Does Clutch Hitting Exist?

Pete Palmer, August, 2008

This is a reprieve of a famous study Dick Cramer did over 30 years ago, before the availability of play-by-play data. Eldon and Harlan Mills published a book in 1969 called Player Win Averages. They developed a method based on the changes in win probability before and after each at-bat. If the probability went up, a player got plus points, otherwise minus points. PWA was the sum of plus points over all points, so .500 was average. Although they never published anything for 1970, we were able to get a copy of the data. Dick came up with an approximation of what the PWA should be based on full season stats and compared it to the actual value. A higher PWA would mean the player did better in key situations, lower the opposite. Then he compared the results for the two years and found no correlation. A player who was above average one year was no more likely to be above the next year as anyone else. Thus the conclusion that clutch hitters did not exist.

Now thanks to Bill James, who started Project Scoresheet and Gary Gillette, who kept it going, plus Dave Smith and the others at Retrosheet, we have play-by-play of almost every game back to 1957. I had created a win probability program back in the 70s, which was based on actual run scoring from each base-out situation. I tabulated these probabilities for each league each year for the past 50 years and output data for each at-bat, showing the batter, pitcher, score, outs and runners before and after each play and the gain or loss in win probability and also gain or loss in potential run scoring. I also entered my plus or minus linear weight run value based on full season statistics.

I did study where I took all players with 500 appearances in a given year and measured their performance using OPS (on-base average plus slugging percentage) divided into 10 samples. The first test was by the last digit of the day of the game (0 to 9) and the second was by what I call stress level. I measure stress level by taking the average increase or decrease in win probability in any situation for a typical mix of batting events (out, single, double, triple, homer and walk) weighted by the expected frequencies. Some people call this leverage. This gives an average change in win probability for each situation based on inning, score, outs and runners on base. The most stressful situation is one run behind in the last of the 9th with 2 outs and the bases loaded. The average change in win probability is around 41%, depending on the season and league. An out is minus 30%, scoring one run is plus 37% and scoring 2 runs in plus 70%. A typical value is around 3%. I then sorted each player into 10 samples in order by stress level. I measured the standard deviation of batting average, on-base average, slugging percentage and OPS in both cases. The expected standard deviation for batting and average and on-base average is easy to calculate. It is simply the square root of p times q over n, where p is the probability of success, q is the probability of failure (equal to 1 minus p) and n is the number of samples. I worked with Trent McCotter on figuring slugging and OPS which are more difficult. For slugging, first you take p1 times q1 plus 4 times p2 times q2 plus 9 times p3 times q3 plus 16 times p4 times q4, where 1, 2, 3 and 4 are probabilities of a single, double, triple and homer. Then you subtract a bunch of terms for combinations of each, that is 4 times p1 x p2, 6 times p1 times p3, 8 times p1 times p4, 12 times p2 times p3, 16 times p2 times p4 and 24 times p3 times p4. Then you divide the whole thing by at-bats and take the square root. For ops, if both parts were independent, the standard deviation would be simply the square root of the sum of the squares of the two parts, but since slugging and on-base are highly correlated because hits are a major part of each, there is a more complicated formula, which I won't go into here, based on the covariance of the two, which ends up being about 20% higher than the simple case.

One problem with player win averages is that performance in a few key at-bats can outweigh the entire rest of the season. For example, on June 30, 2006, Adam Dunn hit a grand slam in the 9th with two out and his team 3 runs behind. He increased the team win probability from 9% to 100%, a gain of 91%. He season mark was 129%. He moved from 150th place to 82nd based on just one at bat. This is an extreme case, but picking up 40% or more in one at-bat is not unusual. This can't happen with scoring potential or full season stats. Anyway, since it takes about 10 runs to produce a win, if performance was equal in all situations, the full season linear weight and run potential values should be the same, about 10 times the win value.

Generally, the top 2 percent of appearances in stress level account for 10% of player win average, while the bottom 35% account for the bottom 10%, so this leads to distortions by over valuing some appearances and under valuing others.

At any rate, the two tests looked identical. What I did was calculate the expected standard deviation for each player based on his season stats. A slugger would have a higher variation in slugging percentage than a singles hitter. Then I divided each measured difference by the expected value. If you sum these figures (called z-scores) squared, then divide by the number of samples and take the square root, it should come out to one if the distribution is normal. Here are some typical values by season:

	random by date				by stress level			
	avg	oba	slg	ops	avg	oba	slg	ops
1957	.96	.99	.94	1.01	.97	.97	.96	1.01
1967	.96	.97	.95	1.01	.97	.98	.98	1.04
1977	.94	.95	.97	1.02	.96	.97	.96	1.02
1987	.95	.97	.94	.99	.96	.95	.94	.99
1997	.95	.96	.95	1.00	.92	.93	.91	.96
2007	.97	.98	.97	1.02	.92	.93	.92	.96

This shows that the two sets are quite similar. For the record, for 50 appearances in each sample, the typical standard deviation for batting average was about 63 points, which is the square root of .28 times .72 over 50. On-base average was also 63 points, slugging was 120 points and OPS 163. The last two are more volatile because they vary with the number of hits of each type.

The above study shows that the overall distribution of batting versus stress is pretty random. Now let's look at just the top 10% rated situations compared to the average. Here I took all players with at least 3000 appearances from 1957 to 2007 who had at least 200 highest stress situations combined in all the years that they had 500 total appearances. I found 770 players. I just looked at OPS in high stress versus the overall value. The average OPS for the top 10th of situations was .779, compared to .771 overall. This is misleading though because it is the result of a high number of intentional walks which raised on-base percentage by about 10 points. This chart showed 3 players beyond 3 sigma, and 33 beyond 2 sigma, where the expected number would have been 2 and 38. The leader in this case was a familiar name, Mickey Mantle. Mantle had an OPS of 1.155 in 634 stress situations, compared to .985 overall. The standard deviation for his sample of 569 appearances at the top 10th on the stress chart is 54 points, approximately 1/3 of the 50 appearance value. His OPS difference of 162 OPS points more than average divided by 54 gives a z-score of 2.98. This covers only 1957-

date, so it is only a partial for Mickey. Jim Ray Hart, with a z-score of -3.40 brought up the rear, along with Andy Van Slyke at -3.09 and Richard Hidalgo at -3.28. However, this study is based on small sample. Normally, statisticians like to say that when looking at one sample if the probability of something occurring by chance is only 5%, then they have 95% confidence that there is a real difference. But if you are looking at all samples, and find just 5% beyond 2 sigma, then this is just as expected and there is no reason to believe the 5% cases are unusual.

In what I consider a much better test, I looked at all players with 3000 plate appearances from 1957-2007, a total of 897. Here I compared their overall player win averages, expressed in wins above average times 10 runs per win, with their linear weight runs, I found a standard deviation of plus or minus 3 runs per 500 appearances. In order to figure what the random variation should be, I programmed a simple simulation with 18 batters all equal and played the equivalent of 5500 appearances, the same as the average of the sample. I got a standard deviation of about 2.5 runs per 500 appearances, just slightly less then the measured value. It may be that real life variation could be a little different from the simulated value, but at any rate, the two are pretty close.

Just for the record, there were two players outside the 3-sigma limit of 9 runs per 500 appearances. Probability theory says there should be 1 out of 400, so the results were not statistically significant. Scott Fletcher averaged 11 runs better in player win averages than linear weights, while Richard Hidalgo was 10 runs worse. Fletcher was only a little over one sigma above the mean in the previous test looking at only the top 10% of situations, since I counted all of his appearances, and he did quite well in his other above average situations. Hart's overall figure was just about average and Van Slyke was minus 2 runs per 500 appearances, just a little below. Hidalgo looked bad on both tests, but again, not outside the expected range.

Another study I did was to compare late/close performance to overall performance in general. Late/close is defined as 7th inning or later, either 1 run ahead, tied, or behind with the tying run on-base, at-bat or on-deck. This is typically about 15% of all at-bats. This test is less reliable than the player win average test because we are comparing only a small selection of appearances to the total rather than using all appearances with different weights. For 1957-2007, the overall performance was .715 OPS, while in late/close situations it was .704. However, weighting the late/close by overall season data, the expected value was .691, indicating that the better pitcher were more apt to be involved. However, again the increase in OPS over expected is due to the high number of IBB in these situations, about double the average, so there is no indication that batters or pitchers in general perform differently in late/close situations. The mix of batters in late/close situations showed a weighted OPS a few points higher than overall, which accounts for the rest of the difference.

Finally, I repeated Dick's correlation study, comparing player win average (expressed in wins over average) with linear weight runs divided by 10 from overall season performance. I used all players from 1957-2007 with at least 250 appearances in adjacent seasons. With over 8000 cases, and an average of 580 appearances in each season, I got a correlation between player win average and linear weight runs/10 of .002, which is next to nothing. When I correlated year to year overall OPS, I got .43, which is pretty high.

The standard deviation for batting average in one season of 580 appearances due to chance alone is about 18 points, for OBA 20 points, for slugging 35 points and for OPS it is 48 points. In the above test, the actual sigma for a single season in OPS was 106 points. If you assume that total is equal to the square root of skill squared plus luck squared, and luck is 48 points, then that leaves 93 points for skill. When comparing two seasons, you have to multiply by the square root of 2 to get the expected sigma due to chance, so it increases to 68 points. The total variation was 81 points, which leaves 44 points actual change in skill from one season to the next.

I did a study of 819 players from 1957-2006 in their first 100 at-bats in their first year. I expected to see an improvement in batting average, but actually there was none. The data showed an average of .260 with no significant difference through all 100 at-bats. This suggests that by the time a player reaches the majors, whatever nervousness there might be has been overcome, and their is no reason to believe that even rookies would be affected by clutch situations.

So if there was such a thing as a clutch hitter, the question would be, why isn't he trying as hard in all situations as he does when the game is on the line. Actually, we believe there is no such thing as a clutch hitter, and that players are trying all the time, although of course there is some variation due other factors, the overall grind of the season, injuries, etc. A typical major leaguer spends less than 40 hours in a season in the batters box, so applying oneself for that amount of time should not be a great strain. If you come up with the bases empty and none out, getting on first should lead to about three times as many runs as making an out, so there is ample incentive for almost every at-bat.