Using Logistic Regression to Test for the Existence of Individual Differences in Clutch Hitting Ability

Summary: Tango's data was analyzed using logistic regression and no evidence of individual clutch hitting ability was found. The data can be explained adequately with a model that allows for different abilities for each batter and an effect for Leverage Index group such that on base average was lower in higher leverage index group situations. The results failed to validate those of Andrew Dolphin and Tango.

Intro:
To test for the existence of individual differences in clutch hitting ability, we will build a model to predict on base average (OBA) as a function of three effects:

1. Clutch situation Effect – It is harder (easier) to hit in clutch situations. Since clutch is measured by groupings of leverage index (LI), it will be assumed that is easier to get on base in low LI situations and harder to get on base in high LI situations.
2. Individual Batter Effect – Different batters have different levels of ability
3. Individual Batter Differences in Clutch Situations Effect – Batters may increase or decrease their relative ability when they bat in high LI situations.

The idea is to test whether the first two effects are sufficient to explain OBA. If the third effect, individual batter differences in clutch situations, provides additional predictive ability, then we will conclude that individual differences in clutch hitting exist. If the third effect doesn't provide additional information, then we can conclude that individual differences in clutch hitting don't exist or that they are too small to be detected with our current sample. Given that the current sample is approximately 340 PA's per batter per LI group, the effect would be so small as to be negligible.

> Null Hypothesis: OBA can be explained by an effect for clutch situations and an effect for individual batter ability.

> Alternate Hypothesis: Individual batters relative ability change in clutch situations and this effect provides information above and beyond the effect for individual batters and the effect for clutch situations.

Logistic Regression:

Logistic regression was used to test the effects. Logistic regression and probit regression are both related to linear regression, but are designed to handle dependent variables that

are categorical such as on base/not on base. Both force the predicted values to lie between 0 and 1 where 1 represents 100% certainty of an event and 0 represents 0% certainty of an event.

Logistic regression takes the form:

$$F(x)=\exp(x)/(1+\exp(x))$$

SAS Proc Logistic was used to perform maximum likelihood estimation of logistic regression.

Specifying the Effects:

Both the effect for LI and Batters are categorical and require the construction of dummy variables. If there K groups, then K-1 dummy variables must be created. Each dummy variable has a value of 0 or 1 and for each case there can only be one dummy variable that equals 1.

Since there are 5 LI groups 4 dummy variables must be created. There is a dummy variable for the 0,1,2 & 3 groups. If the dummy variable for the 0 group equals 1 then the group is 0. If all dummy variables equal 0 then group=4.

The same goes with the batters. Since there are 340 batters, 339 dummy variables must be created. If dummy variables equal 0 then the batter is number 340.

Specifying the effect for individual differences in clutch hitting ability is a little more complex. Both of the previous effects are what we call main effects because they involve only one variable (although categorical) and these effects are constant regardless of the values of other variables. In order to model this effect, four more dummy variables for each batter must be created so that there are five dummy variables for each batter, one for each LI group. This can be done by multiplying the dummy variables for LI group by the dummy variables for the batters. This creates 1,356 new dummy variables. When terms are multiplied, the resulting effect is called an interaction effect because two or more variables are interacting together.

Testing the Effects:

We now have 1,699 dummy variables grouped into three effects. Fortunately we have 584,963 cases which give us an average of 344 cases per variable.

SAS tests each effect separately by grouping the dummy variables and performing a test that at least one of the dummy variables in the group is significantly different from 0.

This is a chi-square with degrees of freedom equaling the number of dummy variables involved. The following is the result from the logistic regression.

Table 1 Logistic Regression All three Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
lig	4	5.3840	0.2501
batterid	339	854.1708	<.0001
lig*batterid	1356	1381.7428	0.3071

Lig represents the test for the five groups of leverage index. The last column is of particular interest. It represents the probability of getting a result this extreme due to chance. Usually if this value is below .05 then we say that it is statistically significant meaning that it is probably a real effect. In this case we can't conclude that LI grouping is statistically significant controlling for batter and the interaction term. The effect for batter is highly significant with Pr <.0001.

The lig*batterid is the test that we are most concerned about. If some batters hit better in the clutch relative to their nonclutch performance, then this effect should be statistically significant. It isn't close to being statistically significant. With four years of data and over half a million cases, the logistic regression couldn't find any evidence that some batters hit relatively better in the clutch.

Wait a minute why isn't the effect for LI groups significant? The reason is that the effect for lig*batterid is correlated to the effects for lig and batterid. These tests are testing whether an effect provides unique information that other effects don't provide. LI groups and the interaction term are canceling each other out. The effect for batter is significant because it's much stronger. Rerunning the logistic regression without the interaction term tells a different story.

Table 2 Logistic Regression Main Effects only

Effect	DF	Wald Chi-Square	Pr > ChiSq
lig	4	32.8416	<.0001
batterid	339	2547.4703	<.0001

In this analysis, both effects are strongly significant at Pr <.0001.

Conclusion:

This analysis found absolutely no evidence that some batters hit better in the clutch relative to their non clutch performance. It doesn't prove that clutch hitting

doesn't exist, no test can do that. It does strongly suggest that if clutch hitting does exist, it's so small as to be negligible.

The null hypothesis was that individual differences in batting ability and the LI group could fit the data as well as a more complicated model that includes an effect for differences in clutch hitting. The alternate hypothesis was that a model with an effect for differences in clutch hitting would fit the data better. This effect was not found to be statistically significant or even close, despite the fact that four years of data was used that included over a half a million PA's. As such the results of this analysis stand in strong contradiction to the results of Andrew Dolphin and Tango, who's data was used in this analysis.