same man in 1986. The lowest slugging percentage ever for a 30-homer man was .417, by Kingman in 1985.

The fewest games played and fewest at bats ever for a 30-homer man were 102 and 354, by Mike Schmidt in the strike-shortened 1981 season; with that exception those records are held by Rudy York in 1937, with 104 and 375 (as I suppose many of you know, York was on the bench most of that year, but exploded with 18 home runs in August and several more in September to clear the 30-mark comfortably.) The most games played in a

30-homer season is a record shared by two temmates, Ron Santo and Billy Williams in 1965, when each played in 164 games. Kirby Puckett in 1986, batting leadoff most of the year although he rarely walks, batted 680 times in a 30-homer season, breaking a record shared by Jim Rice (1978) and Don Mattingly (1986); they had batted 677 times each.

The fewest hits ever for a 30-homer man was 100, by Ron Kittle in 1984. Two players got to 30 homers with only 9 doubles, those being Gus Zernial (1955) and Kingman (1982). Gus Triandos in 1958 had 30 homers but a total of only 40 extra base hits, a record low. Triandos also scored only 59 runs that year, fewest in the group, although it is noteworthy that the alleged 1979 NL MVP, Willie Stargell, scored only 60. Forty-one of the 555 players hit no triples to go with their 30+ homers, a group which needless to say included both Triandos and Stargell, and fifty-two players (52) have had 30 homers but no stolen bases, a group which includes Stargell but not Triandos, who registered his only career stolen base in that season. Stargell scored only 28 runs when not homering, fewest of any player.

Three players in the group had only ten extra-base hits other than home runs--Triandos, Kingman (1982) and Henry Aaron (1972).

Re Brook Jacoby in the 1988 Abstract I wrote that "Sixty-nine RBI. . .has to be a record low for a guy who hits .300 with 30 homers; I can't believe anybody's ever done that before." Not hardly; as I learned, only one player ever drove in fewer runs with 30 homers, that being Felix Mantilla in 1964 (64 RBI); the second-lowest figure before Jacoby was 72, by Colavito in 1966. Mantilla drove in only 34 teammates on the year, again the fewest in the group, and his ratio of RBI per home run (2.13-1) tied for the lowest; Harmon Killebrew in 1963 had precisely the same ratio with half again as many homers (Mantilla 30-64, Killebrew 45-96). Mantilla also scored only 69 runs, so he participated in only 133, lowest in the group.

The fewest total bases for a 30-homer season was 210, by Gus Zernial in 1955; Zernial had 120 bases on home runs, but only 90 on singles, doubles and triples. This, however, is not a record for the highest percentage of bases on homers, nor even does the record belong to Ron Kittle in 1984, when he had two more homers than Zernial (32) but only one more base (211); Kittle had 128 bases on homers, only 83 from other sources. The 83 bases from non-homers ties Dave Kingman (1982) for low in the group, but as a percentage, the highest percentage of bases on homers ever was by Roger Maris in 1961; Maris' 244 bases on 61 homers were exactly twice as many as he had on singles, doubles and triples. He is followed on that chart by Kingman, 1982 (64%).

Kittle can, however, claim the fewest runs created for a man hitting 30 homers; he created only 61 runs. He is followed by Kingman, 1986 (62 runs created) and Zernial, 1955 (64). The most runs created was by Ruth, 1921 (233, using the basic runs created method), followed by Ruth, 1923 (216) and Gehrig, 1927 (211). Ruth in a different season, 1920, created 19.3 runs per 27 outs (highest), while Kingman in 1986 created only 3.75 per 27 outs (least).

The fewest walks ever drawn by a man hitting 30 homers were 22, by Bob Horner in 1979. The most strikeouts ever was 187 by Bobby Bonds in 1969. On the good side of that one, Joe DiMaggio in 1941, when he hit in 56 straight, struck out only 13 times in a 30-homer season, by far the fewest; second on that list is Lefty O'Doul, 1929, with 19, and then DiMaggio in 1939, with 20. O'Doul had the most hits in a 30-homer season, 254; Rogers Hornsby in 1922 and Chuck Klein in 1930 had 250 hits each. Klein had 59 doubles among his 250 hits, a record for a 30-home run season; Medwick in 1937 had 56. Not counting his 40 home runs, Klein that

summer had 285 total bases and scored 120 runs, both group highs. Two players have hit 20 triples and 30 homers in the same year, those being Jim Bottomley (1928) and Willie Mays (1957). Lou Gehrig had a total of 70 doubles and triples (52-18) in 1927, a record for a player hitting 30 or more homers. Of course, the most extra base hits total is the all-time record, 119 by Ruth in 1921, followed by Gehrig.

Hornsby and Ted Williams are the only .400 hitters who hit 30 homers; Hornsby did it twice, but Williams' .406 average was the highest.

Leon Wagner in 1964 made 479 outs in a 30-homer season, the most ever; oddly enough, Wagner and the man who made the fewest outs, Mike Schmidt in 1981, had identical extra base hit totals of 19-2-31. Other than Schmidt, the fewest outs made in a 30-homer season were by Mantle in 1962 (another MVP) and Ted Williams in 1957, his .388 year.

The most stolen bases ever in a 30-homer year is 50, by Eric Davis.

Felipe Alou in 1966 scored 122 runs and drove in only 74, by far the highest ratio of runs to RBI for a 30-homer man. Remarkably, Alou drew only 24 walks and stole 5 bases, one of the lowest run element ratios in the group. Ordinarily, run element ratio and the ratio of runs to RBI are highly correlated; the players with low run element ratios score more runs than they drive in. Alou batted leadoff most of the year despite obvious reasons why he shouldn't, but if it will make you feel any better, his manager was fired late in the year. Alou had the lowest secondary average of any 30-homer man, .249; Ruth in 1920 had the highest, .825.

The highest run element ratio in the group was 1.45-1, by Jimmy Wynn in 1969; Wynn drew a National-League record 148 walks and stole 23 bases. Despite batting cleanup, he scored 26 runs more than he drove in. The lowest run element ratio was by Walker Cooper in 1947 (26/145), while Dave Kingman had the highest ratio of RBI to runs scored (118 RBI, only 68 runs scored, 1984). No player had a run element ratio exactly 1.00.

The most teammates driven in (RBI minus home runs) in a 30-homer season is 143, by Hank Greenberg in 1937; he is followed by the RBI record-holders, Gehrig in 1931 and Hack Wilson. The highest ratio of RBI/homer in a 30-homer season is 4.97-1, by Medwick in 1937 (154 RBI, 31 homers), followed by Vern Stephens, 1950, and then Greenberg in a different season, 1935 (36 homers, 170 RBI.) Medwick accounted for only 32% of his total bases with home runs, lowest percentage in the group.

Harmon Killebrew in 1964 had 49 homers among 60 extra base hits, or 80.3%, the highest ever.

Lefty O'Doul in 1929 hit .348 when not hitting a home run, the highest such average, and hit .284 just on his singles alone, by far the highest ever; Williams is second, almost 40 points behind him. On the other hand, Dave Kingman in 1982 hit just .135 when not hitting a home run, and two players (Gorman Thomas, 1979, and Tony Armas, 1983), hit just .111 when not getting an extra base hit.

The highest batting average when not striking out was .478, by Babe Ruth in 1923; Ruth is the only man who hit .400 in his career when not striking out. The lowest batting average when not striking out was .270, by Ken Harrelson in 1969 (Kingman hit .271).

The worst strikeout-to-walk ratio for a player hitting 30 home runs was 6 to 1 (162-27), by Butch Hobson in 1977; the next-worst was by Cory Snyder. The best strikeout-to-walk ratios were by DiMaggio (5.84-1) and Ted Williams, both in 1941.

I figured the players accounting for the highest and lowest percentage of their runs created with their home runs. The three highest were Kingman, Kingman and Kingman (1986-1982-1976), clearly establishing Kingman as the King of the one-dimensional sluggers. The lowest percentages were for O'Doul, Medwick and Babe Herman. Good-bye.