MINI PROJECT – Par Inc.

**Table of Contents**

1 Project Objective

2 Assumptions

3 Explanatory Data Analysis – Step by step approach

* 1. Environment Setup & Importing Data
     1. Install necessary packages and invoke libraries
     2. Set up working directory
     3. Import & read the data
  2. Variable Identification
  3. Univariate Analysis
  4. Bi-Variate Analysis

4 Hypothesis Formulation and Testing

5 Conclusion

6 Appendix A – Source Code

**1 Project Objective –**

The objective of this report is to analyse the Par Inc. dataset in R and generate insights about the data. This reports will consist of:

1. Importing the dataset into R
2. Understanding the structure of dataset.
3. Graphical Inferences
4. Descriptive Statistics
5. Insights/inferences from the dataset.

**2 Assumptions –**

We assume that:-

1. The purpose of using the hitting machine to drive the ball is to apply equal force on the balls in both cases.
2. The degree of change in the driving distance of 5 yard for the current and the new golf ball is acceptable to calculate the Type I and Type II errors and required sample size.
3. The air-flow during the test is uniform and has no effect on the change in driving distance.

**3 Exploratory Data Analysis (Step-by-step approach) –**

* 1. Environment Set Up and Data Import:-
     1. Install necessary packages and invoke libraries.

Install.packages(“dplyr”, dependencies = TRUE)

Library(dplyr)

3.1.2 Set the working directory:-

Getwd()

Setwd(“C:\\Users\\Aparajita Mukherjee\\Documants\\R”)

3.1.3 Import & Reading data from File:-

golf <- read.csv(“Golf.csv”)

view(golf)

3.2 Variable Identification:–

Dim – The dim function of the R programming language returns the dimension (e.g. the number of columns and rows) of a matrix, array or data frame.

Head – returns the first n rows of a matrix or data frame in R

Tail – returns the last n rows of a matrix or data frame in R

Summary – used to produce result summaries of the results of various model fitting functions.

Str – Compactly displays the internal structure of an R object, a diagnostic function and an alternative to summary(). Ideally, only one line for each ‘basic’ structure is displayed.

Names – Functions to get or set the names of an object.

3.2.1 Variable Identification Inferences:-

dim(golf)

head(golf, 10)

tail(golf, 10)

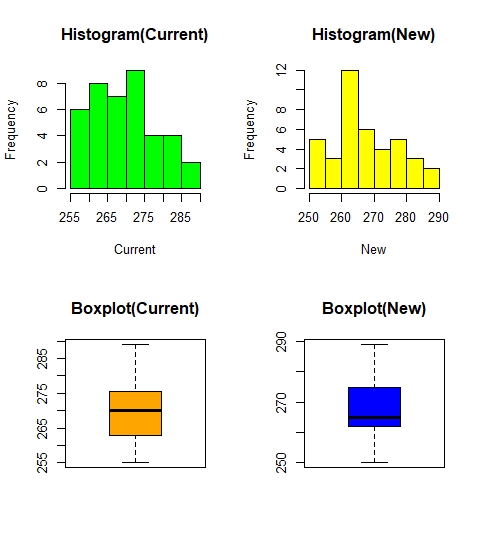
summary(golf)

str(golf)

names(golf)

3.3 Univariate Analysis:-

|  |  |  |  |
| --- | --- | --- | --- |
| **Measure** | **Formula Used** | **Current** | **New** |
| Mean | =AVERAGE(A$2:A$41) | 270.275 | 267.5 |
| Median | =MEDIAN(A2:A41) | 270 | 265 |
| Mode | =MODE(A2:A41) | 272 | 263 |
| Standard Deviation | =STDEV(A$2:A$41) | 8.752984839 | 9.896904463 |
| Standard Variance | =VAR(A2:A41) | 76.61474359 | 97.94871795 |
| IQR 1 | =QUARTILE(A2:A41,1) | 263 | 262 |
| IQR 3 | =QUARTILE(A2:A41,3) | 275.25 | 274.5 |
| IQR | =F8-F7 | 12.25 | 12.5 |
| alpha |  | 0.05 | 0.05 |
| Standard Deviation | =STDEV(A$2:A$41) | 8.752984839 | 9.896904463 |
| Sample(n) |  | 40 | 40 |
| Standard Error | =STDEV(A$2:A$41)/SQRT(COUNT(A$2:A$41)) | 1.383968421 | 1.564837994 |
| Confidence | =CONFIDENCE(F10,F11,F12) | 2.71252826 | 3.067026111 |
| NormSInv | =NORMSINV(1-F10/2)\*F11/SQRT(F12) | 2.71252826 | 3.067026111 |
| 95% Confidence Level | =TDIST(F15,39,2) | 0.009883984 | 0.003917741 |
| T-Test(Independent 2-sample) | =TTEST(A2:A41,B2:B41,2,3) | 0.187989945 |  |



1. Comparing the means of the two sample distributions, we see that even though visually it seems as if New coating has effect on the driving distances but statistically it does not.
2. The difference in mean in the case of new balls can also be attributed to the higher variance compared to Current balls.
3. Statistically there is no effect of new coating on driving distances. Though it is suggested to check the effect on the weights and other characteristics like size and shape of the new balls.
4. Also, the given sample is from only one golf course, It is advisable that test should perform on different kind of golf courses to take care of the differences in grounds.

3.4 Bi-Variate Analysis:-

Independent Two Sample t-test –

t.test(Current, New, alternative = “two.sided”, var.equal = FALSE)

**4 Hypothesis Formulation and Testing –**

1. Null Hypothesis - ,

Alternate Hypothesis -

1. The level of significance (Alpha) = 0.05
2. The sample size N = 40 which is sufficiently large for a Z stat Test.
3. But since the population standard deviation (Sigma) is unknown, we have to use a T stat Test.
4. Since the sole purpose of the test is to check whether there is any effect on driving distances due to the new coating, we could prefer a Two Tailed T Test.

**5 CONCLUSION –**

* In this scenario, the p-value is 0.188 which is greater than 0.05, therefore we will not reject the null hypothesis.
* Thus accepting the null hypothesis, that there is no significant change in driving distances due to the new coating.
* If we compare the means of the two sample distributions, we see that even though visually it seems as if the new coating has effect on the driving distances but statistically it does not.
* The difference in mean in the case of new balls can also be attributed to the higher variance compared to Current balls.
* Statistically there is no effect of new coating on driving distances. Though it is suggested to check the effect on the weights and other characteristics like size and shape of the new balls.

**6 Appendix A – Source Code –**

getwd()

setwd(“C:\\Users\\Aparajita Mukherjee\\Documents\\R”)

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

str(golf)

summary(golf)

names(golf)

attach(golf)

par(mfrow=c(2,2))

hist(Current, main = "Histogram(Current)", col = "Green")

hist(New, main = "Histogram(New)", col = "Yellow")

boxplot(Current, main = "Boxplot(Current)", col = "Orange")

boxplot(New, main = "Boxplot(New)", col = "Blue")

t.test(Current,New, alternative = "two.sided", var.equal = FALSE)