MINI PROJECT
- New Golf Ball

Model Report

By – Nikhil Rawal

GREAT LAKES
INSTITUTE OF MANAGEMENT

Project Overview

Hypothesis testing is an act in statistics whereby an analyst tests an assumption regarding a population parameter. The methodology employed by the analyst depends on the nature of the data used and the reason for the analysis. Hypothesis testing is used to infer the result of a hypothesis performed on sample data from a larger data available.

Par Inc., is a major manufacturer of golf equipment. Management believes that Par's market share could be increased with the introduction of a cutresistant, longer-lasting golf ball. Therefore, the research group at Par has been investigating a new golf ball coating designed to resist cuts and provide a more durable ball. The tests with the coating have been promising. One of the researchers voiced concern about the effect of the new coating on driving distances. Par would like the new cut-resistant ball to offer driving distances comparable to those of the current-model golf ball. To compare the driving distances for the two balls, 40 balls of both the new and current models were subjected to distance tests. The testing was performed with a mechanical hitting machine so that any difference between the mean distances for the two models could be attributed to a difference in the design.

Project Approach

- Exploratory Data Analysis
- Descriptive Statistics
- Data Visualization
- Hypothesis formation
- Selection of appropriate Hypothesis Testing method
- 95% Confidence Intervals
- Need of Larger Sample Size
- Conclusion and Recommendation.

- **Data Exploration**

#Set working directory

#getwd()

- "F:/golf"

Read Input File

- golf=read.csv("Golf.csv")

- Names of the columns

- names(golf)

"Current" , "New"

- Head(mydata)

```
  Current   New
    264     277
    261     269
    267     263
    272     266
    258     262
    283     251
```

- Dimension of data

- Dim(golf)

‘40’ ‘2’

- Structure of data

- str(golf)

```
'data.frame':   40 obs. of  2 variables:
 $ Current: int  264 261 267 272 258 283 258 266 259 270 ...
 $ New    : int  277 269 263 266 262 251 262 289 286 264 ...
```

- Summary of data

- 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
```

- Table(New)

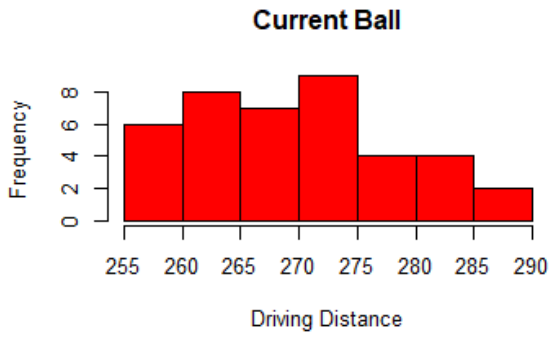
```
250 251 253 255 259 260 261 262 263 264 266 268 269 270 271 272 274
  2   1   1   1   1   2   1   4   4   3   2   1   2   1   1   1   2

276 277 278 279 280 281 283 286 289
  1   1   1   1   1   2   1   1   1
```

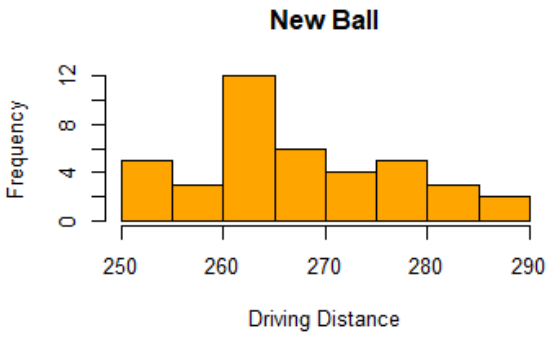
- Table(Current)

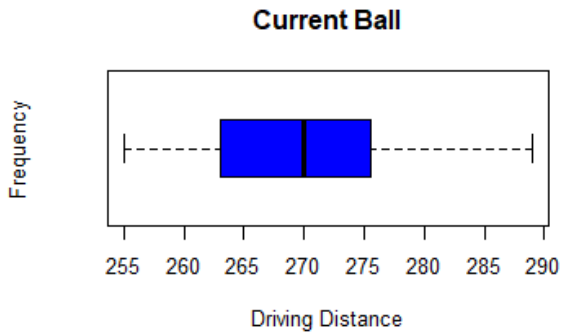
```
255 258 259 260 261 262 263 264 265 266 267 268 270 272 273 274 275
  1   2   1   2   1   1   3   2   1   2   2   1   2   3   1   2   3

276 278 279 280 281 283 284 287 289
  1   1   1   1   1   2   1   1   1
```

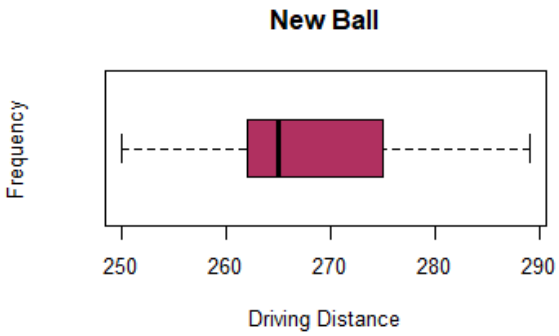
- **Data Visulaisation**

➢ Histogram & Boxplot









- <u>Key Observation</u>

• Hence, The average distance covered by ‘New’ golf ball is lower as compared to ‘Current’ golf ball

• ‘New’ golf ball observed to have relatively higher variation between 260 and 270.

• ‘Current’ golf ball shows the consistent in the frequency till 275.

- **Hypothesis formation**

- For formation of Hypothesis, we need null and Alternative hypothesis.

1. Null Hypothesis

A null hypothesis is a type of hypothesis used in statistics that proposes that no statistical significance exists in a set of given observations. The null hypothesis attempts to show that no variation exists between variables or that a single variable is no different than its mean. It is presumed to be true until statistical evidence nullifies it for an alternative hypothesis.

2. Alternative Hypothesis

In statistics, alternative hypothesis is stated as an alternative to the null hypothesis. The null hypothesis is then tested using one of the many statistical hypothesis tests.

Null Hypothesis (H0): $\mu 1 - \mu 2 = 0$ (i.e. they are the same)
Alternative Hypothesis (Ha): $\mu 1 - \mu 2 \neq 0$ (i.e. they are not the same)
Where,
• $\mu 1$: Mean driving distance of current model golf ball
• $\mu 2$: Mean driving distance of new golf ball

- Confidence Interval – t.test (Current, New)

```
data:  Current and New
t = 1.3284, df = 76.852, p-value = 0.188
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.384937  6.934937
sample estimates:
mean of x mean of y
  270.275   267.500
```

- t.test (Current)

One Sample t-test

```
data:  Current
t = 195.29, df = 39, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 267.4757 273.0743
sample estimates:
mean of x
  270.275
```

- t.test (New)

One Sample t-test

```
data:  New
t = 170.94, df = 39, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
264.3348 270.6652
sample estimates:
mean of x
267.5
```

- Larger Sample Size

- Pooled Standard deviation

```
delta=mean(Current)-mean(New)
> pooledSD <- (((40-1)*(8.75^2)+(40-1)*(9.9^2))/(40+40-2))^0.5
> delta
[1] 2.775
> pooledSD
[1] 9.342711
```

- Power t-test

```
  -  power.t.test(n=40, delta = 2.775, sd=9.342,
+               sig.level = 0.05,type = "two.sample",
+               alternative = "two.sided" )
```

```
     Two-sample t test power calculation

              n = 40
          delta = 2.775
             sd = 9.342
      sig.level = 0.05
          power = 0.258536
    alternative = two.sided
```

Hence, the Power of test is 25.8%, therefore there is only 25% chances that null hypothesis will not be rejected when it is false. We must try more number of samples to increase the power of test.

- Calculating the sample size again

```
  -  power.t.test(power=0.95, delta = 2.775,
+               sd=9.342,sig.level = 0.188,type =
"two.sample",
+               alternative = "two.sided" )
```

```
     Two-sample t test power calculation

              n = 199.2145
          delta = 2.775
             sd = 9.342
      sig.level = 0.188
          power = 0.95
    alternative = two.sided
```

- Conclusion

- From the given data, it may be concluded that, statistically there is no significance change in driving distance due to new coating on golf balls.
- The test be carried out with a larger sample size covering number of golf courses to improve the accuracy of the test results and negating any effect of one type of ground.
- The 95% confidence interval of population mean for Current model is between 267.4757 & 273.0743.
- The 95% confidence interval of population mean for New model is between 264.3348 & 270.6652.
- 2 Tail Hypothesis Test recommends to launch the new ball design into Production, the Power of Test is only 25%.
- With 95% Power of Test, it is recommended to have the number of samples as 200 for both the designs, and then conclude the Hypothesis test recommendations.
- The results need to interpreted and future actions be planned with the understanding of other characteristics like size, shape, weight etc.

Thank You

Source Code

- Set Working direotory

```
getwd()
```

- Read file

```
golf=read.csv("Golf.csv")
golf
attach(golf)
```

- dimension of data

```
dim(golf)
```

- Name of the Variable

```
names(golf)
```

- Structure of data

```
str(golf)
```

- Summary of the data

```
summary(golf)
head(golf)
```

- Finding Missing Value

```
colSums(is.na(golf))
table(New)
table(Current)
```

- Data Visulaiation - Histogram

```
par(mfrow=c(2,2))
hist(Current,main='Current Ball',xlab = "Driving Distance",ylab = "Frequency",col = "red")
hist(New,main='New Ball',xlab = "Driving Distance",      ylab = "Frequency",col = "orange")
```

- Data Visualisation - Boxplot

```
boxplot(Current,main='Current Ball',xlab = "Driving Distance", ylab = "Frequency",col = "blue",horizontal = TRUE)
boxplot(New,main='New Ball',xlab = "Driving Distance", ylab = "Frequency",col = "maroon",horizontal = TRUE)
```

- Mean, Standard, Variance
  (of New and Cuurent Ball)

- Col_Head= c("Mean","Standard deviation","Variance","Remark")
- Current_Stats= c(round(mean(Current),digits = 2),
  - round(sd(Current),digits = 2),
  - round(var(Current),digits = 2),
  - "Current Ball")
- New_Stats= c(round(mean(New), digits = 2),
  - round(sd(New), digits = 2),
  - round(var(New), digits = 2),
  - "New Ball Stats")

- Combined_Stats= rbind(Col_Head,Current_Stats,New_Stats)
- Combined_Stats

- Two sample t-test(Current, New)
  With 95% confidence interval for the difference in means

- T-test(Current,New)

- t.test(Current)
  (One ample t-test to get 95% Confidence Interval)

- t.test(New)
  (One ample t-test to get 95% Confidence Interval)

- Calculation to see the larger sample size

- delta=mean(Current)-mean(New)
- pooledSD <- (((40-1)*(8.75^2)+(40-1)*(9.9^2))/(40+40-2))^0.5
- delta
- pooledSD

- Power t-test

- power.t.test(n=40, delta = 2.775, sd=9.342,
  - sig.level = 0.05,type = "two.sample",
  - alternative = "two.sided")

- Power t-test (Sample size Required)

- power.t.test(power=0.95, delta = 2.775,
  - sd=9.342,sig.level = 0.188,type = "two.sample",
  - alternative = "two.sided")

Thank You

The END
