A Study to Investigate the Factors Influencing the Distance of the Golf Ball

Sukyoung Cho William R. Hagen Meenmo Kang
Stat 424 Final Project

Update: April 26, 2019

1 Problem Statement

Golf is a game that is highly variable and constantly played in different ways due to its variability. In essence, the objective like any game is always the same, hit a small white ball into a hole far away by using different sizes and shapes of stick like objects called clubs in as few hits as possible. It is a very simple concept that is complicated by a plethora of factors. For example, weather conditions, types of grass, and slope and elevation changes are just a few of the factors that can be different from round to round that will affect how the game is played. One of the most important parts about playing golf is hitting the ball far and anticipating the distance it will go based on all the factors at hand. There are a number of factors, like the ones mentioned earlier, that are out of the player's control. One cannot decide how fast the wind is moving or how humid the air is on a given day, but the player can still adjust to these factors and play accordingly. On the other hand, there are a number of factors always under the player's control that can affect how far a golf ball will go. These various factors can be tested in a controlled environment in which the other uncontrollable factors are absent. This can give the player an idea of how far the golf ball will go under only factors up to their discretion. The player can see which factors interact with one another to make the ball go the furthest. The three factors we choose to study that are always under a player's control are the type of ball, the type of club shaft, and using a tee or not. We are curious as to which of these factors or combination of factors will interact with one another to significantly affect the distance of a golf ball struck by a seven iron.

2 Choice of the Factors and Response

Our variables of interest in the experiment are distance, the response, ball, shaft, and tee, which are the factors. Distance was chosen as the response variable because it is simple to measure and understand. It is also one of the most important things to know and think about for the player. The factors were chosen as the ball, shaft, and tee because they are all important factors that a golfer has to make a decision about and because they all affect the distance of a golf ball. There were two levels chosen for each of the three factors. The levels were therefore chosen as best as possible to represent two opposite ends of the gradient for each factor. For the type of ball and type of shaft there is a gradient of levels that could be chosen. For example, there are some golf balls that would be considered "okay", neither very good nor very poor. The same can be said for the type of shaft in which a mix of materials can be used to construct the shaft. These types of levels were not chosen so that the disparity in our two levels is as large as possible in order to better highlight any differences in distance that may exist among a factor. The two levels for the type of ball were taken as a "high performance" golf ball and a "poorly made" golf ball. These two specific balls were the Titleist Pro V1 and the Titleist DT Solo, respectively. The two levels for the type of shaft were graphite and steel. The two levels for the tee were simply using a tee or not using a tee.

3 Execution of the Experiment

We conducted our experiment indoors using a golf simulator, as we wanted perfectly controlled environment. There are many realistic problems in booking a golf course, including external factors which are none of our interest. For example, nuisance factors, such as a wind and a condition of grass, are significant factors but out of human control at the same time. The simulator shows the path the golf ball takes and how far it travels. There are three factors in our experiment: two different irons, two different balls, and shots with or without tee. As for iron and ball, one is superior in terms of quality, price, and performance, while the other one is relatively inferior. We hit shots for each possible combination of the different levels among the three factors, so there were eight total possible combinations. For each combination three swings or replicates were conducted to account for the variability. There are a few things done in order to minimize errors as much as possible. First, the same person conducted swings for the experiment. Different people would result in different distances, and this would have confounded our factors of interest. We also employed the best golfer in our group in order to make swings as precise as possible. For beginners, there would be a huge variation in shots - one good shot and very poor or shanked another. This would result in two very different distances which were not caused from the difference in the treatment. Since it was impossible to have perfectly error-free results, when one shot was totally out of the average due to fatigue, the replicate was repeated to minimize the error. Those human errors in the swing were not what we were interested in. Therefore, we only used replicates that were hit precisely and accurately comparted to the other shots in terms of the distance. The distance was measured in yards.

4 Choice of the Design

Since the factorial designs are known for the most efficient statistical method dealing with two or more factors, and frequently used in purpose of understanding the effects of multiple independent variables upon a single dependent variable, we chose this to be a main method of analysis. 2^k factorial design would allow us to identify useful factors and their interactions, and to exclude unimportant treatments. As we described above, we adopted two levels for each factor: an iron with graphite shaft and steel shaft, a ball of an average performance and higher performance, and putting on tee or not. Low level was coded to be 1 and coded 1 for the upper level treatment. Hence, 2^3 factorial design was conducted with 3 replicates in order to account for the variability in the response. We expected that every single combination of the levels of the factors to be investigated throughout the factorial design, including the interaction up to 3-way interactions.

5 Model Analysis

$$x_1 = \begin{cases} -1 & \text{if} \quad \text{A} = \text{-} \\ 1 & \text{if} \quad \text{A} = + \end{cases} \qquad x_2 = \begin{cases} -1 & \text{if} \quad \text{B} = \text{-} \\ 1 & \text{if} \quad \text{B} = + \end{cases} \qquad x_3 = \begin{cases} -1 & \text{if} \quad \text{C} = \text{-} \\ 1 & \text{if} \quad \text{C} = + \end{cases}$$

$$y = \mu + \frac{A}{2}x_1 + \frac{B}{2}x_2 + \frac{C}{2}x_3 + \frac{AB}{2}x_1x_2 + \frac{AC}{2}x_1x_3 + \frac{BC}{2}x_2x_3 + \frac{ABC}{2}x_1x_2x_3 + \epsilon$$

The equation above is the regression model for our experiment, which is easier to understand. According to the summary table for ANOVA and regrssion, low p-values from the F-test demonstrate that tee, an interaction between ball and tee, and the 3-way interaction of all treatments seem to be only significant factors with 95 percent confidence 0.303, 0.0109, and 0.0125 respectively. Each F_0 tests for the corresponding effect being zero, i.e.

$$H_0 : \text{effect } = 0 \text{ vs } H_a : \text{effect } \neq 0$$

In addition, on the interaction plots below, we can observe intercepts between ball and tee factor, which manifests the interaction effect between them. However, looking at a normal probability plot, an iron seem to be outstanding, contrasting to its p-value on the summary table. It could be because an iron has a collinear effect with other factors. According to the adjusted R-squared value, our model only accounts for 44.73 percent of the variance. The model holds four assumptions behind the factorial design: interval data, normality, homoskedasticity, Multicollinearity assumptions. Looking at the estimated effects table and the dataset, the combination of a high level of b (+), low level for a (-) and c (-) resulted in the farthest distance the ball traveled 186.27 yards in average of three replicates. Our final reduced model will be with a ball, tee and a 3-way interaction among all three factors - lm(dist ball * tee + iron:ball:tee). A ball was included to maintain the hierarchy.

5.1 Strengths

- Simple: It is easy to understand and figure out the effects

- Replicates: it allows us to measure the variation to evaluate the differences, which increases the precision.
- Fixed factors: knowing background and able to make a precise expectation.
- Normality: Shapiro-Wilk's test and QQ plot proves that it holds the normality assumptions
- Homoskedasticity: On the standardized residual plot, the variances across the x-axis are constant, which means it holds the homoskedasticity.
- Multicollinearity: All factors are independent of each other.
- Interval data assumption: Our dependent variable, a distance, is a numeric data not nominal or ordinal.
- Treatment effect: factors would have the same effect no matter who an individual is.

5.2 Weaknesses

- Response: our experiment only took the distance into consideration, which is one minor features in playing golf. There are many other important things such as spin, accuracy, and others.
- Factor level: There was a nuisance factor which we did not expect. Iron, for example, had a different shaft but not others constant. For example, two irons also had different club heads.
- Human error: swings would not be precise and accurate. Although we took rest in the meantime, it is unpredictable.
- Accuracy: We used a golf simulator to collect the data, so it may be different from actually playing outside.

6 Conclusion

The experiment investigated how those three treatments affect the distance of a golf ball flight. There are a few limitations in our model that it only takes the distance into consideration while it is one of the minor aspects in playing golf. Also, we failed to control all the errors such as human errors and factor level. It could be different from playing outside with wind and other external factors. In addition, according to the adjusted R-squared value, our model only explains for 44.73 percent of variance. However, the experiment successfully discovered not only which factor is significant but also all possible interactions between the treatments. Moreover, the factors are fixed, which makes the result more reliable and applicable. In overall, it suggests a simple rule of thumb to hit the ball farther as well. Put the higher performance golf ball on the ground and swing with an iron of graphite shaft. It will send the ball the farthest you can.

7 Tables

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
iron	1	0.13	0.13	0.00	0.9603
ball	1	71.41	71.41	1.35	0.2620
tee	1	298.22	298.22	5.64	0.0303
iron:ball	1	16.34	16.34	0.31	0.5859
iron:tee	1	110.08	110.08	2.08	0.1682
ball:tee	1	438.61	438.61	8.30	0.0109
iron:ball:tee	1	418.33	418.33	7.92	0.0125
Residuals	16	845.33	52.83		

Table 1: ANOVA Output Table

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	167.5667	4.1965	39.93	0.0000
iron1	11.1333	5.9348	1.88	0.0790
ball1	18.7000	5.9348	3.15	0.0062
tee1	14.1333	5.9348	2.38	0.0300
iron1:ball1	-13.4000	8.3931	-1.60	0.1299
iron1:tee1	-25.2667	8.3931	-3.01	0.0083
ball1:tee1	-33.8000	8.3931	-4.03	0.0010
iron1:ball1:tee1	33.4000	11.8696	2.81	0.0125

Table 2: Linear Model Output Table

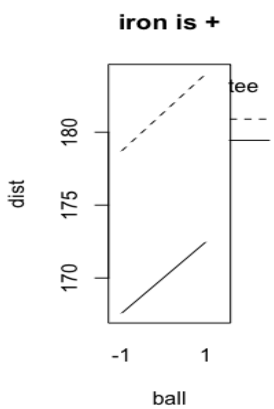
	2.5 %	97.5 %
(Intercept)	172.46	178.75
iron	-3.07	3.22
ball	-1.42	4.87
tee	-6.67	-0.38
iron:ball	-2.32	3.97
iron:tee	-5.29	1.00
ball:tee	-7.42	-1.13
iron:ball:tee	1.03	7.32

Table 3: Confidence Interval for Regression

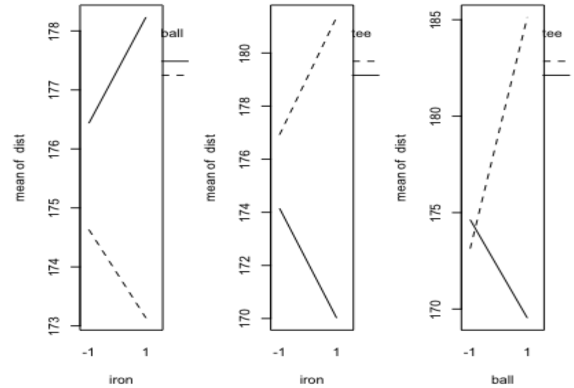
iron	**ball**	**tee**	**iron:ball**	**iron:tee**	**ball:tee**	**iron:ball:tee**
0.15	3.45	-7.05	1.65	-4.28333	-8.55	8.35

Table 4: Estimated Effects

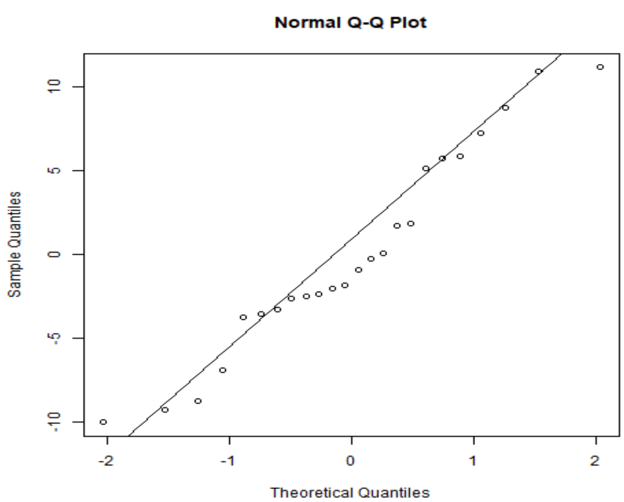
8 Graphics



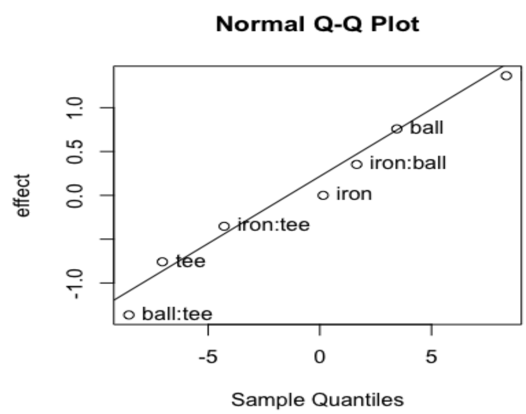
iron is +
iron is -
tee
dist
ball



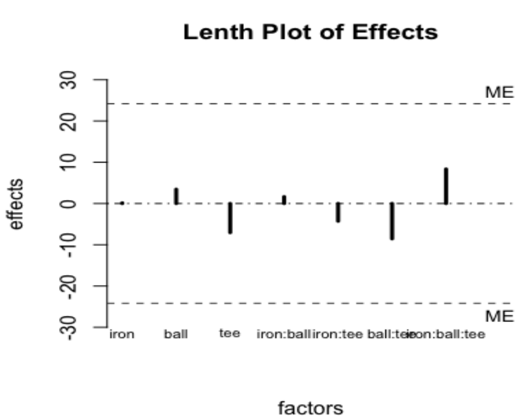
mean of dist
iron
ball
tee



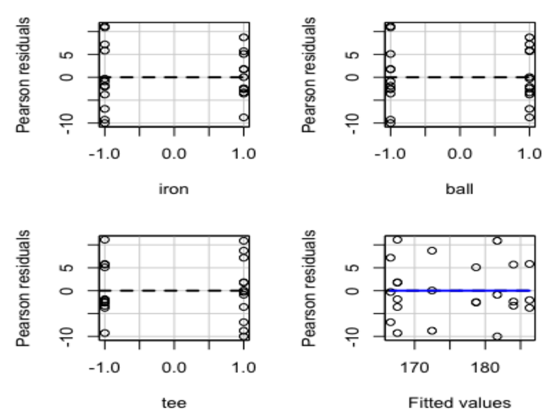
Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles



Normal Q-Q Plot
effect
ball
iron:ball
iron
iron:tee
tee
ball:tee
Sample Quantiles



Lenth Plot of Effects
ME
ME
effects
iron
ball
tee
iron:ball
iron:tee
ball:tee
iron:ball:tee
factors



Pearson residuals
iron
ball
tee
Fitted values