By the Numbers

The Newsletter of the Statistical Analysis Committee of the Society for American Baseball Research
Volume 4, Number 3 October, 1992

Committee News

News from the Convention. I know I'm getting this to you a little late--the Convention was, after all, four months ago--but punctuality has never been my strong suit. The Committee met on Sunday morning, very early, and we had 27 people present, including about 10 new members. The chief outcome of the meeting was a resolution to re-start the Statistical Analysis Bibliography Project. David Nichols has agreed to coordinate the bibliography project, and you should send information about articles incorporating statistical analysis of baseball to him at 2555 Bell Ave., Mountain View, CA 94045. You can also reach him on Internet, where his address is nichols@prc.xerox.com. I will maintain a file of all the articles people identify, so you can send those to me; my address is below.

In conjunction with the StatComm Bibliography Project, you should be aware that SABR has a Bibliography Committee, chaired by Robert McAfee. He maintains a file of articles sent to him by SABR members, and his committee publishes a newsletter that incorporates listings of the articles received recently. If you want to get involved with that work, write him at 5533 Coltsfoot Ct., Columbia, MD 21045.

Changing Chairs. This is the last issue of ***By the Numbers*** that I will edit. Beginning with Vol. 4, No. 4, Rob Wood (2101 California Street, #224, Mountain View, CA 94045) will be doing this. Also, he is the new chair of the Statistical Analysis Committee. Address changes, requests for information, and articles for the newsletter should be sent to him.

Requests for back issues of ***By the Numbers*** (Vol. 1, No. 1 through Vol, 4, No. 3) should be sent to me. I can provide these for $2.50 per issue (which covers duplicating and postage costs).

This Issue. This issue is our convention issue, incorporating articles based on presentations at SABR 22 in St. Louis last June. We have a piece by David Smith, analyzing the "quality start," and concluding that it is a useful measure of pitching performance. Our second piece is by Harold Brooks, and looks at the effects of playing every day on performance in September--is fatigue a factor managers need to be aware of? He finds that it is.

Richard David Adams looks at the genesis of home field advantage, concluding (among other things) that it is inversely related to variables measuring expertise. Alden Mead examines methods of measuring relative performance, finding that we can reach different conclusions depending on our measure of relative performance. He suggests a correction for one such measure. Finally, Mark Pankin provides us with a further extension of his work on batting order effects.

Future Issues. As I mentioned above, Rob Wood will edit future issues of ***BTN*** and you should send your articles to him. I have a feeling he will generally be in the same situation I was, needing material for each issue (there's never a backlog of stuff to print). So fire up your computers or calculators, write up your findings, and send them to Rob (2101 California Street, #224, Mountain View, CA 94045) TODAY.

Donald A. Coffin
Indiana University Northwest
3400 Broadway
Gary, IN 46408

The Quality Start Is A Useful Statistic

by David W. Smith

The quality start is a relatively new statistic devised in an attempt to evaluate the performance of starting pitchers in terms other than the traditional values of ERA and wins and losses. A starting pitcher is credited with a quality start if he pitches at least six innings and allows three or fewer earned runs. The original motivation for this study was to pursue a statement by Bill James in the ***1987 Bill Lames Baseball Abstract***.

While discussing the pros and cons of the quality start as a statistic, he noted that there had been criticism (by Moss Klein, writing in ***The Sporting News***) of the quality start in that it would be possible for a pitcher to go exactly six innings and allow exactly three earned runs in every start, compiling an ERA of 4.50, although each start would be categorized as a quality start. Klein thought that thus possibility invalidated the entire concept. James thought Klein's criticism was an unreasonable one, based on using an extreme example, and he ventured that "...I doubt that any pitcher had an ERA higher than 3.20 in his quality starts." My intuition on this point agrees with his, so I decided to pursue the question by using the Project Scoresheet data base, which covers all Major League games played from 1984 through 1991. James was right about the ERA of pitchers in Quality Starts, but there are in fact other interesting conclusions as well.

First, let's address the initial question: What is the ERA of pitchers in their Quality Starts? Table 1 is a summary of the results from all games, from 1984-1911, 16,831 games total, of which about 52% were quality starts. The information in this table is divided two ways. First, each league is presented individually, and, second, the games are divided by Quality Starts and non-Quality Starts. There are three main points to make from this table:

1. The ERA of starting pitchers when they have a Quality Start is over ***five*** runs per game better than in games they do not have a Quality Start (1.91 vs. 7.50).
2. The winning percentage of pitchers when they have a Quality Start is more than twice what it is when they done have one (0.674 vs. 0.311).
3. The innings pitched per start is also substantially different in Quality Starts and in non-Quality Starts (7.45 vs. 4.80).

The conclusion seems clear: Quality Starts, taken in the aggregate, reflect much better than average performance, with the result that the team is much more likely to win. Furthermore, these good performances are also of longer duration, meaning that the bullpen is given some rest when a Quality Start is taking place.

Table 1: Summary of Quality Start Data, 1984-1991

	ALQS	ALNQS	NLQS	NLNQS
Starts	8935	9195	8522	7010
%QS		49.3%		54.9%
Wins	6184	2878	5590	2927
Losses	2749	6313	2927	4833
W/L PCT	0.692	0.313	0.656	0.309
IP/Start	7.54	4.85	7.37	4.73
ERA	1.93	7.65	1.89	7.28

What about Moss Klein's concern about the "minimum" Quality Start--six innings, three earned runs, 4.50 ERA? Over the last eight years, there have been 17,457 Quality Starts, and 989 of them have been exactly six innings and three earned runs. This is 5.7% of the Quality Starts, what I call the "Klein Percentage." As shown in Table 2 (on the following page), there are five categories of Quality Starts that occurred more frequently, all with substantially better ERAs. It's pretty clear that Klein has missed the significance of Quality Starts by his concentration on the extreme case.

The level and quality of Quality Starts have been consistent over the past eight seasons (charts available by sending a SASE to the Editor of ***BTN***). The two leagues are different from each other, reflecting the higher level of scoring in the AL, but the two leagues tend to move together, indicating that the same forces are at work in both leagues. There is also little year-to-year variation.

Table 2: Leading Categories of Quality Starts, 1984-1991

IP	ER	ERA	Number	% of all QS
7.0	2	2.57	1348	7.7%
9.0	0	0.00	1343	7.7%
7.0	1	1.29	1254	7.2%
9.0	1	1.00	1163	6.7%
6.0	2	3.00	1000	5.7%
6.0	3	4.50	988	5.7%

Another approach to this question is to look at the performances of individual pitchers. Table 3 is a summary of the 18 pitchers with the most Quality Starts since 1984. The results are presented in two ways--total number of Quality Starts and Quality Starts as a percent of starts. The first list rewards longevity and we see some clear differences in the two lists. Frank Viola's total of 177 is impressive, but his 62.1% QS puts him in the middle of this select pack. Power pitchers, such as Gooden and Clemens, are near the top, while finesse pitchers (Tanana, Blyleven, Boddicker) are at the bottom.

Table 3: Individual QS leaders, 1984-1991

A. QS Leaders: Total Quality Starts

	QS	TS	%QS
Frank Viola	177	285	62.1%
Dwight Gooden	167	236	70.8%
Ron Darling	164	251	65.3%
Roger Clemens	161	240	67.1%
Bob Welch	160	259	61.8%
Nolan Ryan	158	251	62.9%
Charlie Hough	154	268	57.5%
Orel Hersheiser	154	216	71.3%
Mike Scott	153	233	65.7%
Dave Steib	150	242	62.0%
Frank Tanana	149	260	57.3%
Tom Browning	146	255	57.3%
Bret Saberhagen	145	226	64.2%
Mike Boddicker	144	264	54.5%
Dave Stewart	135	230	58.7%
Jimmy Key	132	217	60.8%
Bert Blyleven	131	231	56.7%
Doug Drabek	118	183	64.5%

QS=Quality Starts; TS=Total Starts; %QS=QS/TS

Table 3 (Continued)

B: QS Leaders: QS as a Percent of All Starts

	QS	TS	%QS
Orel Hersheiser	154	216	71.3%
Dwight Gooden	167	236	70.8
Roger Clemens	161	240	67.1
Mike Scott	153	233	65.7
Ron Darling	164	251	65.3
Doug Drabek	118	183	64.5
Bret Saberhagen	145	226	64.2
Nolan Ryan	158	251	62.9
Frank Viola	177	285	62.1
Dave Steib	150	242	62.0
Bob Welch	160	259	61.8
Jimmy Key	132	217	60.8
Dave Stewart	135	230	58.7
Charlie Hough	154	268	57.5
Frank Tanana	149	260	57.3
Tom Browning	146	255	57.3
Bert Blyleven	131	231	56.7
Mike Boddicker	144	264	54.4

QS=Quality Starts; TS=Total Starts; %QS=QS/TS

I note with some surprise that Nolan Ryan is only marginally ahead of Viola on the percentage list and that Orel Hersheiser is the overall percentage leader, even though he is not a true flame-thrower.

The last thing to look at in examining the individual pitchers is exceptional performance in a single season. Table 4 gives two glimpses at some great performances. In Part A of Table 4, we see that only three pitchers have had as many as 30 Quality Starts in a single season (since 1984), led by Dwight Gooden's unbelievable 33 in 1985, or 94.5% of his starts. Part B presents the best ERA seasons for those with at least 20 Quality Starts in a season. Four pitchers had a Quality Start ERA of less than 1.00, led by Gooden at 0.95, which wasn't even in the year he had his 33 Quality Starts! (By the way, his Quality Start ERA was 1.34 in 1985.) Note that even in these excellent seasons, these pitchers had non-Quality Start ERAs above 6.00, with Gooden at 8.56. At a minimum, this table tells us that a pitcher who has a Quality Start has really done something special, and that a non-Quality Start, even for the top pitchers, is a far inferior performance.

Table 4: Exceptional Individual Seasons for Quality Starts, 1984-1991

A. Top Individual Seasons: Number of Quality Starts			
	QS	TS	%QS
Dwight Gooden (1985)	33	35	94.5%
Mike Scott (1986)	32	37	86.5%
Bret Saberhagen (1989)	30	35	85.7%
B. Best Season ERAs in Quality Starts			
	QS/TS	QSERA	NQSERA
Dwight Gooden (1984)	21/31	0.95	8.56
Mike Moore (1989)	22/35	0.95	6.53
John Tudor (1985)	27/36	0.98	6.60
Jack Morris (1986)	20/35	0.99	7.39
Orel Hershiser (1985)	25/34	1.01	7.38
Dennis Martinez (1991)	21/31	1.07	6.54
John Tudor (1988)	21/30	1.16	6.54
Mike Boddicker (1984)	20/34	1.18	6.07
Bob Ojeda (1986)	21/30	1.20	7.52

So what does this all mean about the meaningfulness of the Quality Start statistic? That question translates to the effect of a Quality Start on winning the game, which is, after all, the purpose of the competition. The answer is complicated somewhat by the fact that both starting pitchers in a game could very well have a Quality Start, but only one team will win. From 1984 to 1991, there were 4,793 games in which both pitchers had Quality Starts, a very large 54.9% of all Quality Starts (since both pitchers had Quality Starts in these games, these represent 9,586 out of the 17,457 total Quality Starts in the Major Leagues). To sort this out, one might suppose that the thing to do is look only at those games in which one pitcher has a Quality Start, to see if his performance was important. Well, I did that, but I'm not going to present those numbers, because it's really not a meaningful analysis[1]. Think about what those games represent: One pitcher does very well and the other one does not. It is hardly surprising, or meaningful, that, in those games, the pitcher with the Quality Start rarely loses. The other variable to take into account is that pitching changes are done differently in the two leagues, due to the designated hitter. We are all familiar with the argument that National League pitchers are likely to be pinch-hit for in close games, even if they are pitching well, a possibility that American League pitchers do not face. Another the potentially confounding effect, which I have not explored here, is the changing use of relief pitchers, with the increasing use of middle men, set-up men, and closers.

It is nonetheless clear from Table 1 that, in both leagues, teams whose pitchers have a Quality Start have outstanding records n those games. Remember that this record is even better than it appears, since most of the Quality Start losses were in games in which the other starting pitcher also had a Quality Start.

I believe that the greatest value of the Quality Start, however, is not simply in predicting which team will win a given game, although I understand that many would like to interpret it in that way. The greater, and less measurable, value is rather in the durability of the starter in keeping his team in the game for longer periods and in avoiding overuse of the bullpen.

****

A REMINDER

Send your new articles for ***By the Numbers*** to

Rob Wood
2101 California Street, #224
Mountain View CA 94505

For back issues (Vol. 1, No. 1 through VOl., 4, No. 3), write (enclosing $2.50 per issue)

Donald A. Coffin
Indiana University Northwest
3400 Broadway
Gary, IN 46408

****

[1] Editor's note: Anyway, if you want to do it, you can, simply by subtracting 4,793 wins *and* losses from the Major League totals from Table 1.

Playing Every Day and September Performance

by Harold Brooks

Introduction. Consider the following two players. Both play, very well, a crucial defensive position. Offensive statistics for eight years out of the heart of their careers have been averaged for 700 plate appearances, approximately what they averaged per year. Which player would you rather have on your team?

Table 1: Performance Data

	Player A	Player B
AB	625	636
H	177	162
2B	33	31
3B	4	2
HR	26	24
BB	75	64
K	68	69
BA	.283	.255
SA	.471	.424
OBA	.360	.323

Clearly Player A is a better offensive performer than is Player B, hitting for more power and walking more often, as well as hitting for a higher average. The only problem in the choice is the fact that these two players are actually the same man. Player A is Cal Ripken's performance from April through August in 1984-1991. Player B is Ripken in September and October for those same years. Since Ripken is the only man to play every game for all eight of those years, a logical question to ask is whether playing every day is the reason behind the late-season decline. In the specific case of Ripken, we probably can't come to a definite conclusion. Three further questions of a more general nature come up, however, that might shed light on Ripken's situation:

1. Is Ripken alone or of other *every day* (or nearly every day players suffer a similar decline?
2. If other players decline as Ripken does, do they have anything in common other than getting very few days off?
3. Again, if other players decline as Ripken does, how much rest is required to lessen the chances of a player declining late in the season?

To evaluate these questions, I will look at players who started almost every game in a season from 1984 through 1991, using monthly data from ***The Elias Baseball Analyst*** and ***The Great American Baseball Stat Book*** from those years. the players are initially divided into three groups by number of games started. The three groups are 158-162 starts, 153-157 starts, and 148-152 starts. The measures of performance I use are batting average (BA), slugging average (SA), on-base average (OBA), and the simplest form of runs created per game (RC/G), as described by Bill James in his ***Baseball Abstracts*** from 1982 to 1984. RC is given simply by:

RC = (Hits+Walks)*(TotalBases)/(At-Bats+Walks)

Using a base of 25.5 outs per game, as suggested by James to represent the number of outs included in (AB-H), RC is divided by the number of games' worth of outs used by a hitter. That is, a player making 255 outs would have used up 255/25.5 = 10 games. If in doing so he had created 45 runs, he would have 4.5 RC/G.

In order to provide some context, we should know that, on average, overall batting performances decline in September. Over the eight-year period, major league BA, SA, OBA, and RC/G [B/S/O/R] decline by .002, .006, .002, and 0.11 respectively after the end of August. As a result of this, all of the discussion of players will consider how they performed relative to the league/season they were in. In general, the numbers to be presented will report the mean value of the group of players meeting the criteria for games started and position played, as well as the percentage of players declining, and the percentage of players falling within the categories of RC/G. These categories are major decline (more than 2 RC/G, minor declines (between 1 and 2 RC/G) and neutral (less than 1 RC/G), minor rises, and major rises.

Results. As a starting point, let's look at all of the players starting at least 148 games over eight years, regardless of position played (see Table 1). N represents the number of players in each category and the values in the other columns represent changes from April-August performance to September-October performance relative to the league; a negative number means that a player declined more than the league average. There is a slight tendency for players starting more games to perform worse late in the season.

Table 1: Performance Change For All Players Starting At Least 148 Games, 1984-1991

Starts	N	BA	SA	OBA	RC/G
158-162	77	-.002	.000	-.007	-.13
153-157	100	-.001	.007	.002	.09
148-152	137	.007	.015	.008	.33
Total	314	.002	.008	.002	.13

Similarly, a larger fraction, 25% of all players starting at least 158 games, suffers major declines late in the season than those with more rest. Meanwhile, those getting 10-14 games off are much more likely to improve by 2+ RC/G in September than to decline (19% versus 9%) (see Table 2)[2].

Table 2: Declines in RC/G

	Decline			Rise	
Starts	Major	Minor	Neutral	Major	Minor
158-162	25%	14%	29%	18%	14%
153-157	19%	15%	32%	13%	21%
148-152	9%	16%	36%	20%	19%
Total	16%	15%	33%	17%	18%

Segregating players by position played is also of interest. During this time period, only one catcher, Benito Santiago in 1991, has started as many as 148 games in a position at catcher. Obviously, this is not a large enough sample to consider the effects of rest on performance. Similarly, there are not many first basemen in the sample. Only nine first basemen started at least 158 games. There is a suggestion that the amount of rest has little effect on first basemen's offensive performance at the end of the season, but the sample is too small to make a definite determination. As a result, two groups have been selected for further analysis. The first is the "throwing" infielders--second base, shortstop, and third base. The second is the outfield. Discussion is restricted to players making the requisite number of starts at these positions. For example, a player starting 100 games at third base and 60 games at second will not be included, but a player starting 100 at third and 60 at second will be. The results are summarizes in Tables 3-6, which present information similar to that in Tables 1 and 2, except for "throwing" infielders (Tables 3 and 4) and for outfielders (Tables 5 and 6).

Table 3: Performance Change For All Players Starting At Least 148 Games as Throwing Infielders, 1984-1991

Starts	N	BA	SA	OBA	RC/G
158-162	36	-.015	-.014	-.016	-.49
153-157	39	.004	.021	.010	.41
148-152	58	.003	.011	.005	.22
Total	133	-.002	.007	.002	.07

Table 4: Declines in RC/G, "Throwing" Infielders

	Decline			Rise	
Starts	Major	Minor	Neutral	Major	Minor
158-162	28%	22%	19%	25%	6%
153-157	13%	13%	36%	15%	23%
148-152	9%	21%	34%	16%	21%
Total	15%	19%	31%	18%	17%

Looking at players segregated by position, the dominant feature is the poor late-season performance of throwing infielders getting fewer than five games off. Their batting, slugging, and on-base averages all drop by

[2] (Editor's Note) We can test whether the observed distributions differ from expectations by performing a Chi-squared test. This test measures whether the ***observed*** distribution of performance differs from the ***expected*** distribution. We can use the performance distribution for the total sample as our expected distribution. When we calculate the Chi-squared statistic, it is 12.32, compared to a critical value of 11.07, which suggests that the observed distribution is significantly different from the expected distribution.

approximately 0.15, on average (see Table 3). Half of them suffer a decline of more than 1 RC/G and, as a whole, the group drops by 0.49 RC/G (10% of the pre-September value). Only 2 of the 36 players have major rises in the late season (Table 4). This is in sharp contrast to the infielders getting more rest, where major rises are more common than major declines. The no-rest effect is not just due to the presence of Cal Ripken in the data set. More than 20% of the non-Ripken seasons have major declines, while only 7% have major increases. With more rest, the fraction of infielders having major offensive declines decreased. Players receiving more rest have a much greater chance of having an improved late-season performance and a much lesser chance of having a poor finish.

Table 5: Performance Change For All Players Starting At Least 148 Games in the Outfield, 1984-1991

Starts	N	BA	SA	OBA	RC/G
158-162	16	.007	.012	-.002	.16
153-157	27	-.003	-.019	-.003	-.33
148-152	43	.009	.015	.013	.45
Total	86	.005	.004	.005	.15

Table 6: Declines in RC/G for Outfielders

	Decline			Rise	
Starts	Major	Minor	Neutral	Major	Minor
158-162	19%	13%	31%	13%	25%
153-157	33%	7%	33%	7%	19%
148-152	9%	21%	30%	26%	14%
Total	19%	15%	31%	17%	17%

The outfield group shows a somewhat different pattern, with less extreme behavior at the low-rest end of the scale. While the fraction of players declining is approximately 50% at the two lower rest intervals, 33% of the outfielders starting between 153 and 157 games suffer major declines, most of which results from a loss of .019 in SA. This is in contrast to the other two groups of outfielders and makes it difficult to reach definitive conclusions about the effects of rest on outfielder performance. By the 10-14 game rest interval, however, the fraction of outfielders declining has fallen to 43%; and 40% have at least a minor rise in September. As with infielders, the chance of a major decline has fallen. Curiously, the highest rest class for both infielders and outfielders shows a tendency for fewer major performance changes, hinting at a possibility that more rest results in more consistent behavior.

Conclusions. Some attempts to answer the questions posed in the Introduction can be made, based on the data presented above:

1. Cal Ripken is not alone in suffering declines when playing every day of the season. Indeed, every-day players seem to be more likely to decline than to improve.
2. Throwing infielders (2B, SS, 3B) appear to suffer more than other players when they play every day. Ripken declines slightly more than his counterparts, but that may be a result of his getting ***no*** days off while the group studied here included players receiving up to four days off. Interestingly, despite the adverse effect on infielders of everyday play, they are more likely to receive little rest than are outfielders (4.5 infielders per year in the majors, compared to 2 outfielders).
3. As little as five days off alleviates many of the problems in September. Certainly, by 10 days off, the chances of a September offensive disaster are small.

There are some interesting implications of these results. Playing players every day increases the chances of poor late-season performance, a factor that may be critical for teams entering post-season play. Only 1-2 days off per month may be sufficient to counteract the apparent effects of fatigue. Anecdotally, Cal Ripken's best September between 1984 and 1991 was in 1985, when he received two additional days off in August because of a players' strike.

On a final note, it is possible that September declines for every-day players are a relatively recent phenomenon. The current major league baseball schedule consists of 162 games in 182 days. The players thus get 20 days off, plus the number of doubleheaders played. With the near-death of the scheduled doubleheader and the decrease in the number of rainouts requiring additional doubleheaders, teams rarely get more than 25 days off in a season. This number used to be much higher. For example, the 1969 Mets had 36 days off, or more than ten days more than any team in 1991. This is approximately the number of days off that appears necessary to lessen September declines. As a result, it is possible that September declines were less frequent in the past.

(Acknowledgement: Greg Spira, of Whitestone, NY, helped gather data for 1989 and 1990.)

****

Home-Field Advantage Reconsidered: The Effect of Expertise on Peak Performance

by Richard David Adams and Susan Jeanne Kupper

In a previous study, Irving and Goldstein[3] noted that home-field advantage seemed to increase as the superiority of performance increased and found home-field advantage to have a statistically significant relationship with no-hit major league baseball games. This paper consists of six sections. The first section provides a review of Irving & Goldstein's study. The second section provides an alternative explanation for the effect of home-field advantage on the performance of no-hit pitchers. The third section proposes three hypotheses in support of this explanation. The fourth section presents the data and methods used to test the hypotheses. The fifth section supports the results of these tests. The sixth section presents conclusions and suggestions for future research.

A Review of Irving & Goldstein. Irving & Goldstein collected data on no-hit games from the ***Sports Encyclopedia of Baseball***[4] and predicted that a significant number of no-hitters occurred on the home field of the winning pitcher. They excluded no-hitters of less than nine innings and no-hitters that were lost in extra-innings. They determined the number of no-hitters won at home versus on the road to be 111 to 64. They used a Chi-squared goodness-of-fit test (with an expected probability of .500). Their results are summarized in Table 1.

Table 1: Irving and Goldstein's Results

	Expected Probability	Expected Wins	Actual Wins
Home	0.500	87.5	111
Road	0.000	87.5	64

Chi-squared: 12.62

When we replicated Irving & Goldstein's work, we discovered some discrepancies. First, home-team wins were found to be 117 to 58 rather than 111 to 64. This difference was confirmed from multiple sources. Second, the home-field probability of .500 used by Irving and Goldstein was found to be inaccurate. The historic home-field advantage for major league baseball is .5426. Third, since the variable under observation is binomial (an event being won at home or on the road), the binomial probability is known (P_{home} = .5426), and the events are independent of each other, a binomial probability distribution is more appropriate than a Chi-squares goodness-of-fit-test.

The results of our replication are summarized in Tables 2 and 3. When reviewing these tables, please note that the positive effects of the home-field advantage are still statistically significant.

Table 2: Corrections to Irving & Goldstein's Results

	Expected Probability	Expected Wins	Actual Wins
Home	0.5426	94.95	117
Road	0.4574	80.05	58

Chi-squared: 11.19

[3] P.G. Irving and S.R. Goldstein, "Effect of Home-Field Advantage on Peak Performance of Baseball Pitchers," ***Journal of Sports Behavior***, Vol. 13, No. 1, 1990, pp. 23-27.

[4] Neft and Cohen (eds.), ***Sports Encyclopedia of Baseball***, St. Martin's Press (New York, NY: 1988).

Table 3: Results Using a Binomial Distribution

Number of Trials:	175
Number of Successes	117
Number of Failures:	58
Probability of Success:	.6686
Expected Probability of Success:	.5426
Probability Of Observing This Difference If the True Probability of Success Is .5426	.00048

An Alternative Explanation. The alternative explanation is that when variables of expertise are considered, levels of competitive performance can be reliably differentiated among cognitive dimensions.[5] This is, in fact, the general finding in studies of both intellectual and physical competition.[6] The contribution of Irving & Goldstein is that rather than differentiating performance, they held performance constant and differentiated environment. It is our contention that when variables of expertise are considered, it can be shown that environment is not a significant factor in the performance of individuals who demonstrate expertise.

Hayes-Roth, Waterman & Lenat[7] described expertise as consisting of knowledge about a particular domain, an understanding of the domain-specific problems, and a skill for solving those problems. Dreyfus & Dreyfus[8] consolidated this description when they referred to expertise as a competence level within an operative domain. Arkes & Freeman[9] demonstrated empirically that the competence level of expertise can be constrained by environmental dependencies.

Territorial superiority as described by Irving & Goldstein is a common phenomenon in nature. However, the requirement of multi-territorial competition creates problems for attempts to generalize this phenomenon to sports competition. Unlike organisms which are able to limit their competitive endeavors to territories in which they exhibit superiority, sports competitors are generally required to compete in multiple arenas under varying conditions. In these situations, what appeared to be a territorial superiority is actually an environmental dependency, i.e., a difficulty of maintaining performance independent of environment. Some examples of environmental dependencies in baseball are Astroturf vs. natural grass; day games vs. night games; and playing at home vs. playing on the road.

Environmental dependencies are expected to have a negative effect on all levels of individual performance (e.g., game, season, and career). In the absence of environmental dependencies, individual performance can be expected to optimize within the constraints of ability and motivation. One of the best ad hoc comments on the optimization of performance independent of environment comes from Hall of Fame third baseman and perennial Gold-Glove winner Brooks Robinson. When asked, prior to the 1970 World Series, what problems he might have fielding on Astroturf for the first time, he is reported to have said, "I'm a major league third baseman. If you want to play in the parking lot, I'm still supposed to make the play."

Hypotheses. Given that consistency of performance is a metric for expertise[10], two metrics, well-accepted in the expertise literature, are proposed for evaluating winning pitchers of no-hit games:

1. The ability to replicate specific performance and

[5] D.J. Garland and J.R. Barry, "Sport Expertise: The Competitive Advantage," ***Perceptual and Motor Skills***, Vol. 70, 1990, pp. 1299-1314.

[6] For a complete bibliography of these studies, send a SASE to the Editor of ***BTN***.

[7] F. Hayes-Roth, D.A. Waterman, and D.B. Lenat, "An Overview of Expert Systems," in F. Hayes-Roth, D.A. Waterman, and D.B. Lenat (eds.), ***Building Expert Systems***, Addison-Wesley (New York, NY: 1983).

[8] H.L. Dreyfus and S.E. Dreyfus, ***Mind Over Machine***, The Free Press (New York, NY: 1986).

[9] H.R. Arkes and M.R. Freeman, "A Demonstration of the Costs and Benefits of Expertise in Recognition Memory," ***Memory and Cognition***, Vol. 12, No. 1, 1988, pp. 84-89.

[10] For a complete bibliography of these studies, send a SASE to the Editor of ***BTN***.

2. The ability to sustain career performance.

Given a general hypothesis that home-field advantage is inversely related to expertise, when these metrics are applied to the data, three effects should be observable:

H1: Home-field advantage should not be statistically significant for repeat performance pitchers, i.e., those who have won multiple no-hitters.

H2: Home-field advantage should not be statistically significant for high-career-performance pitchers, i.e., those who have a relatively high number of career wins.

H3: Home-field advantage should be statistically significant for low-career-performance pitchers, i.e., those who have a relatively low number of career wins.

Data and Methods. The data set for this paper was extracted from ***The Baseball Encyclopedia***, and career wins for current pitchers was updated from ***The Official Major League Baseball 1992 Stat Book***. The data set consists of all major league no-hitters from 1990 through 1988, using the criteria adopted by Irving & Goldstein (a total of 175 no-hitters). Each observation included the league, the teams, the date of the game, the winning pitcher, his career wins, whether the home team won or lost, and whether or not the winning pitcher won other no-hitters.

Since it is well-established in the expertise literature that expertise is an effect rather than a correlation, the appropriate method is to test for the effect in homogeneous subsets of the data rather than testing for a correlation across the full data set. Therefore, three data sets were extracted, one for each hypothesis. The data set for repeat performance pitchers consists of 48 observations. The data sets for high and low career wins were created by rank-ordering the data in descending order of career wins and extracting the first and last 35 observations.

There are two data-related issues requiring experimental control. First, the data sets have observations in common ($n1 \cap n2 = 18$ and $n1 \cap n3 = 2$). This results in multiple tests of the same data and requires control of the experiment-wide error rate. Thus an alpha level of .15 (.05*3) is used. Second, nonparametric methods in general, and the binomial probability in particular, are sensitive to the size of the data set. It can be demonstrated that given a constant ratio, the binomial probability will increase as the number of trials increases. Consider the following example, using a 60% success ratio:

Trials:	200	40
Successes:	120	24
Expected Probability:	.500	.500
p <	.006	.269

To demonstrate that variations in significance are due to the data and not to the partitioning of the data set, we will use a two-test approach. This approach is designed to answer two questions:

1. Can the null hypothesis be rejected on the basis of the extracted data set at $\alpha = 0.15$?
2. Would the null hypothesis have been rejected if the size of the extracted sample were equal to the size of the full data set (175) at $\alpha = 0.15$?

For hypotheses H1 and H2, which attempt to fail to reject the null hypothesis, both tests must fail to reject. For hypothesis H3, both tests must reject. the parameters for the second test require a linear transformation setting the number of trials to 175.

Results. Hypothesis H1 states that home-field advantage should not be statistically significant for repeat performance pitchers. There have been 21 pitchers who have won two or more no-hitters (48 no-hitters total). Of these, 28 were won on the home field of the winning pitcher (.583), which is not significantly different from the overall home-field advantage of .5426 (see Table 4).

Table 4: Testing H1

	Test 1	Test 2
No-Hit Games	48	175
Won at Home	28	102
Won on Road	20	73
Probability	.5833	.5833
Expected Prob.	.5426	.5426
p <	.3381	.1603

Hypothesis H2 states that home-field advantage should not be statistically significant for high-career performance pitchers. Of the 35 non-hitters won by pitchers in the upper quintile of career wins, 20 were won on the home field of the winning pitchers. H2 is supported (see Table 5).

Table 5: Testing H2

	Test 1	Test 2
No-Hit Games	35	175
Won at Home	20	10
Won on Road	15	75
Probability	.5724	.5724
Expected Prob.	.5426	.5426
p <	.4338	.2461

Hypothesis H3 states that home-field advantage should be statistically significant for low-career-performance pitchers. Of the 35 no-hitters won by pitchers in the lower quintile of career wins, 24 were won on the home field of the winning pitcher. H3 is supported by the data (see Table 6).

Table 6: Testing H3

	Test 1	Test 2
No-Hit Games	35	175
Won at Home	24	120
Won on Road	11	55
Probability	.6857	.6857
Expected Prob.	.5426	.5426
p <	.0616	.0001

Conclusions and Future Research. Support of the first two hypotheses indicates that, based on the available data, the performance of repeat no-hit pitchers and no-hit pitchers with relatively high career wins were not significantly affected by home-field advantage. That is, they did not exhibit difficulty in maintaining their performance independent of environment. Support for the third hypothesis indicates that the performance of no-hit pitchers with relatively low career wins is significantly affected by home-field advantage. These findings support the findings of Garland & Barry that when variables of expertise are considered, levels of competitive performance can be reliably differentiated along cognitive dimensions. While these findings support the findings of Irving and Goldstein that these is a significant relationship between home-field advantage and no-hit performance, they indicate that this relationship is consistent with the view that home-field advantage is inversely related to variables of expertise.

The basis for both this study and the Irving & Goldstein study is no-hitters. Whether or not no-hitters represent peak performance is an empirical question which neither study addressed. If the use of peak-performance is essential to sports behavior research, then further research is necessary to develop well-calibrated metrics for peak performance.

****

Commentary

As those of you who have read this newsletter regularly know, I generally don't offer commentary on the work done. Here, however, there are a couple of comments I think are necessary to fully understand Adams and Kupper's findings.

First, they always use final career wins as their measure of expertise. This is an ex post measure of expertise, rather than a current measure of expertise. To put it another way, if a pitcher pitches a no-hitter in his fifth major league start, with a current record of 3-0, and winds up with a career record of 175-160, he is credited with 175 career wins for purposes of determining his expertise. Adams and I had a lengthy discussion about this after his presentation, and while he had some persuasive arguments in favor of this approach, it still troubles me. When we observe a player early in his career, we may think we can gauge how that career will turn out, but there are many confounding variables, which may not be related to true expertise, that intervene (injury and military service come to mind). I'm not sure we would reach the same conclusions if we used current rather than career wins as our measure of expertise.

A related issue is that all the no-hitters of multiple no-hit pitchers are included in the sample for Hypothesis H1. Surely this is wrong. When a pitcher pitches the first no-hitter, we do not know--and can not know--that it will be the first of two or more. A more appropriate approach would include only subsequent no-hitters in this sample. (Again, Adams and I talked about this, and neither could persuade the other that he was right).

Nonetheless, I think this is an excellent piece of work, and the general point that home-field advantage may be less important for more skilled players is surely worth investigating. Several possibilities come to mind. One is seasonal and career batting performance; another is pitcher wins. I think there is a challenge here that some one should take up.

****

Remarks on Comparing Statistics From Different Years

by Alden Mead

It's generally agreed that a direct comparison of statistics such as batting average from different years can be misleading if one fails to take account of what constituted normal performance in each year, as represented by league averages. An oft-cited case is Bill Terry's .401 batting average (BA) in 1930. Surely, as has been pointed out before, this must be viewed with some skepticism in view of the fact that the entire National League, including pitchers, batted .303 that year. Eleven years later, for example, Pete Reiser led the NL with a .,343 BA, much lower than Terry's BA, but the league batting average (LBA) had been reduced to .258.

Which of these performances was "better"? Certainly, the question cannot be decided just by saying that .401 is higher than .343, since the great difference in the LBA for the two years seems to indicate that it was easier to achieve a high average in the NL in 1930 than it was in 1941.

One possible way to make a comparison, which has been used in the past, is through the Batting Average Ratio (BAR), obtained just by dividing the player's BA by the LBA:

BAR = BA/LBA.

In the example we were considering, the BARs work out as follows:

Terry, 1930, BAR = .401/.303 = 1.323.
Reiser, 1941 BAR = .343/.258 = 1.329.

The apparent conclusion from this is that Reiser had a higher batting average, ***relative to the league***, than did Terry.

Or did he? The same information that the batting average contains can equally well be represented by the "Out Average" (OA), defined as the ratio of hitless AB (outs, or O) to total AB:

OA = O/AB = (AB-H)/AB = 1.000 - BA

Obviously, a low OA means exactly the same thing as a high BA; an OA under .600 is the same as a BA over .400, etc. A player's OA may be compared with the league's OA (LOA). In terms of OA, the Terry-Reiser matchup looks like this:

Player/Year	OA	LOA
Terry, 1930	.599	.697
Reiser, 1941	.657	.742

If guides listed OA instead of BA, we'd phrase the question this way: How much did Terry's ***lower*** OA compare with Reiser's higher one, in view of the fact that the NL had a much lower OA in 1930 than in 1941? To make the comparison, we might define an Out Average Ratio (OAR) as the ratio of the LOA to that of the player:

OAR = LOA/OA

With this definition, a high OAR still represents a good performance relative to the league. Let's see how our heroes compare:

Terry, 1930 OAR = .697/.599 = 1.164
Reiser, 1941 OAR = .742/.657 = 1.129.

This time, Terry seems to win! Moreover, there's no clear basis for claiming that either of these methods is better than the other, since both treat exactly the same information with exactly the same philosophy. ***There is clearly a logical contradiction!***

A solution to this logical contradiction (though not necessarily to all problems of comparing statistics) is to define a ratio which treats hits and outs in a symmetrical way. One can do this through the combined average (CA) defined as

CA = H/O = BA/(1.000 - BA)

A Combined Average Ratio (CAR) is the ratio of the player's CA to the league CA (LCA):

CAR = CA/LCA = (BAR)*(OAR)
=(BA*LOA)/LBA*OA)

Terry's 1930 CAR works out to 1.540, Reiser's in 1941 to 1.501, so Terry wins the comparison after all.

Does this prove that Terry's achievement was greater than Reiser's? I think that would be too rash a claim. However, if one has no information other than the batting averages of the players and the leagues, and if one in convinced that the player's achievement must be measured by comparison to the league average, this is a more rational way of doing it than either BAR or OAR.

The CAR method can also be used to compare other averages that are bounded by .000 and 1.000, such as fielding averages, on base averages, averages for bringing runners in from third base, etc.

I think, though, that a certain amount of skepticism and good judgment needs to be applied to any comparison of individual to league statistics. For example, Babe Ruth's 60 HR in 1927 amounted to 13.7% of the AL total, while Roger Maris's 61 in 1961 were only 5.0% of the AL total (adjusted for the change from 8 teams to 10). It may be true that Ruth's achievement was greater than Maris's, but I don't think these numbers alone prove it. In 1927, most players were still playing as if they were in the dead ball era, unwilling to sacrifice average for power. By 1961, many were swinging for the fences, which would increase the league HR total whether or not it has actually become *easier* to hit home runs.

In conclusion, comparisons of player's with league statistics can be useful and informative, but such comparisons should be designed to be free of contradictions and even then a bit of judgment needs to be used.

Two tables of BAR, OAR, and CAR accompany this article. To get them, send a SASE to Donald A. Coffin, Indiana University Northwest, 3400 Broadway, Gary, IN 46408.

****

****

Remember the Bibliography Project

Send listings of articles using statistical analysis to David Nichols, 2555 Dell Ave., Mountain View, CA 94045. Send copies of the articles themselves to Donald Coffin, Indiana University Northwest, 3400 Broadway, Gary, IN 46408 **AND** to the Chairman of the Bibliography Committee, Robert McAfee, 5533 Coltsfoot Ct., Columbia, MD 21045

****

Finding Better Batting Orders: Simulation Tests

by Mark D. Pankin

For several years, I have been developing techniques to find batting orders that are expected to score the most runs for a given nine starting players. An article describing the models and the basic results appeared in ***By The Numbers,*** Vol. 3, No. 5 (December, 1991). A natural question is "Do these lineup optimization techniques really work?" Short of a controlled experiment using major league teams (a most unlikely event!), a definitive answer is impossible. However, various models can be employed to try to find answers. The article discussed the results of applying the Markov Process model that underlies the lineup optimization techniques. Those results, for 1990 major league teams, showed the optimized lineups had a small, but definite advantage, over those typically used by major league managers. However, there is a type of circular reasoning taking place. The optimization techniques were based on data generated by the Markov model, so it is hardly surprising that the Markov model validates the optimized lineups. In this article, I will discuss the results of some recent tests using a baseball simulator.

The simulator used is the APBA computer baseball game, which is published by Miller Associates. By using its companion Micro-Manager product, series of up to 255 games under the control of a computer manager are possible. The capability for automatic play of long series is necessary for testing the run scoring potential of different batting orders. The APBA game is particularly well suited to the testing because it incorporates both my lineup optimization techniques and Markov model.

The basic idea behind simulation testing is to play many games matching a team against itself using two different lineups, one standard and one optimized. Everything else should be the same. The nature of the computer manager can also affect the meaning and outcome of the test. For these tests, I used "Blackie Dugan", a computer manager supplied with MicroManager. Blackie is supposed to be a savvy major league manager, although my impression is that he overmanages. The important point is that Blackie changes pitchers and uses pinch hitters as he sees fit. This is in contrast to the basic formulation of the Markov model, which performs its calculations as if the same nine batters played the entire game and hit against average 1990 pitching all the time. Consequently, this APBA simulation provides a much different test of the lineup optimization techniques than the previous Markov model tests.

I performed the tests on 1991 major league teams. I chose the highest rated lineup for each team. (The APBA software displays the five highest rated lineups, and it is not unusual for one of the lower rated ones to have slightly higher expected scoring according to the Markov model.) For each team, four series of 255 games, a total of 1020, were specified. (Due to rainouts and rain shortened tie games, the total number of games played to a decision was slightly less.) Each type of lineup was the home team in half the series. To gain additional uniformity, each team in a league used the same three-man, right-handed, starting rotation. One was an above average pitcher (Appier in the AL, Cone in the NL), one was average (Navarro, Smoltz), and one below average (Terrell, Burkett) according to their APBA ratings. Each team had five relief pitchers available. The team's primary relievers were included, and if necessary, additional relievers were "drafted" and some of the team's own were dropped in order to get a total of five that Blackie would use out of the bullpen.

Tables 1 and 2 summarize the results of the tests. We see that the AL played to a virtual standoff both in games won and runs scored. For the NL, the optimized lineups had a small advantage. The average difference of 14.4 runs per 162 game season is consistent with the 0.510 winning percentage, which translates to 83-79 for a season. It is worth noting that the specific formulas employed by the lineup optimizer are different quantitatively, but not qualitatively, depending on whether or not the designated hitter is used. This difference may account for performance difference between the two leagues; and it certainly identifies an area for additional investigation.

Table 1: American League

Lineup Type	Won	Lost	WPCT	Runs per 162 games
Optimized	7115	7120	0.4998	712.6
Standard	7120	7115	0.5002	711.8

Table 2: National League

Lineup Type	Won	Lost	WPCT	Runs per 162 games
Optimized	6230	5981	0.510	696.9
Standard	5981	6230	0.490	682.5

Results for individual teams are not shown because, based on some of the outcomes and a small amount of additional testing not discussed here, I feel that 1000+ games are not enough to be meaningful when comparing the effects of two lineups. I also do not present the results by starting pitcher because there are no clear patterns and I am in doubt as to the significance of those results with regard to lineup efficacy. However, scoring in games started by the three classes of pitchers did follow the expected pattern, and the differences in total game scoring between adjacent classes were over one run per game.

To summarize, I feel that the tests show the lineup optimizer may well lead to higher scoring lineups, and these lineups do not appear to do any worse than conventional ones. Keep in mind that the tests may have been influenced by the actions of the computer manager, Blackie Dugan. One indication such might be the case is the relatively small difference between AL and NL scoring. Actual 1991 AL scoring averaged 726.6 runs per team per 162 games, and the NL averaged 664.3. I would expect simulated scoring to be higher than the actuals because any injuries to the regular players, who are usually better than their replacements, lasted only for the duration of the game. The effect of the uniform starting pitching rotations is unclear.

My judgment is the 14000+ and 12000+ games for the league tests are large enough to produce meaningful results. The best way to address some of the issues raised here is to perform additional testing using a variety of scenarios and different simulators. Because such testing can be laborious and consume large amounts of computer and calendar time, I am looking for all the help I can get. If anyone is interested and has a suitable computer baseball simulation, please contact me (1018 N. Cleveland St., Arlington, VA 22201, 703/524-0937). There is a lot of fun and insight to be gained here. Any takers?