Strokes Gained Analysis of Professional Golfers

Benjamin R. Allen
Department of Sport Analytics, Falk College of Sport and Human Dynamics, Syracuse University

Syracuse University
Falk College of Sport & Human Dynamics

ABSTRACT

As with any sport, it is essential for professional golfers to understand the factors that contribute to their success on the course. As a result, professional golfers are constantly tweaking their games in hopes of maximizing productivity. From equipment changes to swing changes to coaching changes, the ability to thoroughly understanding the historical trends of a golfer in conjunction with contextual information is essential in identifying the true success of a player. In this analysis, I will investigate this research question through a strokes gained performance analysis, while leaving the door open for additional research based on the same methodology.

INTRODUCTION

In golf, the strokes gained metric is a performance analysis tool that applies a numeric value to each shot a golfer takes based the pre- and post-shot lie (fairway, sand, rough) and yardage from the hole. In professional golf tournaments, each pair of sequencing shots is recorded, evaluated, and compared to the rest of the field as the tournament progresses, resulting in the player's strokes gained total (Broadie 2010). Over time, the summation of any given player's total strokes gained as a telescoping sum of positive and negative values will provide a fairly an accurate view of their overall skill. Furthermore, the strokes gained conceptual model also allows us to analyze golfer performance throughout each area of the golf game in subcomponents known as skill categories. The four skill categories that encompass total strokes gained are as follows: Off the Tee (OTT), Approach (APP), Around the Green (ARG), and Putting (PUTT). In answering the research question outlines above, I will be using these subcomponent skill categories as the central driver of my analysis.

The purpose of this analysis is twofold. First, I will investigate the returns to skill of all four skill categories as it relates to the probability that a golfer win a tournament. From there, I investigate the historical trends of the individual player sample to identify and evaluate the 'types' or 'clusters' of players that exist in professional golf. The purpose of this secondary analysis is exponentiated, when considering the results of the initial analysis.

METHOD

Data Methodology

All data used in this analysis was scraped from the datagolf.com historical strokes gained database. The reasoning for a strokes gained based modeling without the inclusion of other stats like fairways, greens, putts, etc. is the vast array of evidence supporting the efficiency of strokes gained models, based on their low AIC values (Courchene 2018). A principal component model that combines strokes gained and traditional stats was also experimented with but ultimately removed. In any case, the following table gives a brief description of the data used in this analysis.

Top 15 PGA Tour Golfers by Average Strokes Gained Per Round

Reminder: 'Total Strokes Gained' is the Sum of Each Individual Skill Category

	Player Name	Total	Putting	Around the Green	Approach	Driving
1	Rahm, Jon	2.17	0.42	0.23	0.72	0.79
2	Fitzpatrick, Matthew	2.05	0.99	0.47	–0.09	0.66
3	Cantlay, Patrick	1.95	0.46	0.28	0.57	0.62
4	Oosthuizen, Louis	1.65	0.75	0.31	0.54	0.05
5	Casey, Paul	1.65	0.01	0.21	1.13	0.30
6	Pendrith, Taylor	1.64	0.50	–0.17	–0.56	1.88
7	DeChambeau, Bryson	1.56	0.45	–0.19	0.21	1.10
8	Migliozzi, Guido	1.53	0.46	0.14	0.75	0.18
9	Spieth, Jordan	1.45	0.46	0.38	0.60	–0.01
10	Ancer, Abraham	1.40	0.46	–0.06	0.64	0.36

**Some International Rounds are Excluded based on SG Tracking Capabilities*

Modeling Methodology

Logistic Regression was used for Part 1 and k-means clustering was used in part 2. For the logistic regression, the four skill categories (putting, around the green, approach, off the tee) were tested in comparison with our binary variable "Winning." Mean Strokes Gained (MSG) from every player from every tournament was used to evaluate the probability significance in skill categories.

In part 2, I chose to chose to cluster the athletes in my sample based on these same skill and investigate the impact of player "type." In this model, individual player data was concatenated to create a set of unique golfers for cluster analysis.

RESULTS Part 1

Returns to Skill Logistic Regression

As mentioned in the method section, the first model is a logistic regression that shed lights onto the impact that each skill category has on a golfer's probability to win a tournament. For our model, **Mean Strokes Gained** of each skill category was used in evaluating play throughout each tournament. This simply means that round-by-round data was compressed into tournament-by-tournament data in order to ensure accurate probabilities. The output of the logit model is as follows:

Strokes Gained Returns to Skill Model

	Dependent variable:
	Won
avg_putt	2.044^{***}
	(0.123)
avg_arg	1.762^{***}
	(0.163)
avg_app	1.954^{***}
	(0.118)
avg_ott	2.532^{***}
	(0.185)
Constant	-8.963^{***}
	(0.323)
Observations	22,864
Log Likelihood	-488.932
Akaike Inf. Crit.	987.864

To help interpret these coefficients, we can transform the log odds from the output into true odds in the table below.

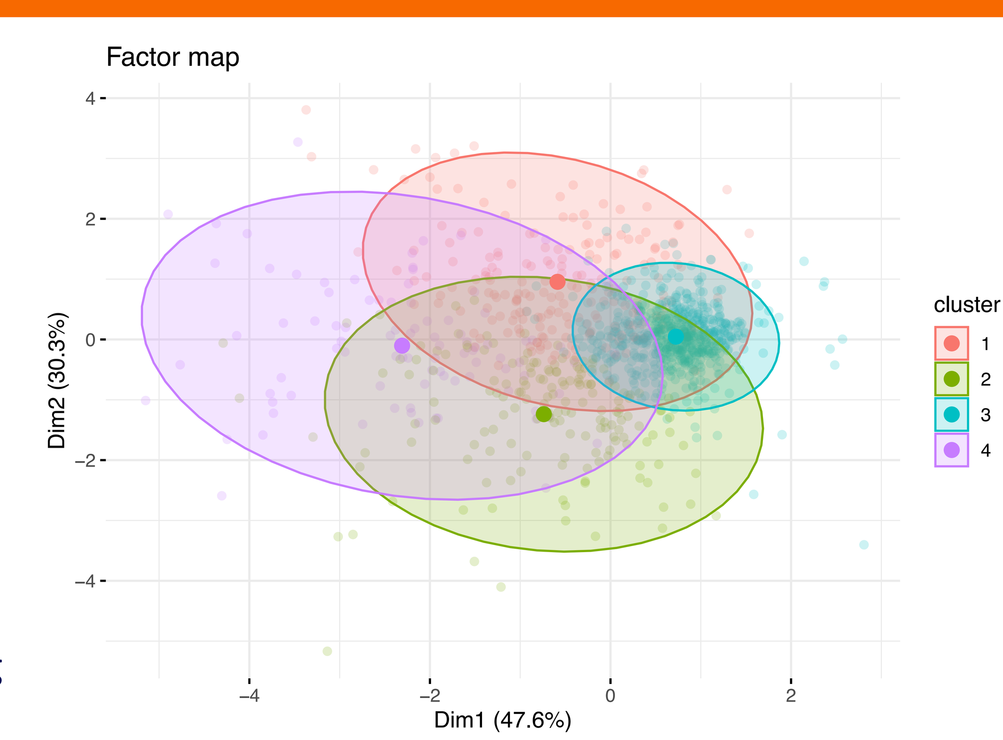
Transformed Log Odds

1	avg_putt	7.724
2	avg_arg	5.823
3	avg_app	7.055
4	avg_ott	12.580
5	Intercept	0.0001

As both tables above depict, the value of an additional mean stroke **off the tee** is significantly higher than any other skill category. One interpretation of these results is that high level players who comparatively gain more strokes from driving are more likely to contend in any given golf tournament. This information is useful in a variety of contexts, but especially when looking at everyday changes a golfer makes to his game. In prepping for any given tournament, season, etc., a player may emphasize putting or short game at the potential expense of his ball striking. Because the strokes gained tradeoff supports driving as the more significant factor, the results of this model advise against that sort of thinking.

RESULTS Part 2

The next model is a k-means clustering algorithm, where each input point represents a different golfer with varying skillset in terms of strokes gained. Because we determined certain skill categories increase the probability of winning more than others, a logical next step would be to cluster and evaluate golfers that possess similar proficiencies. Upon running the model, we get a plot that depicts our four clusters along with the associated central point.



Each colored section represents a different type of golfer based on their most prominent strokes gained skill category. Players like Bryson DeChambeau are in cluster 4 based on their prodigious strength off the tee, while players like Dean Burmester are in cluster 1, based on their putting prowess.

CONCLUSIONS

The overarching conclusion to this analysis is that the relationship between a player's "type" and the ability for them to contend in a golf tournament is closely intertwined. Based on the kind of player you are, analyst and coaching can more easily predict the success you'll have on the course.

The long-term value of this project comes not only in the results of my model, but the reader's ability to replicate the modeling and visualization techniques on any golfer in the time frame. It is my hope to expand this research in order to have a more concreate understanding of strokes gained performance as a hole.

REFERENCES

1. M. Broadie (2010) *Assessing Golfer Performance on the PGA Tour* Columbia University

2. Courchene, W., & Courchene, M. (2018). *Predictive Model Methodology.* Data Golf.

ACKNOWLEDGEMENTS

Special thanks to Dr. Rodney Paul for guiding me through the entire research process!