Analyzing Disc Golf: 2020 Idlewild Open

Authors: Khala Dally, Ethan DellaMaestra, Sneha Kottakkudy, Patrick Loekman, Hannah Shimshock

Introduction

This paper will attempt to analyze the sport of disc golf. For those unfamiliar with this sport, players use flying discs and attempt to throw these discs into a basket or target. The rules are very similar to regular golf where the number of shots it takes for each player to make it into the basket is recorded and the lower the score, the better it is. Specifically, this paper will be looking at the 2020 Idlewild tournament played from August 7th to August 9th in Burlington, Kentucky. This tournament is one of the most difficult tournaments on the calendar. Our research team gathered data from the JomezPro Youtube coverage of the tournament across the 3 rounds. We hoped to gather data that helped determine different factors that could potentially influence a player's score for the tournament. This kind of information could be extremely useful for players attempting to strengthen their game or companies looking at which potential athletes they might want to sponsor. We recorded observations at a player level looking at how each player on the Youtube coverage performed on a particular hole for each of the 3 rounds. The variables that we observed were the **Par** of the hole, the **Distance in Feet** of the hole, the player's **Score** on the hole on a scale from -2 for eagle to 3 for triple bogey, the number of **Out of Bounds** shots the player had on a hole, the number of **Obstacles** the player hit on each hole (mostly trees/branches), and the **Type of Shot** each player threw off of the tee. Each member of the research team watched either 9 or 18 holes of the tournament coverage on JomezPro and recorded their observations. There were a total of 216 observations across the 3 rounds of the 18-hole tournament. The figure below gives a preview as to what our dataset looks like. The **2020 Idlewild Tournament Data** table below lists the **Round**, **Hole**, and **Player** of each observation and the recorded data of our remaining variables.

2020 Idlewild Tournament Data

Round	Hole	Player	Par	Distance ft.	Score	OOB	Obstacles	Shot Type
1	1	Nikko Locastro	4	638	0	0	0	Backhand
1	1	James Conrad	4	638	0	0	0	Backhand
1	1	Paul McBeth	4	638	-1	0	0	Backhand
1	1	Zach Melton	4	638	-1	0	0	Backhand
1	2	Nikko Locastro	4	601	-1	0	0	Backhand
1	2	James Conrad	4	601	0	0	2	Backhand
1	2	Paul McBeth	4	601	-1	0	0	Forehand
1	2	Zach Melton	4	601	-1	0	0	Backhand
1	3	Nikko Locastro	3	407	1	1	0	Backhand
1	3	James Conrad	3	407	-1	0	0	Backhand
1	3	Paul McBeth	3	407	-1	0	0	Backhand
1	3	Zach Melton	3	407	0	0	0	Backhand

Summary

In this section, we will give readers a summary of what our collected data turned out to be. First, we will discuss the results of the Score variable grouped by the 18 different holes. The Score variable was recorded on a scale from -2 for Eagle to 3 for Triple Bogey. The **Score Distributions by Hole** Table below shows the frequencies and relative frequencies of each score. The last column of the table also shows the Average Score of each hole to show which of the holes were the easiest and the hardest for the players on the JomezPro coverage. The average scores ranged from -1.0 (1 stroke under par) at the easiest on hole 16 and 0.5(half stroke over par) at the hardest on hole 18. The highest frequency score for every hole was either Birdie or Par which makes sense because of the highly talented field of players in this well-respected

Score Distributions by Hole

	Eagle	Birdie	Par	Bogey	Double Bogey	Triple Bogey	Eagle %	Birdie %	Par %	Bogey %	Double Bogey %	Triple Bogey %	Average Score
Hole 1	0	8	3	1	0	0	0.00	66.67	25.00	8.33	0.00	0.00	-0.58
Hole 2	0	4	7	0	0	1	0.00	33.33	58.33	0.00	0.00	8.33	-0.08
Hole 3	0	7	4	1	0	0	0.00	58.33	33.33	8.33	0.00	0.00	-0.50
Hole 4	0	4	7	1	0	0	0.00	33.33	58.33	8.33	0.00	0.00	-0.25
Hole 5	0	4	7	0	1	0	0.00	33.33	58.33	0.00	8.33	0.00	-0.17
Hole 6	0	4	6	2	0	0	0.00	33.33	50.00	16.67	0.00	0.00	-0.17
Hole 7	0	6	2	3	1	0	0.00	50.00	16.67	25.00	8.33	0.00	-0.08
Hole 8	2	6	4	0	0	0	16.67	50.00	33.33	0.00	0.00	0.00	-0.83
Hole 9	0	7	4	1	0	0	0.00	58.33	33.33	8.33	0.00	0.00	-0.50
Hole 10	0	1	8	3	0	0	0.00	8.33	66.67	25.00	0.00	0.00	0.17
Hole 11	0	7	4	1	0	0	0.00	58.33	33.33	8.33	0.00	0.00	-0.50
Hole 12	0	3	7	2	0	0	0.00	25.00	58.33	16.67	0.00	0.00	-0.08
Hole 13	1	6	4	1	0	0	8.33	50.00	33.33	8.33	0.00	0.00	-0.58
Hole 14	0	5	4	3	0	0	0.00	41.67	33.33	25.00	0.00	0.00	-0.17
Hole 15	0	4	6	2	0	0	0.00	33.33	50.00	16.67	0.00	0.00	-0.17
Hole 16	3	6	3	0	0	0	25.00	50.00	25.00	0.00	0.00	0.00	-1.00
Hole 17	0	5	3	3	1	0	0.00	41.67	25.00	25.00	8.33	0.00	0.00
Hole 18	0	1	6	3	2	0	0.00	8.33	50.00	25.00	16.67	0.00	0.50

tournament.

Another variable we recorded was the type of shot a player threw off of the tee. There are two options for this variable: backhand or forehand. To summarize this variable, we grouped the shot types by player and recorded the counts and relative frequencies. This information can be seen in the **Shot Types by Player** table below. There were 3 players (Adam Hammes, Eagle McMahon, Jeremy Koling) that threw a majority of forehand shots. The remaining 6 players threw predominantly backhand shots with 2 players (James Conrad, Patrick Brown) not throwing a single forehand shot off of the tee.

	Shot Types by Player			
	Backhand Shots	Forehand Shots	Percent Backhand	Percent Forehand
Adam Hammes	7	11	38.89	61.11
Eagle McMahon	8	10	44.44	55.56
James Conrad	18	0	100.00	0.00
Jeremy Koling	11	25	30.56	69.44
Jordan Castro	13	5	72.22	27.78
Nikko Locastro	14	4	77.78	22.22
Patrick Brown	18	0	100.00	0.00
Paul McBeth	44	10	81.48	18.52
Zach Melton	14	4	77.78	22.22

We also analyzed several numeric variables, specifically, Par, Distance, OOB (Out of Bounds), and Obstacles. For each of the variables mentioned, the minimum, 1st quartile, median, mean, 3rd quartile, max, and standard deviation were calculated. Most of the holes played had a Par of 4 with very little variation. The minimum Par was 3 while the maximum was 5. On the other hand, the distance of the courses varied greatly, with a standard deviation of almost 161 feet and values ranging from around 200 feet to almost 1000 feet. The summary statistics also show that few discs went out of bounds and few obstacles were hit on average, which makes sense due to the skill level of players observed and the prestige of the tournament.

	Variable Summaries						
	Min	1st Qu.	Median	Mean	3rd Qu.	Max	SD
Par	3	3	4.0	3.7780	4	5	0.6299994
Distance	242	418	485.5	510.8000	601	969	161.9672854
OOB	0	0	0.0	0.1204	0	2	0.3789236
Obstacles	0	0	0.0	0.5833	1	3	0.7850344

After analyzing our observations, the overall difficulty for each hole in the 2020 Idlewild Open can be summarized using the figures below. Figure 1 provides a visual overview of different performances in the eighteen holes. Each bar represents a hole, and the colored distribution presents the player's scores. In the context of disc golf, the more negative scores are better while the more positive ones are worse (since they show a greater number of attempts). Each hole has its own par value so the absolute number of attempts need to be converted relative to the specific par value creating one of the scoring options below (-2, -1, 0, 1, 2, and 3). For example, a player who scores 2 on a hole with a par value of 3 will be awarded -1 in their score report. In disc golf lingo, a -2 on a hole is called an "Eagle", -1 is a birdie, 0 is a par, 1 is a bogey, 2 is a double bogey, and so forth. On this professional stage of disc golf, scores are not worse than a triple bogey and are not better than an eagle. Figure 2 shows the number of errors produced in each of the eighteen holes (each bar represents a hole). Our team defines an error as either a throw that resulted to be out of bounds or a throw that hit an obstacle such as a tree to severely deviate its intended flight path. Analyzing errors may also help determine which holes are more challenging as some may naturally have more trees, unbalanced terrain, or a curved path.

Figure 1

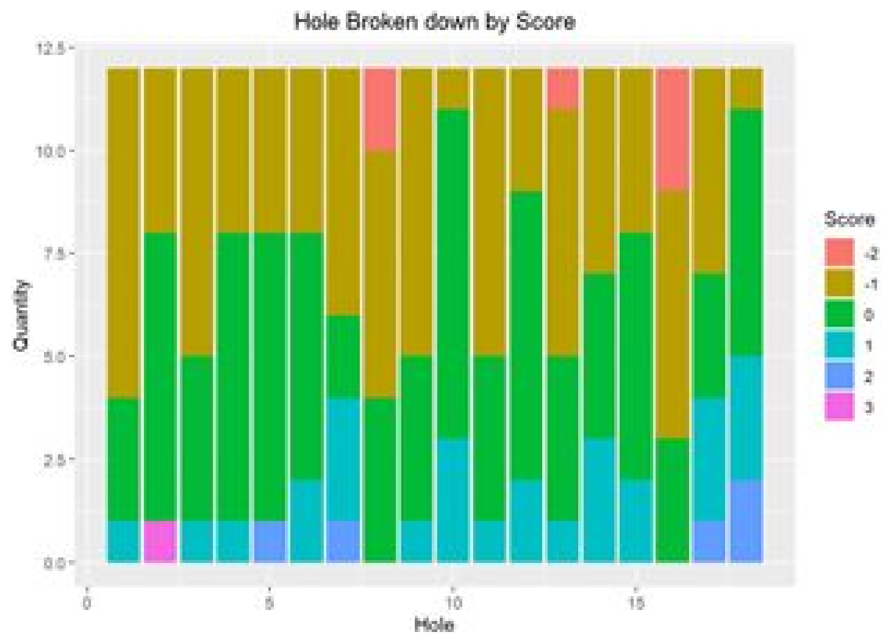


Figure 2

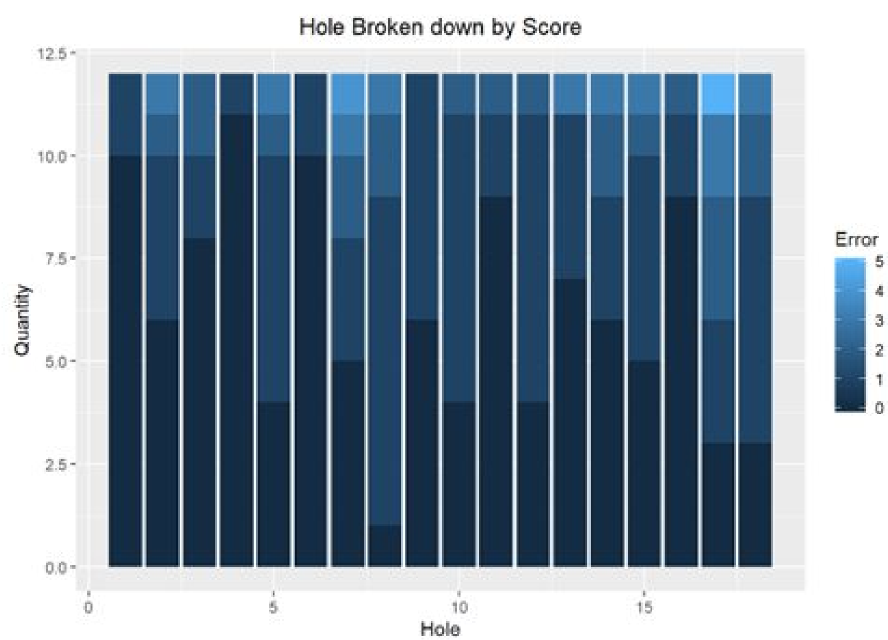
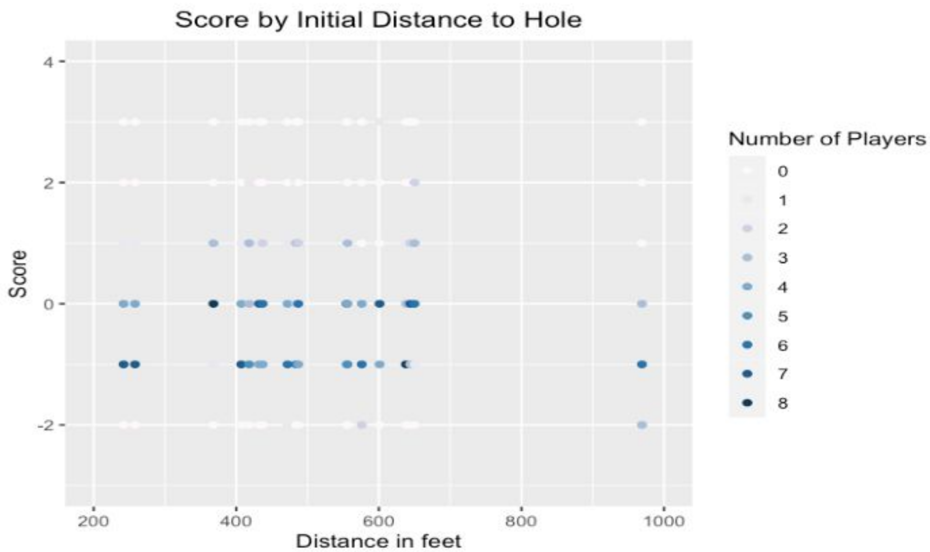


Figure 3



In Figure 3 above we can see a scatterplot comparing the score to the initial distance to the hole. With both Score and Distance in feet being discrete variables, we can see the number of players who got a certain score at a certain distance by looking at the color of the point.

For our analysis, we first split up the data by randomly assigning 70% of the observations to a train dataset that would be used to create our model and the other 30% was assigned to a test dataset that would be used to test the predictive accuracy of our model. We then constructed a multivariate logistic model using stepwise regression to predict score based on the most significant variables which we found to be Par, Distance, Out of Bounds, and Obstacles. The coefficients for each of the significant variables in our model, and their respective p-values indicating their significance level are shown in the Coefficients Table below. We will discuss some of the insights we gained by looking at this model in the next section of our paper.

	Coefficients Table				
	Intercept	Par	Distance	OOB	Obstacles
Coefficients	0.7289132	-0.6303789	0.0018583	0.8839548	0.546621
Significance	0.0208650	0.0000017	0.0002840	0.0000000	0.000000

After creating our regression model, we tested it on the remaining observations in the test data to see how well our model was at predicting score. We then analyzed the mean, standard deviation, and the coefficient of variation of the residuals to determine how far off our model's predictions were from the actual data. This information is summarized in the Residuals Table below. We found that the residuals have a mean close to zero, a relatively small standard deviation, and a CV less than one. This tells us that our logistic model found through stepwise regression does a relatively good job predicting Score.

Residuals Table	
	Model Stats
Residual Mean	0.1228811
Residual SD	0.6810077
CV	0.1804402

Insights

Figure 1 illustrates that all 18 holes have an overwhelmingly large proportion of either birdies or pars. As this is a professional disc golf tournament, it makes sense to observe that the players score a majority of birdies and pars for each hole. If data was collected from recreational players, it would be expected for perhaps a majority of scores to be par and bogey, or bogey and double bogey. Three holes have scores that are better than birdie; they are holes 8, 13, and 16. These excellent scores may reflect the low difficulty of the holes, but perhaps some players had “lucky” shots and therefore performed as outliers. However, this is not likely to be the case because holes 8 and 16 also follow a high scoring distribution; no player scored worse than par. For hole 13, one person scored worse than par; he scored a bogey. Interestingly, hole 8 had a majority of between 2 and 3 errors despite having a high score. One possible explanation may be the sudden transition from open terrain to the forest. Additionally, hole 8 is a par 5 showing that a higher par might not mean a higher score. For hole 16, the distribution shows a majority of no errors. Because hole 16 has the best average score and low error rate, it may be described as the easiest hole. It is noteworthy that hole 16 is also a par 5 hole with a distance of 969 ft; this is yet another example showing that a higher par does not necessarily imply a harder hole. With this analysis, players should try to take advantage of holes 8 and 16 to boost their scores. In contrast, one may observe that holes 7 and 18 are the most challenging. Both follow a similar distribution with four or more players scoring worse than par. Hole 18 has the worst average score by seeing that only one person scored a birdie. Furthermore, holes 7 and 18 also have a high error rate with a majority of more than 2 errors. Players should play more conservatively on holes 7 and 18 as they are likely to lower scores.

Initial thinking led us to believe that distance would have a large influence on a player's score. That being said, the farther the hole, the harder it would be to get a low score. However, looking at Figure 3, we can see that the score tends to stay within the -1 to 1 range regardless of distance. If we look further we can see in our regression analysis that distance has a p-value of .0028 and a coefficient of .00186 in our model for predicting score. Since the p-value is less than .05 we can conclude that it has a significant impact on predicting score, however, it might not be as effective as some other variables such as OOB and Obstacles hit. This could be because a hole with a closer distance may have more obstacles or have a more difficult path.

Our final insight comes from the regression model we created to answer our initial question about which factors influence a player's score in the tournament. As we can see from the Coefficients Table, our model found the most significant variables to predict a player's score in this tournament to be Par, Distance, OOB, and Obstacles. Looking back at the coefficients for each of the prediction variables gave us greater insight into the game itself because we were then able to logically explain the significance behind some of these coefficients based on our experience gathering the observations. For instance, we noticed that par has a negative coefficient in our model, indicating that higher par values lead to better scores. We found this interesting because our initial thinking was that higher par values indicated a harder hole, which we thought would worsen a player's score. However, after further consideration, our team realized that higher pars typically mean that players are better able to correct earlier mistakes made on the hole. While gathering our observations, we noticed that low par holes meant that OOB and Obstacle errors had a much higher impact on a player's score than it did on higher par holes because players had a better chance of fixing their mistakes. We also thought the coefficient for OOB was interesting because it is 0.88 which we translated to mean that for every OOB error, the score was negatively impacted by almost 0.88 points. We thought this was interesting because, by definition, going out of bounds assesses the player 1 penalty stroke so we would expect the coefficient to be 1. However, it is slightly below 1. This could be because, oftentimes, when a person throws their disc out of bounds, they are attempting to get aggressive. As we discussed earlier, on longer holes, players can correct their mistakes on early shots so this could be why the OOB coefficient is not quite a full stroke.

Our analysis of disc golf is not without its faults and many things could have been done differently to get a better quality analysis. One thing we could have done was to collect data from more than one tournament. It would have been beneficial to collect data from the 2019 Idlewild Open, 2018 Idlewild Open, and so on. Many of the same players would be playing in those tournaments and it would give us more possibilities within the same pool of talent to base our analysis off of. We also could have looked into other disc golf tournaments on the calendar to compare our data across different courses and times of the year. Another potential improvement for our analysis is to analyze the possible confounding effects of the hole and round number in our analysis. Since the Idlewild Open is consecutive, only the top select players from each round make it onto the coverage for the following round. Furthermore, as holes progress, it seems as though it gets harder with the 18th hole often being the most difficult which we noticed in our insights. In respect to our actual analysis and the model we found to be a good indicator of predicting score, we could have done further verifications to test whether the regression model we came up with was truly a good regression model. We looked at the residuals but other methods could have been used to ensure the quality of the model. There are also potentially other variables that we did not record that could have an impact on a player's score such as fairway hits or putting percentages. For future analysis, we could record other variables to see if they are significant as well.