Golf Project
Stat 795

Aditya, Josh, Kelso

December 10, 2018

Executive Summary

The golf ball manufacturer, Titlest, have claims against smaller companies that some of their brand new balls are not performing to standard. In addition, there is a lawsuit saying that the companies are taking nearly new balls or slightly used balls and either washing or refurbish them, then selling them off as new golf balls. The model of ball in question is ProV1392.

Introduction

In golf, players want to minimize the number of strokes/hits. The best players will have scores or strokes below par. Par is the standard number of strokes needed to make the ball into the hole. In tournaments, players will have negative scores which means that players have repeatedly par. That's a good thing. In fact, the least amount of stroke made during a game or tournament wins. Tournaments will consistently see negative scores, but a tampered or old ball will affect a player's game and their number of strokes needed to make the ball. The golfing associations which govern the size of golf balls, R&A and United States Golf Association, specify that the mass of a ball must be no more than 1.620 oz and have a diameter at least 1.680 inches. Any golf balls that do not meet those regulations cannot be used in tournaments. The company Titleist are suing other smaller companies because those smaller companies are washing and refurbishing balls and selling them as new. There are complaints that the "new" balls are performing below standard. This analysis is to find out if those smaller company is truly washing and refurbishing their balls (ProV1392 model) and selling them as new.

Methods

Hit number is just the observation number. Block is which group the ball is in. Total is the total distance the ball has moved. Carry is the distance from hitting the ball and where it initially lands. Ball model is whether the ball is a new ball, washed ball, or a refurbished ball. C_ground is the how far the ball rolled after the initial landing of the ball. Traj is the angle which the ball's path has taken. Total dispersion is the distance away from where the ball would have gone if its trajectory was straight or centerline. Ballnum is the id of the ball. There were no observations excluded nor included. This analysis is going to use R/RStudio software.

Results

Experimental Design

In order to test to see if there is a difference in performance between brand new golf balls and the washed or refurbished golf balls that are being marketed as new, a well designed experiment must be created. A key element in any well designed experiment is randomization. Care must be taken to ensure that all locations are considered where both the new and refurbished balls are being sold and then a certain amount of both new and refurbished balls be purchased randomly from these locations. The desired sample size is computed from a sample size estimation where 95% power is achieved for an ANOVA model testing to see if there is a difference between the groups. The sample size chosen for this experiment is 84 observations per group for a total of 252 experimental units. Once the sample has been collected, golf balls will be placed into three separate categories. New golf balls are balls that are actually new and never been used before. Washed golf balls are balls that are used, but simply washing them could get rid of any markings on them. Refurbished balls needed more than a simple washing, and needed to be refinished in order for them to give the appearance of being a new golf ball. Each ball is then placed in a small container and given an unique ball number so that they can be randomly selected during testing.

Once the sample has been collected a series of tests can be conducted to see if there truly is a difference in performance between the new balls and the refurbished balls. We will conduct one test at a time because all that is needed to show that there is, in fact, a difference between the new balls and the refurbished ones is that they perform different in regards to a single metric. The easiest metric to measure is distance, so we will begin by measuring the distance off the tee for each of the groups of golf balls. In order to reduce the amount of error in the model, a machine will swing the golf club to ensure that the golf balls are being struck with a similar amount of force. Multiple measurements will be taken to assess the distance traveled for each ball. Carry distance will measure how far the ball travels in the air and total distance will measure how far the ball travels after it is done rolling. As wind conditions can be a confounding factor in the experiment, a randomized complete block design will be used so that the wind conditions will be similar within each block. Two balls of each kind will be included in each block, and the machine will be recalibrated between blocks to ensure that the force used to hit each ball is consistent. The average golfer swings a golf club at about 93 miles per hour, so this was the swing speed chosen when calibrating the machine.

In addition to the force being used to hit the ball, the trajectory of the ball will have an impact on how far each ball will travel. Due to a variety of different reasons, different golfers hit the ball at different angles. Either they have drivers with different amounts of loft, they place the ball more forward or back in their stance or they fail to make contact with the ball at the optimal point in their swing. For these reasons the trajectory of the ball after impact was altered randomly for each shot and then measured. The trajectory for each shot was randomly chosen from a normal distribution with mean trajectory 4.5 and a standard deviation of 1.5. Since trajectory will have a large effect on the distance traveled, it will need to be included as a covariate in our model.

Exploratory Data Analysis
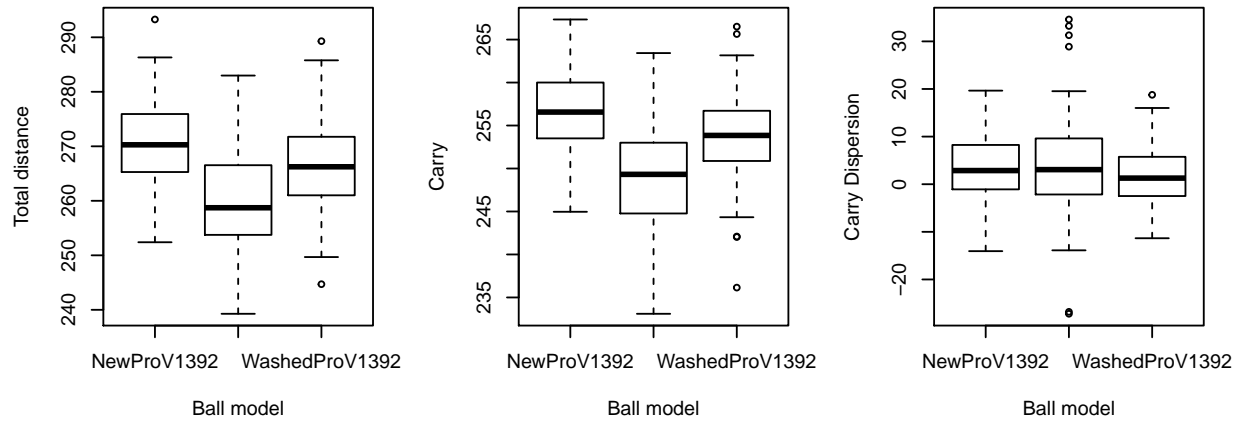
From the box plots, we can see that the total distance is considerably high for new balls compared to washed and refurbished balls. Washed balls travel higher total distance compared to the refurbished balls. The carry is almost same for new and refurbished balls whereas it is considerably less for washed balls as compared to new and refurbished balls. Carry dispersion is almost same for new and refurbished balls whereas it is considerably less for washed balls as compared to new and refurbished balls.
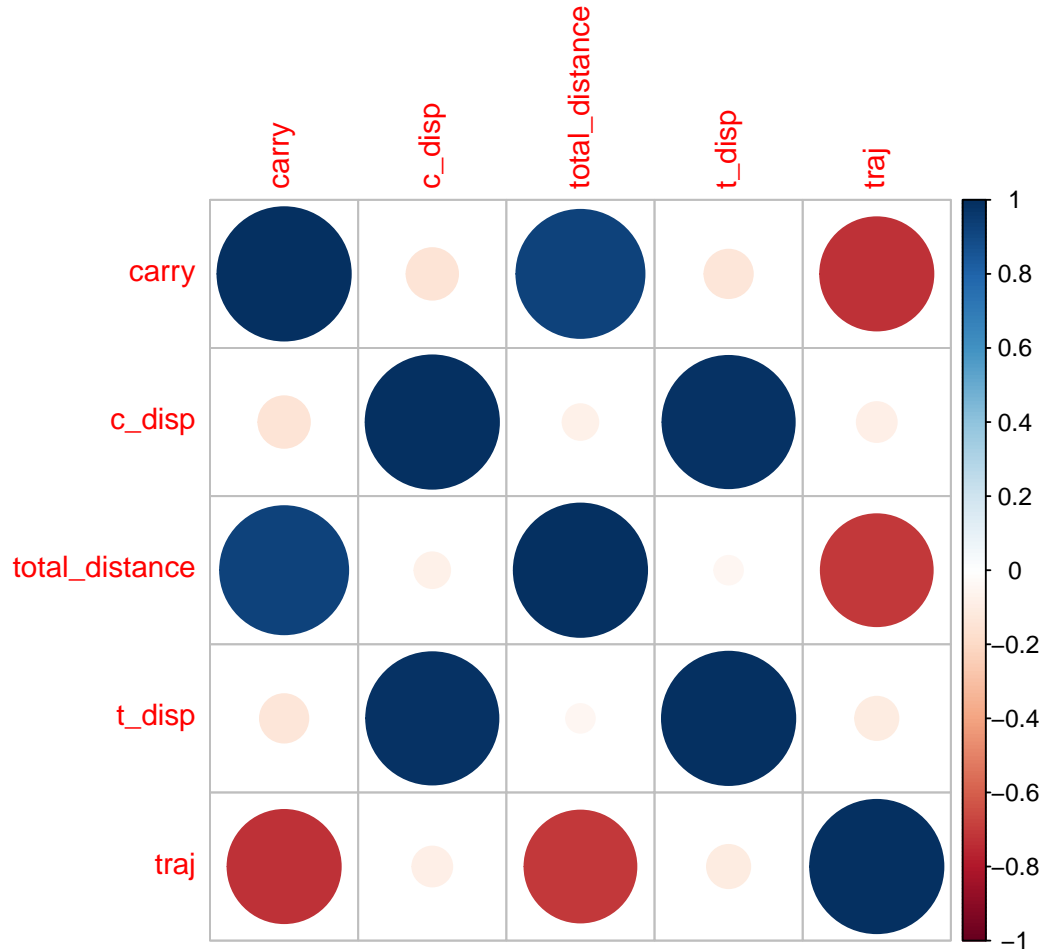
Total distance has a high positive correlation with carry but a large negative correlation with trajectory. Total dispersion has a high positive correlation with carry dispersion. Trajectory has a high negative correlation with carry and total distance which seems logical.

More to be added in EDA

Model Selection/Inferences

We are concerned about the total distance travelled by the "new" balls. Since a controlled experiment was conducted, we have controlled blocks and treatement. Ball model is the treatment effect and block is the block effect. Trajectory which is a continuous variable is considered as covariate. Analysis of covariance model is selected since our explanatory variables are a mox of categorical and continuous variables.

A linear relationship between the dependent variable and covariate is essential for a ANCOVA analysis which is evident from the scatterplot.
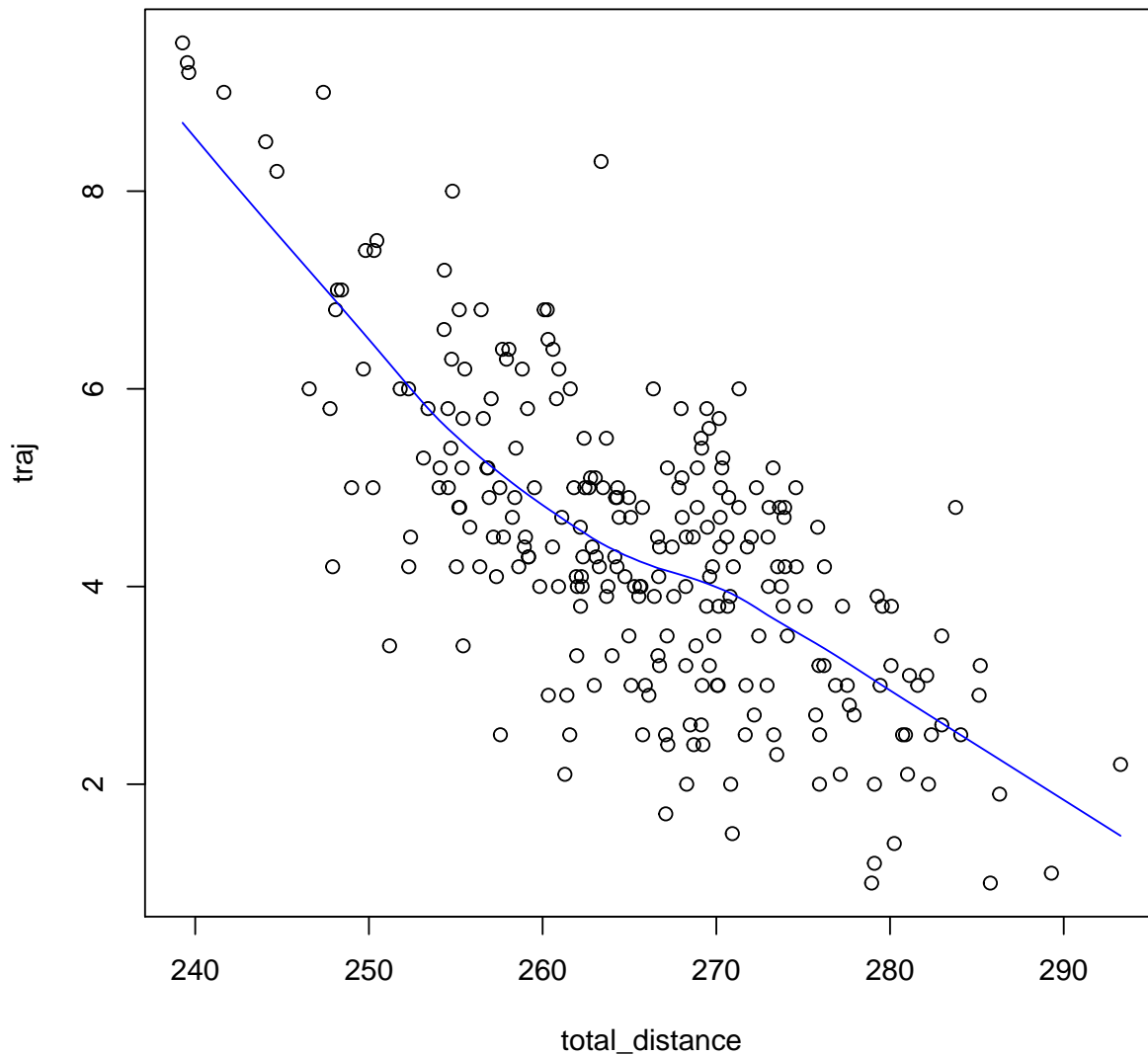
Analysis of Covariance

Df Sum Sq Mean Sq F value Pr(¿F) factor(ballmodel) 2 5803 2901.3 117.45 ¡ 2e-16 *** factor(block) 41 11749 286.6 11.60 ¡ 2e-16 *** c.traj 1 1810 1809.8 73.26 2.56e-15 *** Residuals 207 5114 24.7 — Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Analysis of Variance Table

Response: $total_{d}istanceDfSumSqMeanSqFvaluePr(>F)factor(ballmodel)25802.62901.29117.447 < 2.2e - 16 * * *factor(block)4111749.1286.5611.600 < 2.2e - 16 * * *c.traj11809.81809.8373.2642.564e - 15 * * * Residuals2075113.524.70 - - - Signif.codes : 0'***'0.001'**'0.01'*'0.05'.'0.1''1DfSumSqMeanSqFvaluePr(>$

$F)c.traj11215412154492.014 < 2e-16 ***factor(ballmodel)2107553721.7492.67e-09***factor(block)4161331506.055 < 2e-16***Residuals207511425---Signif.codes: 0'***'0.001'**'0.01'*'0.05'.'0.1''1 Analysis of Variance Table$

Response: $total_{d}istance Df SumSq MeanSq Fvalue Pr(> F)c.traj112154.212154.2492.0143 < 2.2e-16***factor(ballmodel)21074.5537.321.74872.672e-09***factor(block)416132.8149.66.0551 < 2.2e-16***Residuals2075113.524.7---Signif.codes: 0'***'0.001'**'0.01'*'0.05'.'0.1''1$

The ANCOVA table is as follows

SOV	df	SS	MSS	F	p-value
Trajectory	1	12154	12154	492.014	0
Ballmodel	2	1075	537	21.749	0
Block	41	6133	150	6.055	0
Residuals	207	5114	25		
Total	251	24476			

Table 1: ANCOVA table

Ball model, block and trajectory has a significant effect on the total distance traveled by a ball.
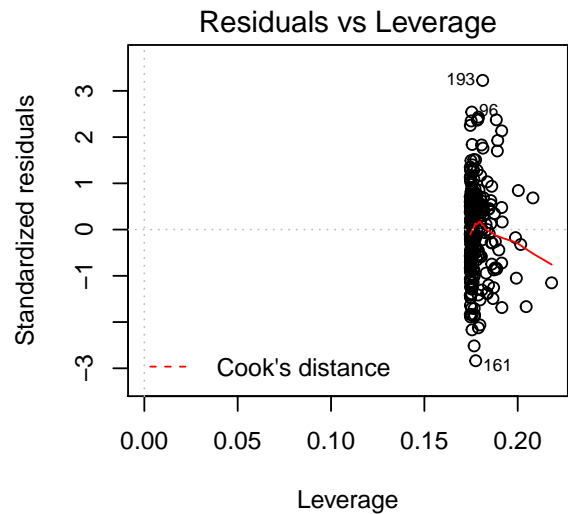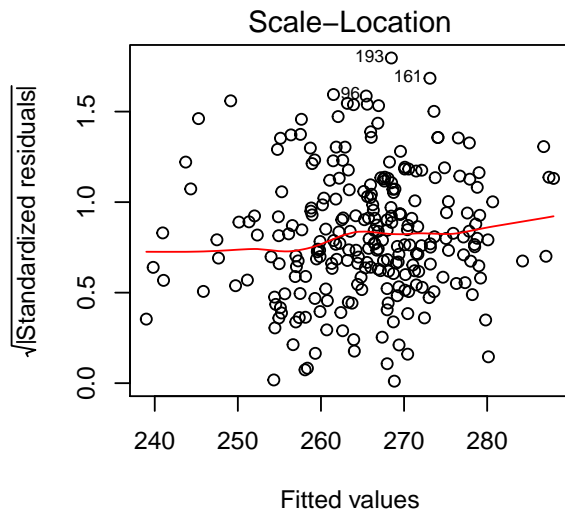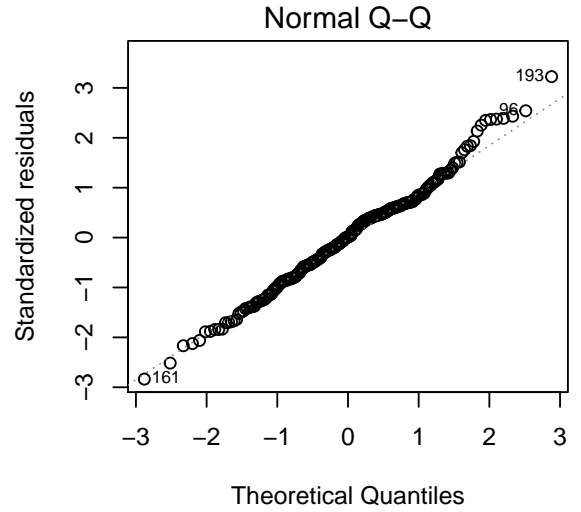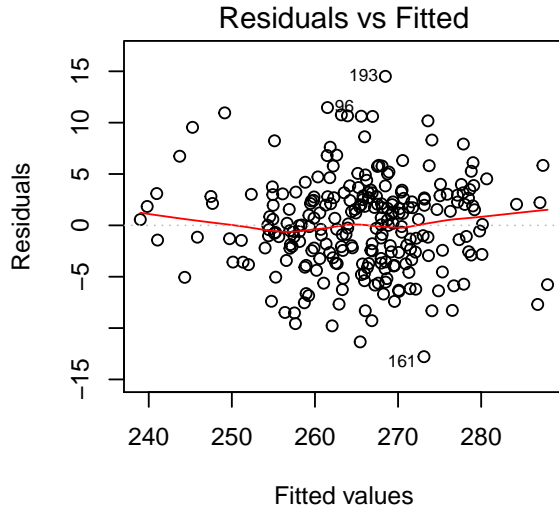
Since the diagnostic plots looks good with a couple of outliers which are not leverage points, residuals are fairly normally distributed.

Conclusion