***Extending our Analysis: The Clutch Factor***

*Introduction*

After completing our analysis of the Hot Hand theory, we wanted to apply a similar analysis to another theory in basketball, the Clutch Factor. There is a perception among players and fans alike that certain players elevate their performance in important moments, such as when the score is close or the game is in its final minutes. The idea is that “clutch” players are more likely to make these big shots in these critical moments. This is a similar perception to the Hot Hand with streakiness, and therefore we believe that a similar analysis will allow us to explore if the Clutch Factor is another fallacy, or may actually be true.

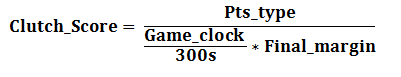
*Defining “Clutch\_score”*

For our analysis, we created a “*clutch\_score*” variable that measures the degree of how clutch a given shot is. Clutch can be defined as the following: During a critical moment of the game, how much will this shot improve the chances of my team winning?

In our given dataset, there are a few variables we believe define a clutch shot. These include

* **'game\_clock'**: How much time is left in the game? In the closing seconds of a game, a made shot would be more clutch. We define clutch time as any point in the game with < 5 minutes (300s) remaining in the 4th quarter, which is the same way the NBA defines it.
* **'pts\_type'**: What type of shot was taken? A 3-point shot would be more clutch than a 2-pointer since it affects your team's score more.
* **'final\_margin'**: What was the ending score of the game? Shots made in a close game are more clutch than in a game that has one team with a big lead. Ideally, we would've liked to use the margin of the score at the time of the game, but our dataset does not provide that information, so we used *final\_margin* instead. We use the absolute value of *final\_margin,* and define shots with a *final\_margin* > 6 as not clutch, since that is greater than a 2 possession game.

We can combine these variables to create our final ‘*clutch\_score*’, which we defined as:



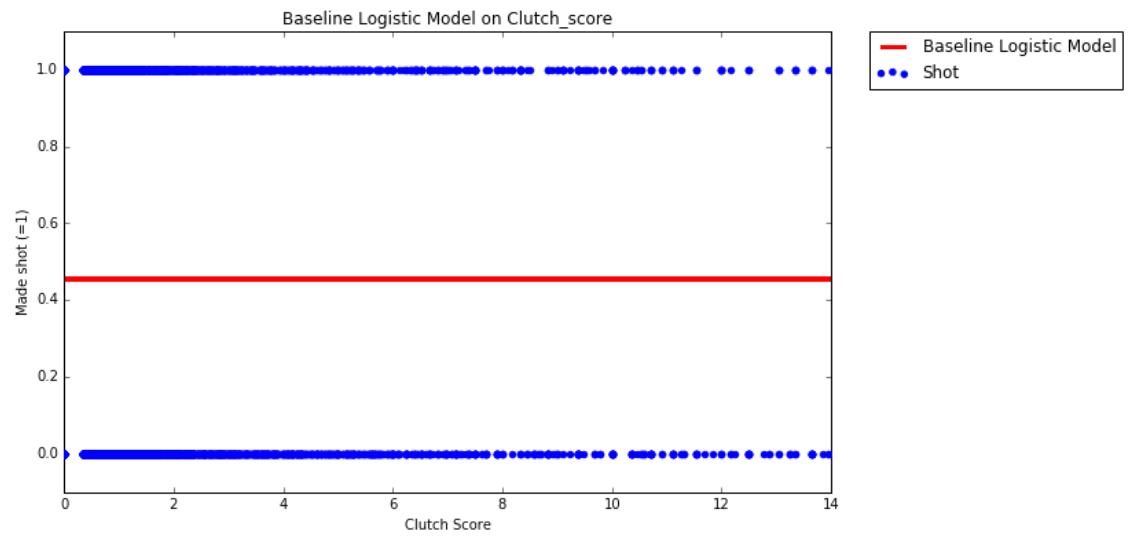
This gives us a dimensionless parameter and ensures that each parameter is given equal weight. A larger number translates to a more clutch shot. For all shots not taken in the 4th quarter with less 5 minutes left, *clutch\_score* = 0. Also, for all shots taken in a game with a *final\_margin* > 6, *clutch\_score* = 0.

*Fitting an Initial Model*

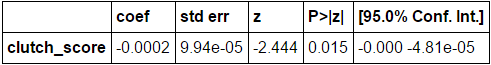
After calculating the *clutch\_score* for all shots in our dataset, we proceeded with fitting a model to our data. Our initial fitting would be testing across the entire NBA is *clutch\_score* was a positive predictor for a made shot. We first fit a logistic regression (just like for the Hot Hand) with the only predictor being *clutch\_score*. Our model gave us the following results:



We see that without any other predictors, the coefficient on *clutch\_score* is estimated to be essentially 0, suggesting it has no effect on if the shot is made. Further, the model barely performs better than a 50% classification rate, suggesting it’s hardly better than just flipping a coin. If we plot the logistic model vs. *clutch\_score*, we get the following:



From this graph we can see that our logistic model is essentially useless, and appears to be almost horizontal. This suggests that *clutch\_score* isn’t a very good predictor. We can also use the *statsmodel* package to gather more information about our *clutch\_score* predictor:



We see that the p-value for our coefficient is 0.015, which is less than the 0.05 cutoff, telling us that this predictor coefficient is significant. This shows us that by itself, *clutch\_score* is a significantly poor predictor of a made shot. It doesn’t appear that an increase in *clutch\_score* means a shot is more likely to go in from a league-wide standpoint.

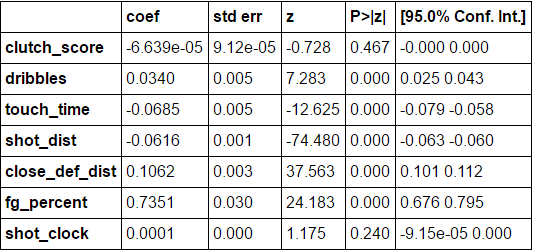
*Incorporating Confounding Variables*

One reason our initial model performs poorly is that we don’t take into account any factors that may be affecting difficulty of the shot or how talented the player who is shooting it is. We want to try to control for these variables, so therefore we will include them as predictors in our model.

Based on our dataset, the variables we’ll add are:

* **Dribbles**: number of dribbles taken before the shot
* **Touch Time:**  time holding the ball before the shot
* **Shot Distance:** how far away the shot was taken from
* **Closest Defender Distance:** how close the nearest defender was
* **Field Goal Percentage:** the shooting percentage for the given player
* **Shot Clock:** amount of time left on the shot clock

When we run a logistic regression model including all these variables along with *clutch\_score*, we get the following information about each variable:



We can examine the coefficient values (*‘coef’*) and p-values (*‘P>|z|’*) to see how each variable affects the probability of the shot going in and if it is a significant predictor. This new model shows us that the *clutch\_score* coefficient is again practically 0, suggesting that once again it is not a positive predictor for shot probability. Its p-value = 0.467, which is much greater than 0.05, showing us that it is not significant in this model. It appears that despite accounting for confounding variables, there is no support for a Clutch Factor for the league as a whole.

*Extending Analysis to Individual Players*

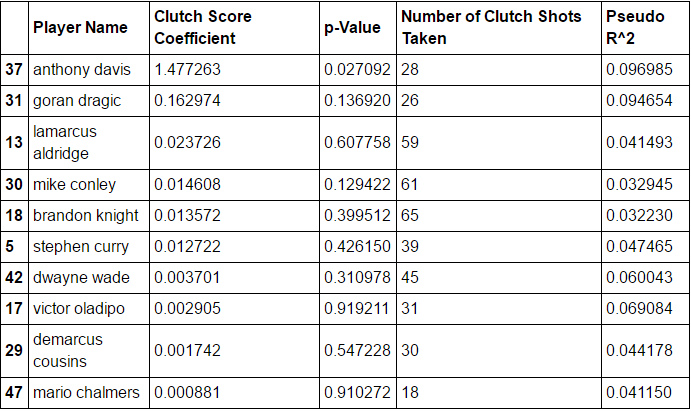
It is not surprising that the clutch factor doesn't appear to exist across the league according to our above analysis. The theory is that there are certain players that are clutch, not that everyone is more clutch. Thus, it is important to extend our analysis to individual players to see if there are individually clutch players, as the theory suggests. We evaluate the model on the Top 50 players, as determined by the same *player\_score* metric determined for the Hot Hand.

**Important Note:** As we evaluate the model on a large number of players, we have to keep in mind the issue of Multiple Comparisons, as we did for the Hot Hand analysis. This idea says that as we're comparing p-values across a large number of independent samples, there is a high chance of at least one having a p-value < 0.05 due to probability distribution alone and not necessarily due to significance. We have to adjust what our p-value cutoff for significance is. In this case, Multiple Comparisons across 50 samples says our proper p-value will be:

****

After fitting a different model for each of the Top 50 players, we can split them up into ‘Clutch Players’, aka players who had a positive coefficient on the *clutch\_score* predictor, and ‘Choke Artists’, aka players who had a negative coefficient on the *clutch\_score* predictor. We examine the Top 10 Clutch Players and Top 10 Choke Artists below.

**Top 10 Clutch Player Results:**

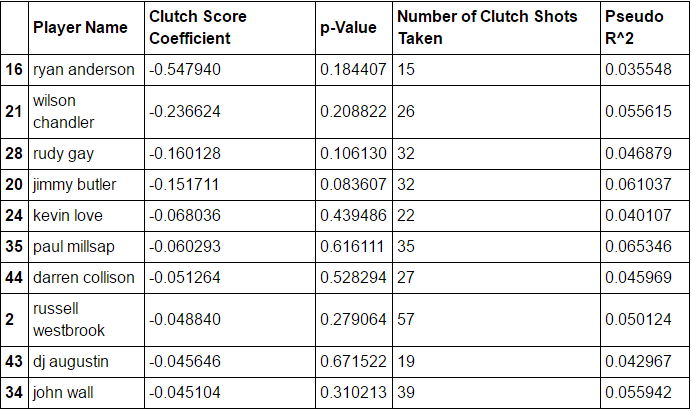


Based on these results, it appears that Anthony Davis is the most clutch of our Top 50 players, with a *clutch\_score* coefficient of 1.477. This suggests that every time we increase the *clutch\_score* of a shot by 1, Anthony Davis’ chance of making that shot increases by 1.477%. The p-value of this coefficient is 0.027092, which doesn’t pass our Multiple Comparisons adjusted criteria of p = 0.00103, so we cannot conclude that Anthony Davis’ model is significant. In fact, when we examine all of our results, we see that none of our players actually have a significant result for the *clutch\_score* coefficient.

In regards to the lack of significance of our results, we see that based on 'Number of Clutch Shots Taken', we are working with quite small sample sizes for our data and therefore it is difficult for our model to draw strong conclusions. If we were working with more data, we could further explore some other potentially clutch players for who we don't currently have enough data for.

With these findings, our model implementation suggests that there is no support for Clutch Factor. No player appears to be individually clutch.

**Top 10 Choke Artist Results:**

****

Looking at players with negative clutch score coefficients, we see that no p-values are < .00103 or even < .05, and therefore we can't determine our results to be significant. Thus for our data and model, we can't declare any player to definitely be a choke artist. *Clutch\_score* doesn’t appear to have any bearing on affecting a player’s shot probability negatively.

*Conclusions*

This analysis of Clutch Factor presents results that, based on our data and implementation of a clutch score, don't support the idea of players being clutch towards the ends of games. With more variables and shot data available, we could try to find stronger support for our findings and potentially find that there may be some players who could be called clutch. Like discussed before, a flaw of our data is that we don't have a variable for the game score at the time of the shot, but only the final margin, so we can't fully measure just how close a game is in a given instance. This information would provide better measurements of our *clutch\_score* and may help further inform our model. Further study should also be done by varying the definition of *clutch\_score*, as there may be a more accurate way to define how clutch a given shot is. For our definition of *clutch\_score*, we didn't find any support for the Clutch Factor theory, but other definitions of *clutch\_score* may provide different results.

For our analysis of the data, we can conclude that the Clutch Factor is just another fallacy.