Not All .300 Seasons are Created Equal – A Permutation Method to Measure Clutch Performance

Shaun Lysen - slysen@wharton.upenn.edu
Department of Statistics, University of Pennsylvania – Wharton School, Philadelphia, Pennsylvania 19104

Introduction

The 2005 American League MVP race was one of the more controversial races in recent history. Alex Rodriguez of the New York Yankees was named the American League MVP despite the belief by many that David Ortiz of the Boston Red Sox should have been. Specifically, the critics noted that Rodriguez often got his hits at less important times, e.g. when no one is on base. Ortiz, on the other hand, got his hits at the most important opportunities, which would in some sense, make him more "valuable" to his team. We will investigate this claim to see who got his hits at more important times.

We will accomplish this by first casting the batting average in a 0/1 loss function framework. Next, we will propose a more sensible loss function based on the incremental expected runs scored over getting out given the current state of the game. This will create different at bat weights for the 24 different states of the game. These weights will be used in calculating the Expected Runs Batting Average (ERBA)

Now with differentially weighted at bats, we can randomly permute which at bats the batter got his hits and compute a new ERBA. By simulating this 1000s of times, we create a null distribution corresponding to the null hypothesis that who is on base and the number of outs has no effect on hitting.

Statistical Framework

Batting averages can be viewed through a loss function framework. Specifically, define a loss function that is 1 when getting out and 0 when getting a hit. Then,

$$BA = 1 - \hat{L} \quad \text{where } \hat{L} = \frac{1}{n}\sum_{i=1}^{n} L(AB_i)$$

That is, the batting average is simply 1 minus the empirical risk This treats getting an out the same regardless of the situation. Thus, a 0/1 loss function is not sensible if we want to take into account the fact that some at bats are more important than others. Clearly, hitting with the bases loaded with 2 outs has more of an impact than hitting with no one on with 2 outs.

Now let us introduce the state-space model for baseball. Note that there can either be a runner on each base or not on each of the three bases for 2 x 2 x 2 = 8 configurations for each of 0, 1, or 2 outs for a total of 8 x 3 = 24 states. For each state, we have the expected number of runs that would score in that inning

		Number of Outs		
		0	1	2
Runners	None	0.49	0.27	0.10
	1st	0.85	0.51	0.23
	2nd	1.11	0.68	0.31
	3rd	1.30	0.94	0.38
	1st,2nd	1.39	0.86	0.42
	1st,3rd	1.62	1.11	0.48
	2nd,3rd	1.76	1.32	0.52
	Loaded	2.15	1.39	0.65

Table 1: Expected Runs Scored in an Inning for the Given Current State of the Game. Source: Albert (2001)

A More Sensible Loss Function

At most stages of a baseball game, a team's objective function is to maximize expected runs. Therefore, we propose a loss function that is a measure of what is truly lost by getting out, expected runs scored.

$$L_{ERBA}(AB_i) = \begin{cases} ER(s,o) - ER(s,o+1) & \text{if } AB_i = out \\ 0 & \text{otherwise} \end{cases}$$

With ER(s,o) denoting expected runs scored corresponding to current baserunner state s, and number of outs o.

This is equivalent to saying the weight for an at bat is proportional to the expected runs lost if the batter were to get out.

Expected Runs Batting Average (ERBA)

The overall layout for this section can, and probably *should*, be
Now that we have our new loss function, we define the Expected Runs Batting Average as

$$ERBA = 1 - \hat{L}_{ERBA} \quad \text{where } \hat{L}_{ERBA} = \frac{\sum_{i=1}^{n} L_{ERBA}(AB_i)}{\sum L_{ERBA}(out)}$$

We note this is analogous to the previous batting average where we sum the losses for each at bat and divide this by the total loss possible. In the previous loss function, since each at bat had loss 1, the total loss possible was n.

Denote $w_i(s,o) = L_{ERBA}(s,o)$ for the i^{th} at bat

Then,

$$ERBA = \frac{\sum_{i=1}^{n} w_i(s_i, o_i) I(AB_i \neq out)}{\sum_{i=1}^{n} w_i(s_i, o_i)}$$

So we simply add up the at bat weights for each hit and divide by the sum of the at bat weights for all at bats

Clutch Performance via Permutation Methods

Now with different weights for different situations, we can propose a general method to evaluate just how good an observed statistic is. We will fix the number of at bats and the number of hits that a batter has and randomly permute which at bats the hits occur.

This corresponds to testing the null hypothesis that the current baserunner configuration and number of outs do not affect batter performance. By running this permutation method 1000s of times, we can create the entire null distribution and calculate a p-value.
This p-value is in some sense, a **measure of clutch.**

2005 AL MVP Race
Alex Rodriguez vs. David Ortiz

The motivation for this idea came from reading sportswriter after sportswriter who believed that David Ortiz was the true MVP for the 2005 season. They contended that Ortiz got his hits at the most "valuable" times to his team. The loss function that we propose better captures this idea since it treats at bats where more runs can be gained or lost on average are weighted more heavily.

We ran this permutation 100,000 times for each player and here are the corresponding histograms

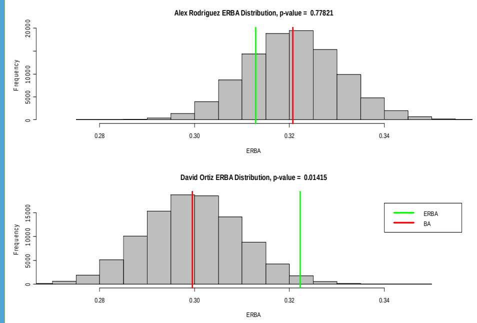

Figure 1: Null Distributions for Alex Rodriguez and David Ortiz – 2005 Season. Red line is the actual batting average while the green line is the ERBA

We see that Ortiz almost could not have done any better as to when he got his hits, while Rodriguez was below average. Even though Rodriguez had a higher BA by over 20 points, his ERBA was below Ortiz's by about 10 points.

Extensions and Conclusions

While this method does not say that a .250 hitter with a p-value of .01 is better than a .350 hitter with a p-value of .99, it allows us to understand just how good of a .250 or .350 season a batter had. When did the batter get his hits? Was it at times when an out would hurt his team the most?

We note that this permutation idea is quite general and could apply to any situation where there are differently weighted at bats. For instance, we could do a similar analysis with On Base Percentage (OBP) or Slugging Percentage (SLG). We might even turn the tables and use this for relief pitchers to measure the importance of getting outs in various situations.

Note that all of this was done with the idea of expected runs scored. Some times during baseball games, this is not the objective function. Late in close games team want to maximize the chance of scoring a run. One extension might be to define the loss function with respect to a loss in probability of winning. This will extend the current idea by also taking into account the current inning and the score which this current idea does not.

We now consider some historical data, specifically, we examine the p-values for the MVPs from 1970 until 2006.

Year	AL MVP	pval	NL MVP	pval
1970	Boog Powell	0.1176	Johnny Bench	0.1908
1971	Vida Blue	*	Joe Torre	0.1636
1972	Dick Allen	0.0529	Johnny Bench	0.5359
1973	Reggie Jackson	0.0889	Pete Rose	0.4602
1974	Jeff Burroughs	0.0150	Steve Garvey	0.1362
1975	Fred Lynn	0.1096	Joe Morgan	0.0700
1976	Thurman Munson	0.1587	Joe Morgan	0.8325
1977	Rod Carew	0.2300	George Foster	0.5256
1978	Jim Rice	0.1266	Dave Parker	0.0650
1979	Don Baylor	0.0320	Keith Hernandez/Willie Stargell	.0971/.7677
1980	George Brett	0.0793	Mike Schmidt	0.0430
1981	Rollie Fingers	*	Mike Schmidt	0.0748
1982	Robin Yount	0.0266	Dale Murphy	0.2823
1983	Cal Ripken	0.6734	Dale Murphy	0.4489
1984	Willie Hernandez	*	Ryne Sandberg	0.1483
1985	Don Mattingley	0.2686	Willie McGee	0.3469
1986	Roger Clemens	*	Mike Schmidt	0.3354
1987	George Bell	0.5266	Andre Dawson	0.0680
1988	Jose Canseco	0.8160	Kirk Gibson	0.6788
1989	Robin Yount	0.0771	Kevin Mitchell	0.4950
1990	Rickey Henderson	0.9069	Barry Bonds	0.0670
1991	Cal Ripken	0.4197	Terry Pendleton	0.4391
1992	Dennis Eckersley	*	Barry Bonds	0.3061
1993	Frank Thomas	0.3448	Barry Bonds	0.0794
1994	Frank Thomas	0.8021	Jeff Bagwell	0.5001
1995	Mo Vaughn	0.0140	Barry Larkin	0.3201
1996	Juan Gonzalez	0.2968	Ken Caminiti	0.0382
1997	Ken Griffey Jr.	0.0596	Larry Walker	0.2065
1998	Juan Gonzalez	0.1370	Sammy Sosa	0.3748
1999	Ivan Rodriguez	**	Chipper Jones	**
2000	Jason Giambi	0.0226	Jeff Kent	0.0832
2001	Ichiro Suziki	0.0084	Barry Bonds	0.0307
2002	Miguel Tejada	0.0007	Barry Bonds	0.2003
2003	Alex Rodriguez	0.8134	Barry Bonds	0.3557
2004	Vladimir Guerrero	0.8506	Barry Bonds	0.3196
2005	Alex Rodriguez	0.7782	Albert Pujols	0.8240
2006	Justin Morneau	0.3660	Ryan Howard	0.8584

Table 2: p-values for MVPs from 1970 to 2006. Source: Retrosheet (note play by play data for 1999 is currently unavailable)

Literature cited

Albert, J. (2001) *Using Play by Play Baseball Data to Develop a Better Measure of Batting Performance*
Albert, J. and Bennett, J. (2001). Curve Ball
Retrosheet . http://retrosheet.org

Acknowledgments

I thank Professor Dylan Small, University of Pennsylvania for helping me with this poster process as well as my abstract and informing me about NESSIS. This is my first time doing this process.
I would also like to thank Professor Tom Cover, Stanford University who was my original inspiration for statistics in sports, both through his classes and one on one discussions.