A Universal Strategy for Measuring Clutch Performance in North American Sports

Brian Macdonald

Mentor: Dr. Ben Fitzpatrick, Clarence J. Wallen, S. J. Professor of Mathematics

Abstract:

The popular conception of athletes who consistently raise their performance in important moments, or “clutch” athletes, is a fixture in sports-related discussions. Fans, media, and players frequently refer to “clutchness” as a trait that certain athletes possess, but statisticians question this assumption. My project aims to determine the extent to which clutch players exist statistically in different sports. For each of the four major North American leagues, I will devise a scale for the importance of in-game situations based on the win probability a league-average player is expected to add to his team. Then, I will choose a performance metric and track it over ten seasons, determining if each player’s metric is correlated with the aforementioned importance scale. The intended result will be a percentage of clutch players in each league. This information could potentially help coaches and executives decide how much importance to ascribe to the invisible “clutchness” characteristic when evaluating players. It could also influence the discussion of clutch players among fans, media, and the general industry.

Introduction:

When professional athletes do well in key moments of games, they often gain the admiration of fans. The popular term used to describe these players is “clutch,” defined by Merriam-Webster as “successful in a crucial situation.” By contrast, athletes who appear to underperform in such moments are known as “choke” players or “chokers.” However, many sports statisticians have emphasized luck as a contributor to clutch performance, arguing that few to no athletes consistently raise their level of play in important scenarios. The resulting debate has pitted the faction of the statistical community that rejects “clutchness” against traditionalists, players, and other statisticians who affirm its existence (Verducci, 2004).

Though fans, media, players, and coaches fixate on clutchness, it is difficult to agree on a specific meaning for the trait, let alone quantify it statistically. However, I aim to develop a definitive framework for this purpose. My question is this: Using a universal statistical strategy, how can I measure the existence of clutch athletes across different sports? Inspired by Fuld (2005), I will place clutch performance in the context of overall performance, and I will rate the importance of in-game situations by the win probability an average player should add to his team. By correlating importance and performance, I wish to determine the prevalence of clutchness in the four major North American leagues.

Background:

For the most part, the “clutch debate” revolves around baseball, where Cramer first argues in 1977 that consistently timely hitters do not exist. While Cramer’s study utilizes only two years of data, Palmer (1990), Ruane (2005), and Birnbaum (2008) confirm his conclusion over a larger sample. However, Dolphin (2004) finds a moderate amount of clutch hitting, while Silver (2007) and Albert (2007) find a small amount. Bill James (2004) remains on the fence, arguing that Cramer’s study and similar works are incapable of disproving the existence of clutch hitters, and that new methods are needed to answer the question.

Usually, statisticians quantify clutchness by selecting a particular game state as “important.” (Fuld, 2005) Often, this takes the form of an in-game period occurring late in the contest and a range of scores wherein the game is close. Players’ clutchness is measured by their statistics under these conditions. In MLB, late-inning pressure situations (LIPS) include “any at-bat in the seventh inning or later where the batter's team trails by three runs or fewer, is tied or is ahead by only one run,” as well as bases-loaded at-bats when down by four. (Major League Baseball) Tango et al. (2007) use a zero-to-three-run deficit in the eighth inning or later. In the NBA, “clutch time” denotes the last five minutes of a game when the score is within five (National Basketball Association). Favale (2015) uses the last two minutes, while Schumaker et al. (2010) look at the last eight minutes. In American football, the fourth quarter is commonly used as the divider. To measure clutchness, Pro Football Focus analyzes fourth-quarter play when the score is within eight (Linsey, 2019), while Schatz (2009) uses seven points. In hockey, the term “close game” is occasionally used, denoting a game that is “within a goal during the first two periods, or tied in the third.” (Sporer, 2015)

These dividers are convenient, but they are arbitrary (Fuld, 2005). Realistically, there is no point in a game when it instantly becomes much more high-stakes. The seventh inning or the last five minutes are only chosen because they feel like appropriate numbers. Thus, I will use the entire game, but I will weigh each situation based on its individual importance. Win probability is the most definitive measurement of importance; the “clutchest” plays occur when the game is on the line. However, I am not asking how much a player improves his team’s win probability; I am asking how he performs when there is more win probability *to be gained*. Performance and importance must be measured separately (Fuld, 2005).

Another shortcoming of clutch performance studies is that they often track players’ high-leverage statistics without comparing them to players’ normal statistics. When I measure clutchness, I want to observe the extent to which athletes *raise* their performance (or possibly lower it) in critical situations, not simply how they perform in those spots. Otherwise, I risk confusing overall talent with clutch ability. For instance, Favale (2015) and Solomonov et al. (2015) measure player statistics in NBA “clutch time” but fail to compare them to players’ ordinary production. The results of Favale’s study are predictable: the best players appear the most clutch.

Methods:

In order to address the aforementioned flaws, I offer a definition of “clutch”: the extent to which a player raises performance as importance increases continuously (Fuld, 2005). By "raises," I mean that critical-moment statistics must be put in the perspective of overall statistics. "Importance" is determined by win probability on a continuous scale, while "performance" is determined by outcomes; the two measures must be distinct. I will use data from the 2010-2019 seasons of the MLB, NBA, NFL, and NHL, computing performance and importance for each league.

To calculate importance, I will employ Fuld’s method, but I will modify it for each sport. The importance of any game situation will be given by the formula:

*Importance Rating* = p1(w1 - w0) + p2(w2 - w0) +...+pn(wn - w0)

where n is the number of positive or beneficial outcomes being considered, wn is the team win probability given outcome n, w0 is the team win probability given a certain base outcome, or “bad outcome,” and pn is the league-wide relative frequency of outcome n (Fuld, 2005). This formula yields the expected increase in win probability an average player would produce in each situation. The values of pn can be calculated for each individual season (to account for year-to-year changes) or the entire ten-year sample (for expedience).

To calculate performance, I will choose a statistic from each sport that provides a numerical scale for the aforementioned outcomes. I will use percentage-based statistics, as weights in the numerator will denote performance while the denominator will tell me which “game events” (such as plate appearances or shots) to record for each player. Players will only qualify for the dataset if they have a minimum number of such game events between 2010 and 2019. For baseball, I will use On-Base Plus Slugging (OPS) (Major League Baseball). For basketball, I select Points Per Possession (PPP) (Fromal, 2012). For football, I will take Passer Rating (QBR) (Silver, 2010), and I will only include quarterbacks in my dataset. For hockey, I will use Expected Goals (xG) (Ryder, 2004) divided by Unblocked Shot Attempts (USAT) (Lee, 2020), and I will only include forwards.

Once I have collected the data, I will carry out a regression on importance and performance for each player. Again, I will either divide this process into seasons (Fuld, 2005) or perform it once across the full ten-year dataset. Fuld (2005) uses a Gompertz regression, but I will consider other models for each sport based on the shape of the data. The process will yield a correlation (r) for each player, with positive values pointing toward “clutch” players and negative values pointing toward “choke” players (Fuld, 2005). To see which correlations are significant, I will use a two-sided test with H0: r = 0 as my null hypothesis. Then, I will use the Benjamini-Hochberg method to identify significantly clutch and choke players; for this procedure, I will choose my false discovery rate based on the size of the datasets.

Expected Results:

The main result of the study should be a percentage (and named list) of clutch and choke players in each league over the ten-year sample. The percentages should provide insight into the prevalence of “clutch ability” as a whole; they should also indicate the relative presence of clutchness among the four sports.

Conclusion:

Though it is hard to define, clutchness is an integral part of the North American sports consciousness, a frequent driver of discussions and evaluations of athletes. My project will define and test a strategy for measuring clutch performance across baseball, basketball, American football, and hockey. Many studies on this subject draw arbitrary lines between “important” and “unimportant” situations, conflate performance and importance, or fail to compare clutch numbers to overall numbers; however, I will avoid these problems by measuring importance by the win probability an average player adds to his team, and measuring performance by a rate-based metric. From there, regression will reveal the players in each sport who exhibit a significant clutch ability over a ten-year sample. These results will have a chance to influence player evaluation in front offices; depending on the sport, executives will be encouraged to place more weight, less weight, or no weight at all on “clutch talent.” Additionally, the study will produce evidence in the unsettled “clutch debate” between traditionalists and statisticians, and it may change the way fans and media discuss players’ critical-moment performances.

Budget and Timeline:

The intended timeframe for my project is the spring 2022 semester. This project requires play-by-play data for the MLB, NBA, NFL, and NHL, not simply players’ total statistics. Between 2010 and 2019, I must record every plate appearance for every MLB player, every shot for every NBA player, every pass for every NFL quarterback, and every unblocked shot attempt for every NHL forward. While this data is readily available in the form of box scores, its sheer volume would make it impossible to record manually in one semester. Therefore, I must use downloadable data, and I will stick to the .csv format. MLB play-by-play data is free to download at <https://www.retrosheet.org/game.htm>. The requisite NBA play-by-play data can be purchased at <https://www.bigdataball.com/datasets/nba/play-by-play/?utm_source=nbastuffer&utm_campaign=glossary&utm_medium=website> for $300, at a rate of $30 per season. NFL play-by-play data from 2010 to 2018 can be found at <https://www.kaggle.com/maxhorowitz/nflplaybyplay2009to2016/version/6?select=NFL+Play+by+Play+2009-2018+%28v5%29.csv>, but to complete the set, 2019 data can be purchased from <https://www.bigdataball.com/datasets/nfl/play-by-play/> for $30. NHL play-by-play data, while harder to find, should be available for free here: <https://www.kaggle.com/martinellis/nhl-game-data>.

I will carry out the study over the semester’s 17 weeks, including spring break. Because my class schedule is more crowded on Tuesday and Thursday than on Monday, Wednesday, and Friday, I will plan half an hour of research for each of the former days and one hour for each of the latter days. Deducting school holidays, this amounts to 60 hours of work in 16 weeks. I will add 10 hours over the one-week spring break, producing a total of 70 hours. I expect to spend 30 hours organizing the data and preparing to run the regression, 20 hours carrying out the regression and significance tests, and 20 hours organizing and reporting the results. At a rate of $15 per hour, I will request a stipend of $1,050. Along with the costs of acquiring data, this amounts to $1,380. No other funding is needed.

References:

Albert, J. (2007, July 12). Hitting in the pinch. In J. Albert & R. H. Koning (Eds.), *Statistical Thinking in Sports* (pp. 111-134). CRC Press.

Birnbaum, P. (2008). Clutch hitting and the Cramer test. *The Baseball Research Journal*, *37*, 71+. Retrieved from <https://link.gale.com/apps/doc/A193140691/AONE?u=loym48904&sid=googleScholar&xid=af21397f>.

Cramer, D. (1977). Do clutch hitters exist?. *The Baseball Research Journal*, 6. Retrieved December 12, 2021, from <https://sabr.org/journal/article/do-clutch-hitters-exist/>.

Dolphin, A. E. (2004, February 2). *Clutch hitting: Fact or fiction?*. Dolphinsim. Retrieved December 12, 2021, from <http://www.dolphinsim.com/ratings/notes/clutch.html>.

Favale, D. (2017, October 3). *The most statistically clutch NBA player at every position*. Bleacher Report. Retrieved December 12, 2021, from <https://bleacherreport.com/articles/2384632-the-most-statistically-clutch-nba-player-at-every-position>.

Fromal, A. (2012, January 26). *Understanding the NBA: Explaining advanced offensive stats and metrics*. Bleacher Report. Retrieved December 12, 2021, from <https://bleacherreport.com/articles/1039116-understanding-the-nba-explaining-advanced-offensive-stats-and-metrics>.

Fuld, E. (© 2005, May 19). *Clutch and choke hitters in Major League Baseball: Romantic myth or empirical fact*. Retrieved December 12, 2021, from <http://cyrilmorong.com/FULD.pdf>.

James, B. (2004). Underestimating the fog. *The Baseball Research Journal*, 33. Retrieved December 12, 2021, from <https://sabr.org/research/article/underestimating-the-fog/>.

Lee, C. (2020, December 31). *Advanced hockey stats 101: Fenwick (part 2 of 4)*. Medium. Retrieved December 12, 2021, from <https://medium.com/hockey-stats/advanced-hockey-stats-101-fenwick-part-2-of-4-a8796b8188d>.

Linsey, B. (2020, May 26). *PFF's all-clutch team: The NFL's best performers in the fourth quarter and overtime in 2019*. PFF. Retrieved December 12, 2021, from <https://www.pff.com/news/nfl-pffs-all-clutch-team-2019>.

Major League Baseball (n.d.). *Late-inning Pressure Situation (LIPS)*. MLB. Retrieved December 12, 2021, from <https://www.mlb.com/glossary/advanced-stats/late-inning-pressure-situations>.

Major League Baseball (n.d.). *On-base Plus Slugging (OPS)*. MLB. Retrieved December 12, 2021, from <https://www.mlb.com/glossary/standard-stats/on-base-plus-slugging>.

Merriam-Webster (n.d.). Clutch. In *Merriam-Webster.com dictionary*. Retrieved December 12, 2021, from <https://www.merriam-webster.com/dictionary/clutch#:~:text=1%20%3A%20to%20hold%20onto%20>.

National Basketball Association (n.d.). *Players clutch traditional | Stats*. NBA. Retrieved December 12, 2021, from <https://www.nba.com/stats/players/clutch-traditional/?sort=GP&dir=-1>.

Palmer, P. (1990). Clutch hitting one more time. *By the Numbers*, *2*(2). Retrieved December 12, 2021, from <http://www.philbirnbaum.com/btn1990-03.pdf>.

Ruane, T. (2005). *In search of clutch hitting*. Retrosheet. Retrieved December 12, 2021, from <https://www.retrosheet.org/Research/RuaneT/clutch_art.htm>.

Ryder, A (2004, January). *Shot quality: A methodology for the study of the quality of a hockey team’s shots allowed*. Retrieved December 12, 2021, from <http://hockeyanalytics.com/Research_files/Shot_Quality.pdf>.

Schatz, A. (2009, September 25). *Fairy tale endings*. ESPN. Retrieved December 12, 2021, from <http://insider.espn.com/nfl/insider/news/story?id=4502503>.

Schumaker, R. P., Solieman, O. K., & Chen, H. (2010, September 6). Research in sports statistics. *Integrated Series in Information Systems*, 26, 29–44. <https://doi.org/10.1007/978-1-4419-6730-5_4>.

Silver, E. (2010, September 5). *Football 101: Introduction to NFL quarterback rating explained with calculator*. Bleacher Report. Retrieved December 12, 2021, from <https://bleacherreport.com/articles/451946-football-101-introduction-to-nfl-quarterback-rating-explained-with-calculator>.

Silver, N. (2007). Is David Ortiz a clutch hitter?. In J. Keri (Ed.), *Baseball Between the Numbers*, Chapter 1-2 (pp. 14-34). Basic Books.

Solomonov, Y., Avugos, S., & Bar-Eli, M. (2015, March). Do clutch players win the game? Testing the validity of the clutch player's reputation in basketball. *Psychology of Sport and Exercise*, *16*(3), 130–138. <https://doi.org/10.1016/j.psychsport.2014.10.004>.

Sporer, E. (2015, February 20). *Primer: Get to know new enhanced stats on NHL.com*. NHL. Retrieved December 12, 2021, from <https://www.nhl.com/news/primer-get-to-know-new-enhanced-stats-on-nhl-com/c-754260>.

Tango, T. M., Lichtman, M. G., & Dolphin, A. E. (2007). *The book: Playing the percentages in baseball*. Potomac Books.

Verducci, T. (2004, April 5). *Does clutch hitting truly exist? Not in the stats community--but don't try telling that to the guys on the field*. Sports Illustrated Vault | SI.com. Retrieved December 12, 2021, from <https://vault.si.com/vault/2004/04/05/does-clutch-hitting-truly-exist-not-in-the-stats-communitybut-dont-try-telling-that-to-the-guys-on-the-field>.