Project -1

Par inc. golf ball

Model Report

Table of Contents

1 Project Objective

The objective of the report is to explore the Golf data set ("Golf") in R and generate insights about the data set. This exploration report will consists of the following:

- Importing the dataset in R
- Understanding the structure of dataset
- Formation of Hypothesis test
- Graphical exploration
- Descriptive statistics

2 Assumptions

- Sample Size: 40, No. of samples: 2, Alpha – 0.5
- Comparison of two data set to be performed to get the output of standard deviation and delta of mean .

- We have to conduct the paired t- test here since both sample are not independent (before /after).
- Also needs to be produced the descriptive statistical summaries of the data for each model

3 Exploratory Data Analysis – Step by step approach

A Typical Data exploration activity consists of the following steps:

1. Environment Set up and Data Import
2. Formulate and present the rationale for a hypothesis test
3. Paired t -test
4. Descriptive statistical summaries
5. Powered t- test

4 Solutions

4.1 Q1 . Question Formulate and present the rationale for a hypothesis test that par could use to compare the driving distances of the current and new golf balls

- The level of significance (Alpha) = 0.05.
- The sample size, N = 40 which is sufficiently large for a Zstat Test.
- But since the population standard deviation (Sigma) is unknown, we have to use a **T test.**
- Degree of Freedom: Since the sample is the same for both Sampling tests, we have **N-1 degrees of freedom : 39**
- Specify the level of significance. = 0.05 so z = 1.96

Hypothesis Formulation

Use the T Test: Paired two Sample for Means:
H0: μCurrent – μNew = 0 (New coating balls did not effect on driving distances significantly)
H1: μCurrent – μNew is not equal to 0 – (New coating balls effected on driving distances significantly)
Test statistics: Two sample t-test

4.2 . Q2. Analyze the data to provide the hypothesis testing conclusion. What is the p-value for your test? What is your recommendation for Par Inc.?

```
        Welch Two Sample t-test

data:  Current and New
t = 1.3284, df = 76.852, p-value = 0.906
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 6.253091
sample estimates:
mean of x mean of y
  270.275   267.500

> |
```

Conclusion:
In this scenario, the p value is 0.906 which is greater than the 0.05. Hence accept the null hypothesis So we conclude that population means driving distances of the current and new golf balls are equal

We will use the P-value approach. Current New 270.28 267.50 mean 8.75 9.90
std. dev. 40 40 n 78 df 2.775 difference (Current - New) 87.282
pooled variance 9.342 pooled std. dev. 2.089 standard error of difference 0 hypothesized difference

T= 1.33 p-value=.1879 (two-tailed) F-test for equality of variance 97.95 variance: New 76.61 variance: Current 1.28 F .4465 p-value

Looking at the descriptive statistics for each model, we can initially conclude that Current model has a longer range of distance based on the 40 samples with a mean of 270.275 compare to 267.500 for new model.

Beside, standard deviation of current is 8.75 and of new is 9.89. Although, two value of mean is different but Sd also is quite large so we can say they are quite same. P-value i=0.1879
> 0.05=α. Our decision rule for this problem is:

Do not reject H0,Mean distance of cut-resistant balls equals mean distance of current-model balls

The new cut-resistant balls have no difference in distance compared to the current-model one Therefore, we

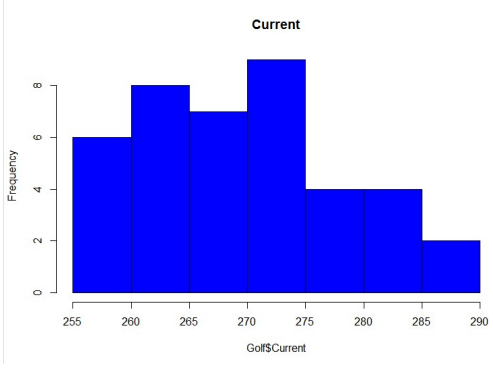
recommend for Par, Inc.: they should not launch this new product because the new model is not an improvement in distance compared to the old mode

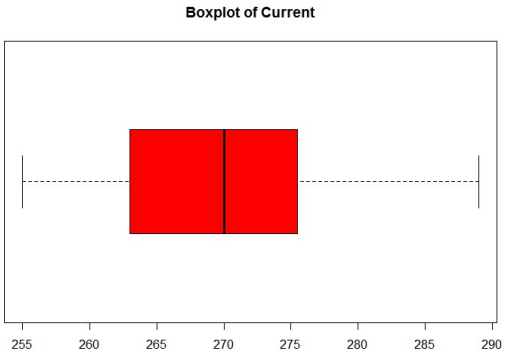
4.3 Q3. Provide descriptive statistical summaries of the data for each model

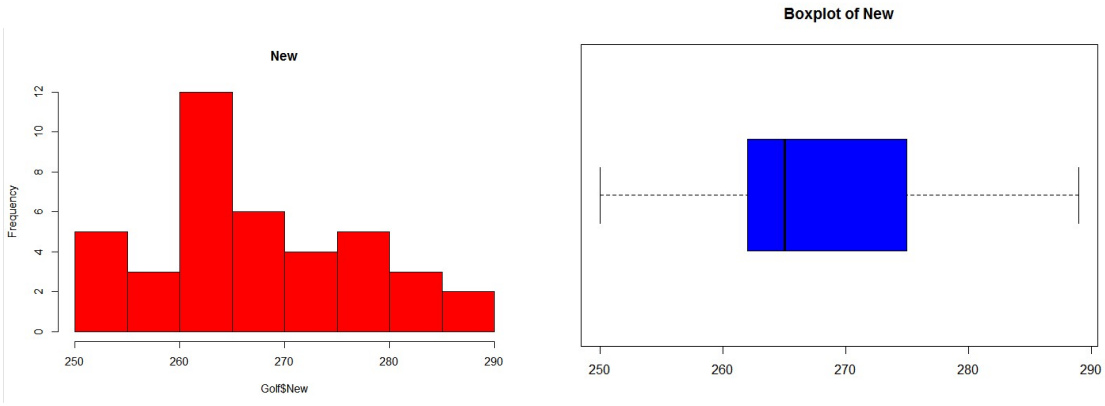
Five Point Summary and Standard Deviation on both the samples

```
> summary(Golf)
    Current          New
 Min.   :255.0   Min.   :250.0
 1st Qu.:263.0   1st Qu.:262.0
 Median :270.0   Median :265.0
 Mean   :270.3   Mean   :267.5
 3rd Qu.:275.2   3rd Qu.:274.5
 Max.   :289.0   Max.   :289.0
> sd(Golf$New)
[1] 9.896904
> sd(Golf$Current)
[1] 8.752985
> var(Golf$Current)
[1] 76.61474
> var(Golf$New)
[1] 97.94872
> SD = sd(Golf$Current-Golf$New)
> SD
[1] 13.74397
>
```

- Histogram and Box Plot for both the schemes:







Observations:
Both the samples seems to be normally distributed. The Current data looks more normally distributed with having minor left skewed, whereas the New data looks right skewed. there is a dip in the performance.

4.4 Q4. What is the 95% confidence interval for the population mean of each model, and what is the 95% confidence interval for the difference between the means of the two population?

For current model:

Confidence interval mean

95% Confidence level

270.275 mean

8.752984839

std. dev.

40 n

1.960 Z

2.713 half-width

272.988 upper confidence limit

267.562 lower confidence limit

For new model:

Confidence interval mean

95% confidence level ,

267.5 mean,

9.896904463 std. dev.,

40 n,

1.960 z

3.067 half-width

270.567 upper confidence limit

264.433 lower confidence limit

There is 95%probability that the true population mean is between 267.562 and 272.988 in current model while 264.433 and 270.567 in new model. And by their confidence intervals, we can say that there is no significance difference between the driving distances of Current Model and New Model

```
> power.t.test(n = 40, delta = 2.77,sd = 13.74, type = "paired", alternative = "one.sided",sig.level = .05)

     Paired t test power calculation

              n = 40
          delta = 2.77
             sd = 13.74
      sig.level = 0.05
          power = 0.3475506
    alternative = one.sided

NOTE: n is number of *pairs*, sd is std.dev. of *differences* within pairs
```

4.5 Q5. Do you see a need for larger sample sizes and more testing with the golf balls? Discuss

```
> power.t.test(power = .95, delta = 5,sd = 13.74, type = "paired", alternative = "one.sided",sig.level = .05

     Paired t test power calculation

              n = 83.09698
          delta = 5
             sd = 13.74
      sig.level = 0.05
          power = 0.95
    alternative = one.sided

NOTE: n is number of *pairs*, sd is std.dev. of *differences* within pairs
```

The larger the sample size the smaller the standard deviations which means point estimator of mean will become more precise. Hence there is no need to take larger sample size. Our recommendation for the company (Par, Inc.) is to continue with further study of the improved cut-resistant. However, besides testing the driving distance, Par Inc. should also consider other factors like costing, availability of suppliers and the like before finalizing and launching the product.

5. Appendix A – Source Code


```
> Golf <- read.csv("C:/Users/IBM_ADMIN/Documents/My Received Files/r file/R Programming/Golf.csv", header= TRUE)
> Golf
   Current New
1      264 277
2      261 269
3      267 263
4      272 266
5      258 262
6      283 251
7      258 262
8      266 289
9      259 286
10     270 264
11     263 274
12     264 266
13     284 262
14     263 271
15     260 260
16     283 281
17     255 250
18     272 263
19     266 278
20     268 264
21     270 272
22     287 259
23     289 264
24     280 280
25     272 274
26     275 281
27     265 276
28     260 269
29     278 268
30     275 262
31     281 283
32     274 250
33     273 253
34     263 260
35     275 270
36     267 263
37     279 261
38     274 255
39     276 263
40     262 279
```

```
> summary(Golf)
    Current          New
 Min.   :255.0   Min.   :250.0
 1st Qu.:263.0   1st Qu.:262.0
 Median :270.0   Median :265.0
 Mean   :270.3   Mean   :267.5
 3rd Qu.:275.2   3rd Qu.:274.5
 Max.   :289.0   Max.   :289.0
> sd(Golf$New)
[1] 9.896904
> sd(Golf$Current)
[1] 8.752985
> var(Golf$Current)
[1] 76.61474
> var(Golf$New)
[1] 97.94872
> SD = sd(Golf$Current-Golf$New)
> SD
[1] 13.74397
> hist(Golf$Current,col = "Blue",main="Current")
> hist(Golf$New, col= "Red",main= "New")
> boxplot(Golf$Current,horizontal=TRUE, col="Red",main="Boxplot of Current")
> boxplot(Golf$New,horizontal=TRUE, col="Blue",main="Boxplot of New")
> attach(Golf)
The following objects are masked from Golf (pos = 3):

    Current, New

> t.test(Current,New,paried = TRUE,conf.level = 0.95,alternative = "less")

        Welch Two Sample t-test

data:  Current and New
t = 1.3284, df = 76.852, p-value = 0.906
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 6.253091
sample estimates:
mean of x mean of y
  270.275   267.500

> delta = mean(Golf$Current -Golf$New)
> delta
[1] 2.775
> power.t.test(n = 40, delta = 2.77,sd = 13.74, type = "paired", alternative = "one.sided",sig.level = .05)

     Paired t test power calculation

              n = 40
          delta = 2.77
             sd = 13.74
      sig.level = 0.05
          power = 0.3475506
    alternative = one.sided

NOTE: n is number of *pairs*, sd is std.dev. of *differences* within pairs

> power.t.test(power = .95, delta = 5,sd = 13.74, type = "paired", alternative = "one.sided",sig.level = .05)

     Paired t test power calculation

              n = 83.09698
          delta = 5
             sd = 13.74
      sig.level = 0.05
          power = 0.95
    alternative = one.sided

NOTE: n is number of *pairs*, sd is std.dev. of *differences* within pairs

>
```