

Effect of Batting Order (not Lineup) on Scoring

By David W. Smith

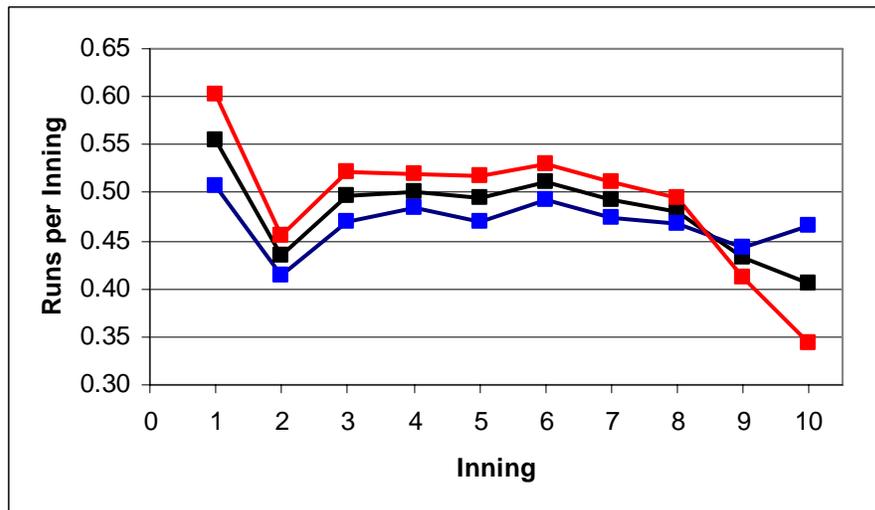
Presented July 1, 2006 at SABR36, Seattle, Washington

The study I am presenting today is an outgrowth of my presentation in Cincinnati in 2004 in which I addressed the pattern of scoring in a game, such as the possible importance of one-run wins, come from behind wins, etc. The first point I established was the rate of scoring by home and road teams in each inning. The results for 1957-2005 are in Table 1 and graphically in Figure 1.

Table 1. Average runs in each inning for home and road teams, 1957-2005

| Inning | Visitor | Home | Both |
|--------|---------|------|------|
| 1 | 0.51 | 0.60 | 0.55 |
| 2 | 0.41 | 0.46 | 0.43 |
| 3 | 0.47 | 0.52 | 0.50 |
| 4 | 0.48 | 0.52 | 0.50 |
| 5 | 0.47 | 0.52 | 0.49 |
| 6 | 0.49 | 0.53 | 0.51 |
| 7 | 0.47 | 0.51 | 0.49 |
| 8 | 0.47 | 0.49 | 0.48 |
| 9 | 0.44 | 0.41 | 0.43 |
| Extra | 0.47 | 0.34 | 0.40 |

Figure 1. Average runs in each inning, 1957-2005. Red: home team. Blue: visiting team. Black: combined. Inning 10 refers to all extra innings combined.



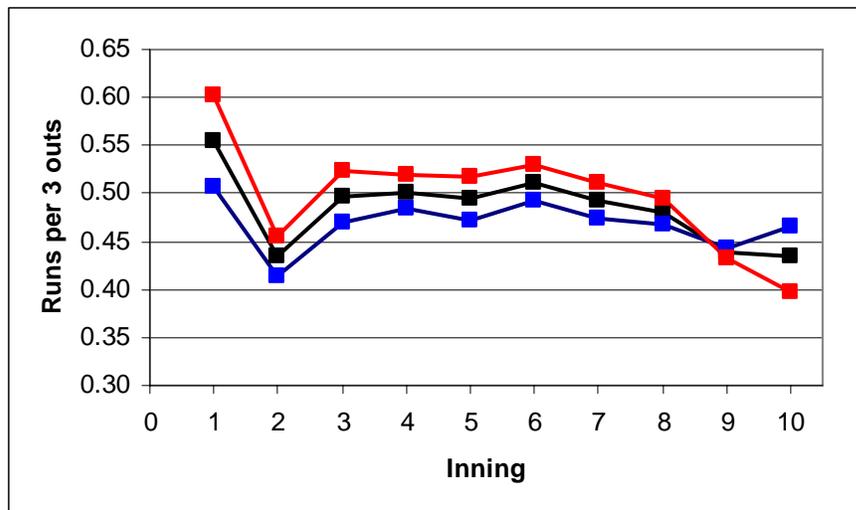
Two points of special note are:

- 1) More runs are scored in the first inning than any other, and this difference is by a wide margin, especially for the home team.
- 2) The home team averages more runs than the visitors in each of the first eight innings, although that pattern reverses in the 9th inning and later.

This second point deserves an additional comment. At first glance, it might seem surprising that the home team scoring drops so dramatically in the 9th and extra innings. However, most of this apparent decline is accounted for by the fact of “walkoff” wins. That is, when the home team gets enough runs to win in the 9th inning or later, they stop batting, with the result that fewer total runs are scored. For the period studied here, 1957-2005, there were 9053 walkoff wins, which works out to one in every 10.5 games. This sizable number would seem to be enough to account for at least a sizable portion of the observed drop. Included in this total are 4646 extra inning wins by the home team, all of which are, of course, walkoffs.

My friend, Clem Comly, suggested a different way to look at scoring rate for these innings, which is to normalize scoring per 3 outs made in each inning. In this way, the partial innings of the walkoffs will be mitigated to some extent. The result of that recalculation is seen in Figure 2.

Figure 2. Average runs in each inning, normalized per 3 outs, 1957-2005. Red: home team. Blue: visiting team. Black: combined. Inning 10 refers to all extra innings combined.



The values for the first 8 innings are, of course, unchanged, as are those for the visitors in the 9th and extra innings. The rates for the home team in these last two categories are increased as expected, but they don’t quite catch up to the visitors because there are still many potential runs that are not scored due to the walkoffs.

It occurred to me and to several who saw these results in Cincinnati that a good place to start in trying to figure out the underlying factors causing this pattern was to consider which batters were likely to bat in each inning. Therefore, I began my journey into the land of lineup studies, territory already well staked out by Mark Pankin (see, for example, <http://www.pankin.com/markov/btn1191.pdf>). However, Mark’s pioneering work has concentrated on the starting lineup and how variations there can affect scoring. He also made great use of sophisticated modeling to explore these questions in very interesting

ways. Tom Ruane has also done some Markov modeling work on the effect of lineups (http://www.retrosheet.org/Research/RuaneT/lineup_art.htm).

My approach here is a bit different in two ways. First, I looked at actual performance data, not model results and second, I focused on the consequences of different lineup positions batting first in a given inning. In this study, it is essential to distinguish clearly between the starting lineup and the batting order in a given inning. When I refer to the starting lineup, I will use the term “lineup slots”, whereas when I refer to the batting order in an inning, I will speak of the first batter, second batter, etc.

This analysis requires play by play data and I used the Retrosheet files from 1957 through 2005, as summarized in Table 2.

Table 2.
Data for present study, 1957-2005

| Seasons | Games | Half-Innings |
|---------|--------|--------------|
| 49 | 95,755 | 1,719,857 |

Note: There were actually 95979 games played from 1957-2005, but Retrosheet has not yet acquired play by play data for 224 games from 1957-1973. Therefore, 99.8% of games played from the last 49 seasons were available.

The basic pattern of which batter leads off an inning is shown in the following two tables, with raw totals in Table 3 and the same information in percentage form in Table 4, which allows much easier comparisons.

Table 3. Number of times each lineup slot batted first in an inning. “Extra” refers to all extra half innings. The 40688 extra half innings reported here occurred in 8971 games, an average of 4.5 per game.

| Slot | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Extra | Total |
|------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| 1 | 191509 | 1992 | 36609 | 16260 | 25562 | 26305 | 20948 | 24872 | 15844 | 5063 | 364964 |
| 2 | 0 | 616 | 20841 | 26610 | 16393 | 26203 | 17268 | 23309 | 15157 | 4661 | 151058 |
| 3 | 0 | 273 | 12578 | 33101 | 11886 | 25889 | 17936 | 22164 | 16361 | 4660 | 144848 |
| 4 | 0 | 68494 | 6678 | 31786 | 11935 | 23353 | 19707 | 19798 | 16758 | 4167 | 202676 |
| 5 | 0 | 52683 | 3957 | 27726 | 17365 | 20528 | 22067 | 18761 | 17514 | 4288 | 184889 |
| 6 | 0 | 34656 | 2205 | 22373 | 24039 | 17558 | 24138 | 18885 | 18032 | 4523 | 166409 |
| 7 | 0 | 19377 | 26877 | 16194 | 28312 | 15324 | 24361 | 19459 | 17348 | 4258 | 171510 |
| 8 | 1 | 9268 | 41042 | 10857 | 29527 | 16627 | 23642 | 20885 | 16138 | 4478 | 172465 |
| 9 | 0 | 4151 | 40723 | 6603 | 26484 | 19526 | 20999 | 22723 | 15239 | 4590 | 161038 |

Table 4. Percentage of times each lineup slot batted first in an inning, 1957-2005.

| Slot | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Total |
|------|-----|------|------|------|------|------|------|------|------|------|-------|
| 1 | 100 | 1.0 | 19.1 | 8.5 | 13.3 | 13.7 | 11.0 | 13.0 | 10.7 | 12.4 | 21.2 |
| 2 | 0 | 0.3 | 10.9 | 13.9 | 8.6 | 13.7 | 9.0 | 12.2 | 10.2 | 11.5 | 8.8 |
| 3 | 0 | 0.1 | 6.6 | 17.3 | 6.2 | 13.5 | 9.4 | 11.6 | 11.0 | 11.5 | 8.4 |
| 4 | 0 | 35.8 | 3.5 | 16.6 | 6.2 | 12.2 | 10.3 | 10.4 | 11.3 | 10.2 | 11.8 |
| 5 | 0 | 27.5 | 2.1 | 14.5 | 9.1 | 10.7 | 11.5 | 9.8 | 11.8 | 10.5 | 10.8 |
| 6 | 0 | 18.1 | 1.2 | 11.7 | 12.6 | 9.2 | 12.6 | 9.9 | 12.2 | 11.1 | 9.7 |
| 7 | 0 | 10.1 | 14.0 | 8.5 | 14.8 | 8.0 | 12.8 | 10.2 | 11.7 | 10.5 | 10.0 |
| 8 | 0 | 4.8 | 21.4 | 5.7 | 15.4 | 8.7 | 12.4 | 10.9 | 10.9 | 11.0 | 10.0 |
| 9 | 0 | 2.2 | 21.3 | 3.4 | 13.8 | 10.2 | 11.0 | 11.9 | 10.3 | 11.3 | 9.4 |

There are several interesting features here. First of all is the surprise that there was actually a first inning in which the batter in the 8th slot batted first! This occurred on June 9, 1961 (game 2) in Boston when the Angels had a batting out of order situation which resulted in the 1st and 8th place batters swapping places the first time through the order. The lineup slot which bats first most often in an inning moves around the lineup fairly smoothly as the game progresses. The highlighted boxes in Figure 4 indicate the three lineup slots which lead off each inning most frequently. Those marked in red are those in which the three most frequent slots to bat first are at least 40% of the total. Those in yellow are less than 40%. In the second inning, the leading three slots (number 4 through 6) comprise over 81% of the total and by the ninth inning, the three most frequent are just under 36%, reflecting the expected randomization in the first slot as the game proceeds.

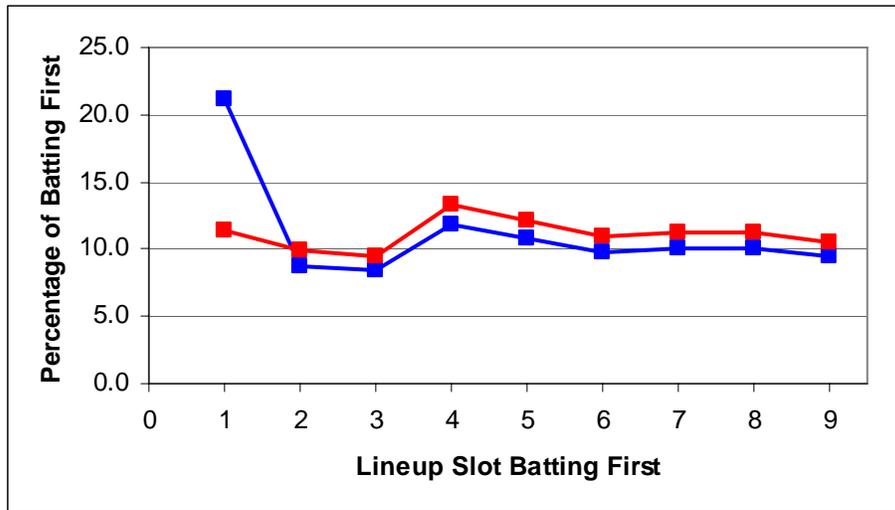
It is not surprising that the 1st three men in the lineup are quite unlikely to lead off the 2nd inning, but the low totals for the number 4, 5, and 6 men to lead off the 3rd are perhaps a bit unexpected. Finally, note that the clear leader in times leading off the second inning is the fourth batter in the lineup. This takes us to a related concept, which is something of a mirror image of which batter leads off an inning, namely which position makes the last out (that statement does not do justice to innings which end on caught stealing or pickoff plays). This means that the most common single result of the first inning (35.8% of the time) is that the side is retired in order or one runner reached base and was retired on a double play, caught stealing or pickoff. This is true even though the scoring rate is highest in this inning.

The last column in Table 4 shows that batter in the leadoff slot is the first batter in 21.2% of all innings with the other 8 positions fairly evenly distributed in comparison. Of course, this predominance of the leadoff batter is greatly influenced by the first inning reality. If the first inning totals are removed and the numbers renormalized, then there is much evenness across the slots, as shown in Table 5. In fact, after the first inning the 4th and 5th slots in the starting lineup are the most likely to bat first. These percentages are seen more easily in Figure 3.

Table 5. Percentage of times each lineup slot led off after 1st inning.

| Slot | % of Total |
|------|------------|
| 1 | 11.3 |
| 2 | 9.9 |
| 3 | 9.5 |
| 4 | 13.3 |
| 5 | 12.1 |
| 6 | 10.9 |
| 7 | 11.2 |
| 8 | 11.3 |
| 9 | 10.5 |

Figure 3. Percentage of time each lineup slot batted first in all innings and in all innings after the first. Blue: all innings, Red, all innings after first.



The possibility of differences between the visiting and home teams must also be considered. I did that comparison and only one of the 90 entries in Table 4 showed even a one percent difference between visitors and home. That is, the fourth place slot batted first in the second inning for the visiting team 37.5% of the time and 34.0% of the time for the home team. This means that the visitors went out in order in the first inning more often than the home team did. This is reflected in the higher average number of first inning plate appearances for the home team: 4.42 as compared to 4.28 for the visitors. This percentage difference appears small, but does represent over 13,000 more home team batters in the first inning over the 49 years studied.

Of course, the most important measurement is scoring. Table 6 presents the average number of run scored in each inning when the first batter in the inning was in the indicated lineup slot. The numbers here are the averages across all innings. As expected, I found some variation between individual innings, but I am not reporting those details here for two reasons. First, as we have seen, each inning/batter combination occurs with greatly varied frequency, so comparisons become less valuable. For example, the batter in the 3rd slot bats first in the 2nd inning only 0.1% of the time, so the runs scored in those

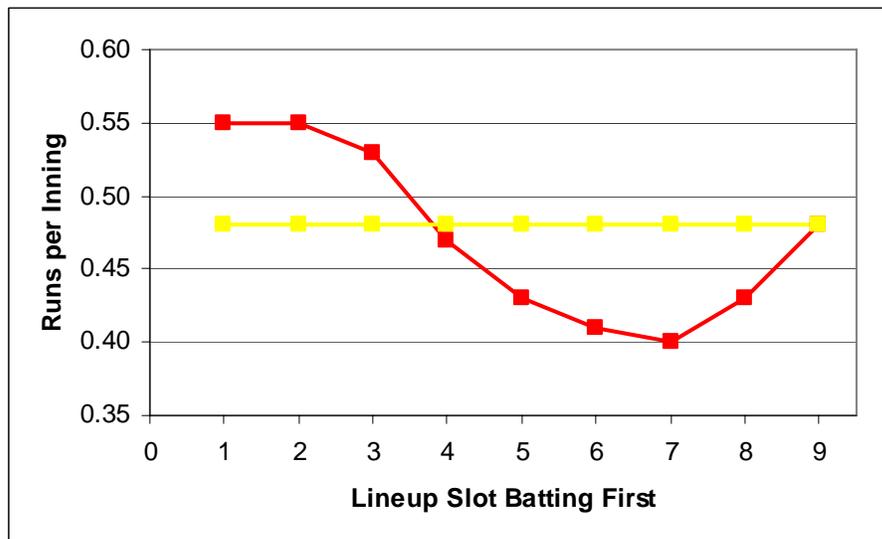
innings don't have much meaning in a comparative sense. Second, tables with 90 data points in them are cumbersome and not easy to read.

Table 6. Average runs scored per inning when indicated batter batted first.

| Slot | 1957-2005 |
|------|-----------|
| 1 | 0.55 |
| 2 | 0.55 |
| 3 | 0.53 |
| 4 | 0.47 |
| 5 | 0.43 |
| 6 | 0.41 |
| 7 | 0.40 |
| 8 | 0.43 |
| 9 | 0.48 |

To enhance the ease of viewing, I converted the data in Table 6 to the graphical format shown in Figure 4. This will be the standard format for the bulk of the presentation.

Figure 4. Average runs scored per inning when first batter occupied indicated lineup slot. Average for all innings shown by horizontal line.



There is a remarkable relationship between the first batter in an inning and the chance of scoring. The first two slots in the lineup are equally valuable as first batters in terms of scoring, followed by a steady but non-linear decline to the 7th place batter, ending with a definite upturn for the last two spots. This pattern may be unexpected at first, especially the observation that when the men in the 4th and 5th place slots bat first, the result is virtually the same as when the inning begins with the 9th and 8th place batters, respectively. However, we must remember that the scoring is the result of everyone who bats in that inning, not just the first batter. On average across all innings there are about 4.3 batters per inning (data not shown). Therefore, when the 4th slot leads off, then the 8th and 9th slots bat as the 4th and 5th men that inning. On the other hand, when the 8th slot

batter leads off, then the inning gets back to the top of the order in three batters which leads to more scoring, even if the man starting the inning made an out. I wanted to be sure that the high results when the leadoff man bats first were not overwhelmed by the first inning effect. When those totals are removed, then the average number of runs scored when the leadoff man bats first in an inning other than the first is 0.56, indistinguishable from the data in Table 6.

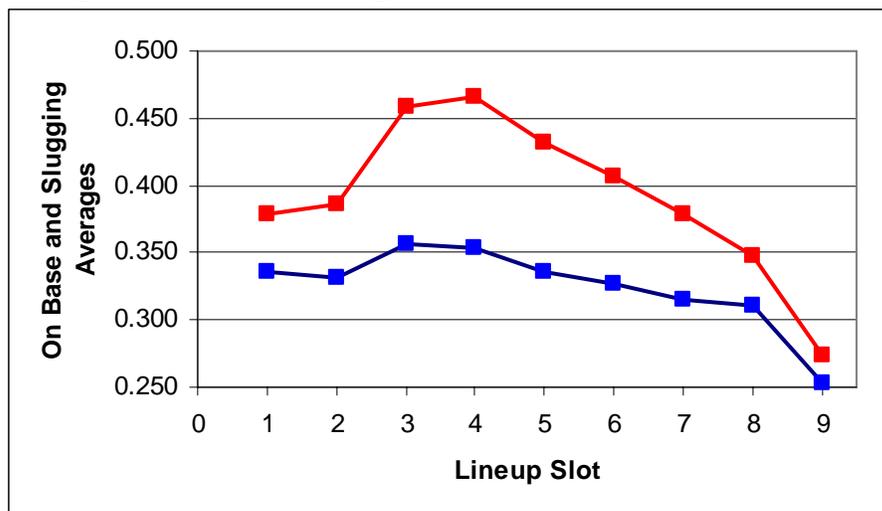
I wish to emphasize that the numbers in Table 6 and Figure 4 represent the combined effect of all the batters in an inning, simply separated by lineup slot of the first batter. The actual performance of batters in the individual slots is shown in Table 7.

Table 7. Batting performance by lineup slot, all games, 1957-2005.

| | OA | SA |
|---------|-------|-------|
| 1 | 0.336 | 0.378 |
| 2 | 0.331 | 0.386 |
| 3 | 0.357 | 0.459 |
| 4 | 0.353 | 0.467 |
| 5 | 0.336 | 0.431 |
| 6 | 0.326 | 0.407 |
| 7 | 0.316 | 0.379 |
| 8 | 0.310 | 0.348 |
| 9 | 0.254 | 0.274 |
| Average | 0.326 | 0.394 |

Figure 5 contains the same information graphically, showing clearly that the performance of each lineup position in isolation follows the expected pattern and is quite different from the batting order analysis above.

Figure 5. Batting performance by lineup slot, all games, 1957-2005. Red: Slugging Average, Blue: on base average.



The next variable I looked at was the effect of the DH. There are ample data from several sources that AL teams score more than NL teams since 1973, but it is very clear

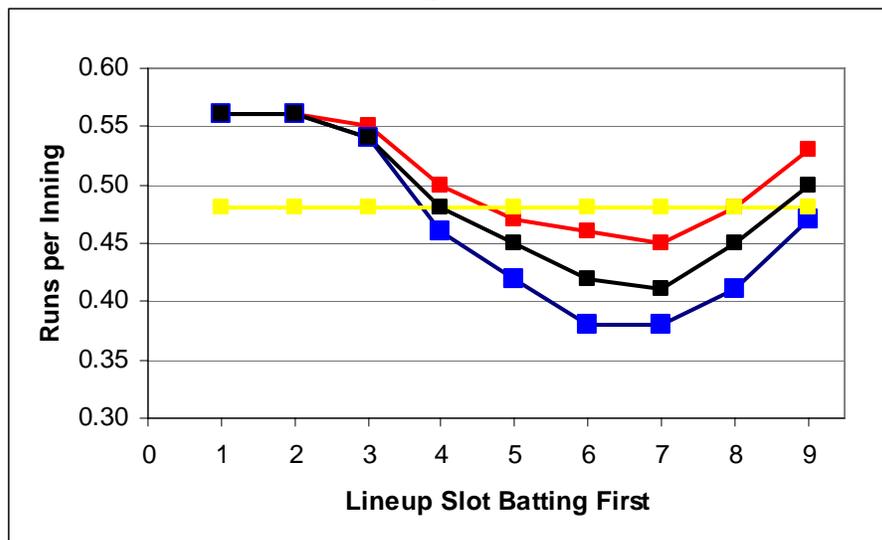
that the difference in totals is accounted for entirely by the DH (see, for example, Table 3 in <http://www.retrosheet.org/Research/SmithD/batlearn.pdf>). Table 8 presents the differences in run scoring between the two leagues. The first column is repeated from Table 6.

Table 8. Average runs per inning, with and without the DH (AL and NL home teams, respectively) in relation to first batter of inning.

| Slot | 1957-2005 | DH | No DH | 1973-2005 |
|------|-----------|------|-------|-----------|
| 1 | 0.55 | 0.56 | 0.56 | 0.56 |
| 2 | 0.55 | 0.56 | 0.56 | 0.56 |
| 3 | 0.53 | 0.55 | 0.54 | 0.54 |
| 4 | 0.47 | 0.50 | 0.46 | 0.48 |
| 5 | 0.43 | 0.47 | 0.42 | 0.45 |
| 6 | 0.41 | 0.46 | 0.38 | 0.42 |
| 7 | 0.40 | 0.45 | 0.38 | 0.41 |
| 8 | 0.43 | 0.48 | 0.41 | 0.45 |
| 9 | 0.48 | 0.53 | 0.47 | 0.50 |

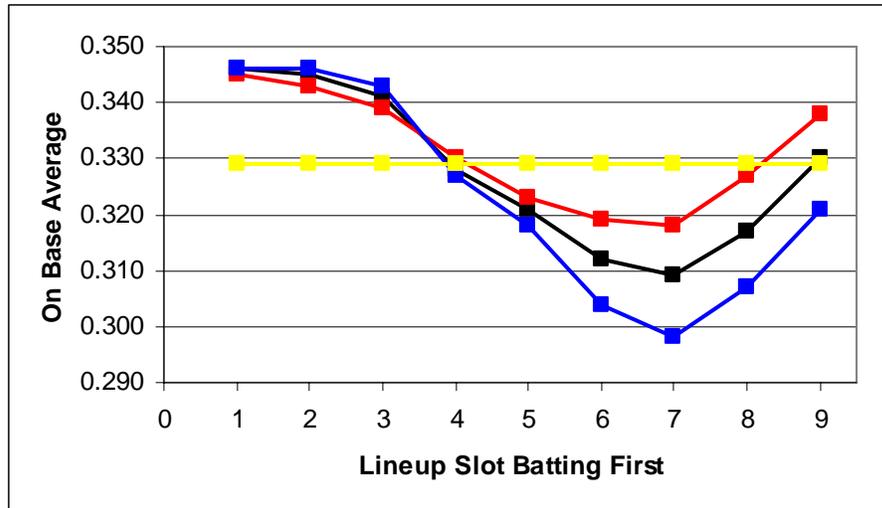
Figure 6 contains the same information graphically and shows the striking effect of substituting the DH for the pitcher, with a major effect of which lineup slot bats first. For the leadoff and number two slots, the leagues are the same, because innings that begin this way are very unlikely to have the pitcher bat. Differences appear through the rest of the lineup, with the widest discrepancy being found when the men in the 6th and 7th slots bat first and the pitcher will very likely come to the plate in NL games and thereby lower scoring in those innings. The overall level of scoring is slightly higher since the advent of the DH (data not shown).

Figure 6. Average runs scored per inning in games with and without the DH in relation to lineup slot batting first in an inning. Red: DH 1973-2005, Blue: No DH 1973-2005, Black: Combined, Yellow: Average across all slots.



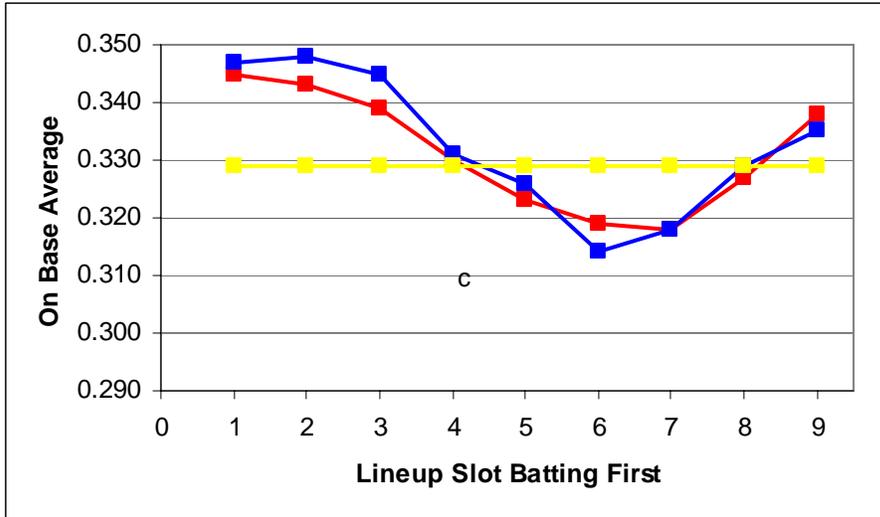
I then investigated how the scoring difference arose by looking at on base and slugging averages for the teams as a function of which lineup slot bats first. Figure 7 has the on base data which show even fewer differences for first 5 slots than the scoring did. However, there are wide differences for innings that begin with lineup spots 6 through 9, with the 7th slot showing the biggest difference.

Figure 7. On base average in each inning when indicated lineup slot bats first with and without DH, 1973-2005. Red: with DH, Blue: No DH, Black: combined, Yellow, average across all slots.



As noted above, almost all production differences between the leagues since 1973 are due to pitcher batting. I therefore recalculated the data of Figure 6 by omitting all plate appearances by pitchers. The results are in Figure 7, where we see that there is now some separation for innings that start with the 2nd and 3rd slots, but that the differences have almost completely disappeared from the 4th slot to the 9th.

Figure 8. On base average in each inning when indicated lineup slot bats first with and without DH, 1973-2005, pitcher appearances removed. Red: DH, Blue: No DH, Yellow: average across all slots.



To finish the study of the DH, I prepared the parallel two figures for slugging average. These are Figures 9 and 10.

Figure 9. Slugging average in each inning when indicated lineup slot bats first with and without DH, 1973-2005. Red: with DH, Blue: No DH, Black: combined, Yellow, average across all slots.

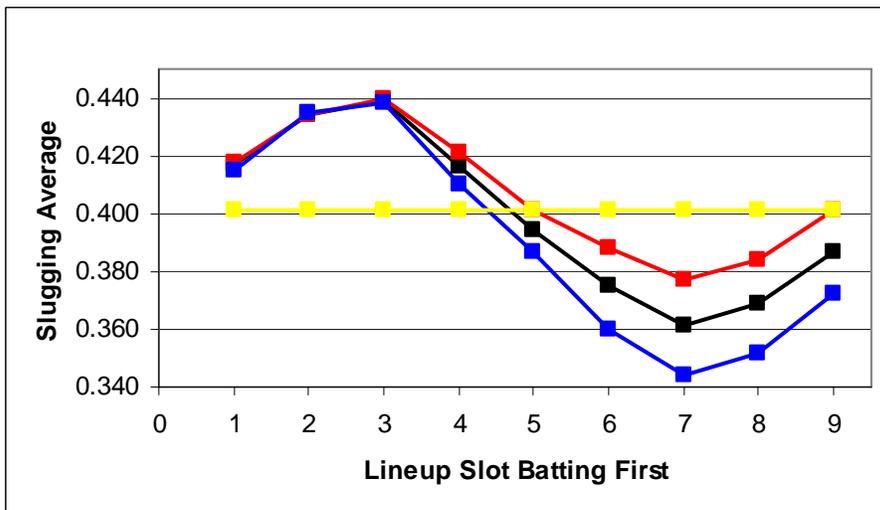
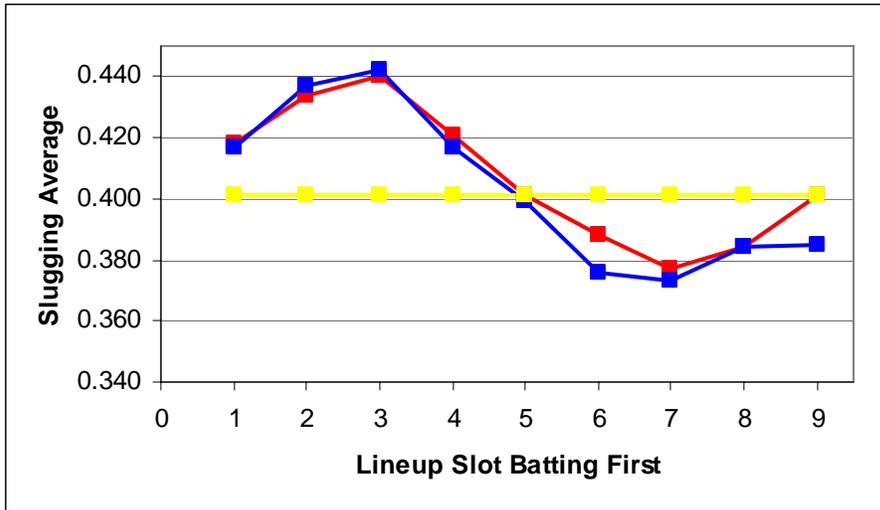


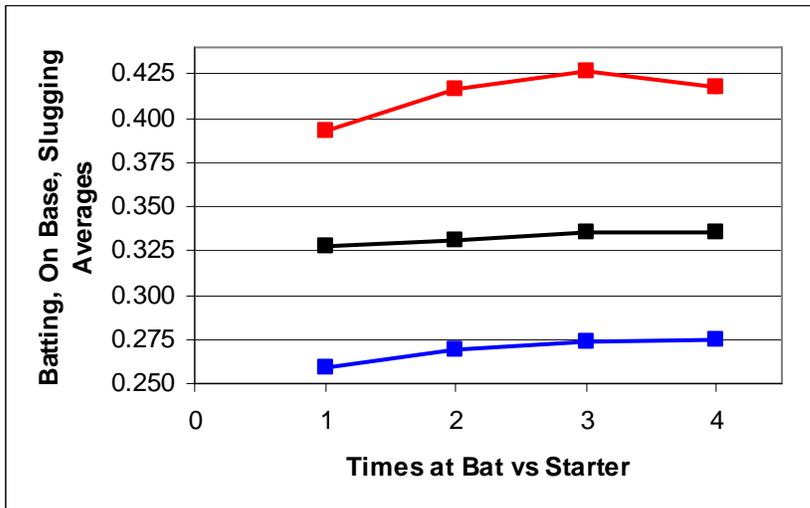
Figure 10. Slugging average in each inning when indicated batter bats first with and without DH, 1973-2005, pitcher appearances removed. Red: DH, Blue: No DH, Yellow: average across all slots.



Once again, the removal of pitcher appearances brings the two leagues very close together, even closer than was seen for on base average. The largest difference is found for the 9th slot, with the low NL value representing a large number of pinch-hitters who have a collective slugging average of .329 over these seasons.

There is one more topic I wish to discuss and that is related to some work I did previously on batter learning. I made a presentation on this at the SABR meeting in Kansas City in 1996 and published an expanded version in the Baseball Research Journal in 2006. Briefly, what I discovered is that there is a dramatic improvement for starting batters against starting pitchers in batting average, on base average and slugging average between their 1st, 2nd and 3rd times at bat. Figure 11 is a summary of these results:

Figure 11. Batting performance by starters in different times at bat, 1960-2005. Blue: Batting Average, Black: On Base Average: Red: Slugging Average.



At first glance, it appears that the scoring and batting order results I presented today do not agree with the batter learning study. As shown above in Figure 1, there is much more scoring in the first inning than in any other, followed by a great decline in the second inning. How can this be if batters are producing at a much lower rate their first time up, especially in terms of slugging average? The answer turns out to be simple and perhaps obvious. There is no disagreement and the apparent differences can be explained entirely by which batters typically bat in the first. Table 9 presents the percentage of plate appearances for each lineup position in each inning.

Table 9. Percentage of plate appearances by starting batters in each lineup slot in each inning.

| Slot | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Extra |
|------|------|------|------|------|------|------|------|------|------|-------|
| 1 | 23.4 | 6.1 | 19.0 | 6.4 | 16.0 | 9.5 | 12.1 | 15.7 | 8.8 | 14.7 |
| 2 | 22.9 | 3.3 | 18.2 | 7.9 | 14.2 | 11.9 | 9.5 | 16.2 | 10.1 | 13.2 |
| 3 | 23.3 | 1.6 | 15.4 | 11.1 | 11.5 | 14.2 | 8.3 | 15.2 | 12.7 | 12.7 |
| 4 | 15.0 | 10.1 | 11.7 | 14.3 | 9.1 | 15.2 | 9.0 | 12.6 | 14.7 | 11.2 |
| 5 | 8.5 | 16.9 | 8.0 | 16.0 | 8.2 | 14.7 | 11.1 | 10.0 | 15.6 | 11.0 |
| 6 | 4.3 | 21.4 | 4.8 | 15.9 | 9.2 | 12.9 | 13.8 | 8.5 | 14.9 | 11.1 |
| 7 | 1.9 | 20.4 | 6.3 | 14.1 | 11.7 | 10.3 | 15.6 | 8.6 | 12.4 | 11.9 |
| 8 | 0.7 | 16.0 | 10.7 | 11.1 | 14.3 | 8.3 | 15.3 | 9.3 | 8.6 | 11.1 |
| 9 | 0.1 | 4.1 | 5.9 | 3.1 | 5.9 | 3.0 | 5.3 | 3.9 | 2.1 | 3.2 |

These numbers differ quite a bit from those in Table 4, which showed the percentage of times that each batting order slot batted first. The bulk of appearances (84.5%) in the first inning are by the best batters (see Table 7) and in the second inning the weaker batters predominate (74.7% of total). In no other inning besides the first two is there such a disproportion in plate appearances between the different lineup slots. So, even though batters in all lineup positions do better the second and third times they bat, the absolute level of the most common batters in the second inning is very low. Since the numbers here represent starting batters, the percentage of batters from the 9th slot is low as the game progresses as starting pitchers are removed.

Conclusions:

1. The lineup slot of the first batter in an inning matters a great deal in a team's average scoring.
2. Lineups appear to be well designed in that the best scoring results are seen when the man in the leadoff slot bats first in any inning.
3. The lowest scoring per inning occurs when the 7th slot leads off the inning.
4. There is a DH effect, but it only affects the detailed rates, not the basic patterns.