**ADVANCED STATISTICS module**

**(Probability & hypothesis testing)**

BUSINESS

REPORT

**Prepared by: AMAYIKA**

**Great Learning.**

## Email- Kashyap.amayika009@gmail.com

September 30th, 2022

Table of Contents

[Problem -1 – Data of A physiotherapist with a male football team 5](#_bookmark0)

[1.1 What is the probability that a randomly chosen player would suffer an injury? 5](#_bookmark1)

[1.2 What is the probability that a player is a forward or a winger? 6](#_bookmark2)

[1.3 What is the probability that a randomly chosen player plays in a striker position and has a foot injury? 7](#_bookmark3)

[1.4 What is the probability that a randomly chosen injured player is a striker? 8](#_bookmark4)

[1.5 What is the probability that a randomly chosen injured player is either a forward or an attacking midfielder? 8](#_bookmark5)

[Problem – 2 Probability that an accident at a nuclear power plant will result in radiation leakage 8](#_bookmark6)

[2.1 What are the probabilities of a fire, a mechanical failure, and a human error respectively? 9](#_bookmark11)

[2.2 What is the probability of a radiation leak? 10](#_bookmark12)

[2.3 If there has been a radiation leak, for which the definite cause is not known. What is the probability that it has been caused by: A Fire , A Mechanical Failure, A Human Error 10](#_bookmark14)

[Problem -3 The breaking strength of gunny bags used for packaging cement 11](#_bookmark18)

[3.1 What proportion of the gunny bags have a breaking strength less than 3.17 kg per sq cm? 12](#_bookmark21)

[3.2 What proportion of the gunny bags have a breaking strength at least 3.6 kg per sq cm.? 13](#_bookmark23)

[3.3 What proportion of the gunny bags have a breaking strength between 5 and 5.5 kg per sq cm.? 13](#_bookmark24)

[3.4 What proportion of the gunny bags have a breaking strength NOT between 3 and 7.5 kg per sq cm.? 14](#_bookmark26)

[Problem – 4 Grades of the final examination in a training course 15](#_bookmark27)

[4.1 What is the probability that a randomly chosen student gets a grade below 85 on this exam? 16](#_bookmark28)

[4.2 What is the probability that a randomly selected student scores between 65 and 87? 17](#_bookmark18)

4.3 What should be the passing cut-off so that 75% of the students clear the exam? [18](#_bookmark23)

[Problem – 5 Zingaro company Data 19](#_bookmark24)

5.1 According to Zingaro Why the unpolished stones may not be suitable for printing? [21](#_bookmark26)

[5.2 Is the mean hardness of the polished and unpolished stones the same? 24](#_bookmark27)

[Problem – 6 Aquarius health club data 25](#_bookmark28)

6.1 Using the sample data provided can you conclude whether the program is successful? [26](#_bookmark23)

[Problem – 7 Dental implant data 30](#_bookmark24)

7.1 Test whether there is any difference among the dentists on the implant hardness. [32](#_bookmark26)

[7.2  Test whether there is any difference among the dentists on the implant hardness. 34](#_bookmark27)

[7.3 What do you conclude regarding whether implant hardness depends on dentists? State your conclusion. 35](#_bookmark28)

7.4 Test whether there is any difference among the methods on the hardness of dental implant, separately for the two types of alloys. What are your conclusions? [37](#_bookmark26)

[7.5 Test whether there is any difference among the temperature levels on the hardness of dental implant, separately for the two types of alloys. What are your conclusions?. 39](#_bookmark27)

[7.6 Consider the interaction effect of dentist and method and comment on the interaction plot, separately for the two types of alloys? Is it possible to identify which dentists are different, which methods are different, and which interaction levels are different? 40](#_bookmark28)

[7.7 Is it possible to identify which dentists are different, which methods are different, and which interaction levels are different? 42](#_bookmark28)

List of Tables

[Table 1. Male Football Team Data 6](#_bookmark7)

[Table 2. Zingaro company data set 19](#_bookmark8)

[Table 3. Summary of Zingaro Data set – before treatment 20](#_bookmark9)

[Table 4. Information of datatype – Zingaro company 20](#_bookmark10)

[Table 5. Null value counts – Zingaro company 20](#_bookmark13)

[Table 6. Aquarius company dataset 26](#_bookmark15)

[Table 7. Aquarius company summary and information 26](#_bookmark16)

[Table 8. Aquarius company null values 27](#_bookmark17)

[Table 9. Aquarius company datatypes 27](#_bookmark19)

[Table 10. Dental Implant Dataset 30](#_bookmark20)

[Table 11. Dental Implant Dataset information 30](#_bookmark22)

[Table 12. Dental Implant Dataset summary 31](#_bookmark25)

[Table 13. Dental Implant Dataset null values 31](#_bookmark7)

[Table 14. Dental Implant Dataset categories and datatypes 31](#_bookmark8)

[Table 15. Dentists and its response with alloy 32](#_bookmark9)

[Table 16. Dentists and its response with alloy 1 33](#_bookmark10)

[Table 17. Dentists and its response with alloy 1- interaction hypothesis 33](#_bookmark13)

[Table 18. Dentists and its response with alloy 2 33](#_bookmark15)

[Table 19. Dentists and its response with alloy 2 – interaction hypothesis 34](#_bookmark16)

[Table 20. Shapiro and Levene Test value table 35](#_bookmark17)

[Table 21. Dentists and its response with alloy 1 & 2 together 35](#_bookmark19)

[Table 22. Dentists and its response with alloy 1 36](#_bookmark20)

[Table 23. Dentists and its response with alloy 2 36](#_bookmark22)

[Table 24. Methods and its response with alloys 37](#_bookmark25)

[Table 25. Methods and its response with alloy 1 38](#_bookmark7)

[Table 26. Methods and its response with alloy 2 38](#_bookmark8)

[Table 27. Methods and its response with alloy - interaction hypothesis 39](#_bookmark9)

[Table 28. Temperature and its response with alloy 1 39](#_bookmark10)

[Table 29. Temperature and its response with alloy 2 39](#_bookmark13)

[Table 30. Temperature and its response with alloy - interaction hypothesis 40](#_bookmark15)

[Table 31. Dentist and Method Interaction for Alloy -1 40](#_bookmark16)

[Table 32. Dentist and Method Interaction for Alloy -2 41](#_bookmark17)

[Table 33. Dentist and Method Together for Alloy -1 42](#_bookmark19)

[Table 34. Dentist and Method Together for Alloy -2 43](#_bookmark20)

[Table 35. Overall Interaction Table 43](#_bookmark22)

List of Equations and Formulas

[Equation 1. Probability Formula 6](#_bookmark8)

[Equation 2. Mutually Exclusive and Disjoint Events – Probability formula 7](#_bookmark9)

[Equation 3. Conditional Probability Formula 8](#_bookmark10)

[Equation 4. Baye’s Theorem 9](#_bookmark13)

[Equation 5. Total Probability Theorem 10](#_bookmark15)

[Equation 6. Test statistics 12](#_bookmark16)

[Equation 7. Test statistics- for two sampled t-test – one tailed 13](#_bookmark17)

[Equation 8. Test statistics- for two sampled t-test – two tailed 23](#_bookmark19)

[Equation 9. Test statistics- for two sampled t-test for question 5.2 25](#_bookmark25)

[Equation 10. Standard Error 28](#_bookmark20)

[Equation 10. Test Statistics for paired T-Test 29](#_bookmark20)

[Equation 11. Test statistics- for paired t-test 29](#_bookmark22)

List of Figures

[Figure 1. Normal distribution curve of breaking strength of gunny bags 11](#_bookmark9)

[Figure 2. Normal distribution curve for breaking strength less than 3.17 kg per sq. cm. 12](#_bookmark10)

[Figure 3. Normal distribution curve for breaking strength at least 3.6 kg per sq. cm 13](#_bookmark13)

[Figure 4. Normal distribution curve of breaking strength between 5 kg per sq. cm and 5.5 kg per sq. cm 14](#_bookmark15)

[Figure 5. Normal distribution curve of breaking strength not between 3 & 7.5 kg per sq. cm 15](#_bookmark16)

[Figure 6. Normal distribution curve of grades 15](#_bookmark17)

[Figure 7. Normal distribution curve of grades (with clarity) 16](#_bookmark19)

[Figure 8. Normal Distribution Area less than 85 marks 16](#_bookmark20)

[Figure 9. Area less than 85 marks 17](#_bookmark20)

[Figure 10. Area that a randomly selected student score between 65 and 87 17](#_bookmark16)

[Figure 11. Area that a randomly selected student score between 65 and 87 (zoomed in) 18](#_bookmark17)

[Figure 12. Area such that passing cut-off so that 75% of the students clear the exam 19](#_bookmark19)

[Figure 13. Unpolished Stone Histogram 21](#_bookmark20)

[Figure 14. Polished and Treated Stone Histogram 21](#_bookmark22)

[Figure 15. Boxplot of Polished and Unpolished treated 21](#_bookmark17)

[Figure 16. Normal Density Curve of Unpolished 22](#_bookmark19)

[Figure 17. Normal Density Curve of Polished 22](#_bookmark20)

[Figure 18. Plot for 5.1- One sampled T-Test 23](#_bookmark22)

[Figure 19. Plot for 5.2- Two sampled Two Tailed T-Test 25](#_bookmark22)

[Figure 20. Boxplot For Aquarius gym Untreated Outliers 27](#_bookmark17)

[Figure 21. Boxplot For Aquarius gym Treated Outliers 28](#_bookmark19)

[Figure 22. Boxplot For Aquarius gym Treated Outliers 28](#_bookmark20)

[Figure 23. Plot for Paired T-test 29](#_bookmark22)

[Figure 24. Dental Response Boxplot with outliers 31](#_bookmark17)

[Figure 25. Dental Response Point Plot for Alloy -1 33](#_bookmark19)

[Figure 26. Dental Response Box Plot for Alloy -1 33](#_bookmark20)

[Figure 27. Dental Response Point Plot for Alloy -2 34](#_bookmark22)

[Figure 28. Dental Response Box Plot for Alloy -2 34](#_bookmark22)

[Figure 29. Interaction Point Plot for Both Alloy 1 & 2 36](#_bookmark25)

[Figure 30. Response Point Plot for Alloy -1 with respect to Dentist 36](#_bookmark22)

[Figure 31. Response Point Plot for Alloy -2 with respect to Dentist 36](#_bookmark25)

[Figure 32. Method Response Point Plot for Alloy -1 38](#_bookmark22)

[Figure 33. Method Response Point Plot for Alloy -2 38](#_bookmark25)

[Figure 34. Interaction between Method & Dentist - Boxplot 40](#_bookmark22)

[Figure 35. Interaction between Method & Dentist- alloy 1 41](#_bookmark25)

[Figure 36. Interaction between Method & Dentist- alloy 2 41](#_bookmark25)

[Figure 37. Method & Alloy – overall 42](#_bookmark25)

[Figure 38. Dentist & Alloy – overall 42](#_bookmark22)

[Figure 39. Method Dentist for Alloy 1– Boxplot 43](#_bookmark25)

[Figure 40. Method Dentist for Alloy 2– Boxplot 43](#_bookmark25)

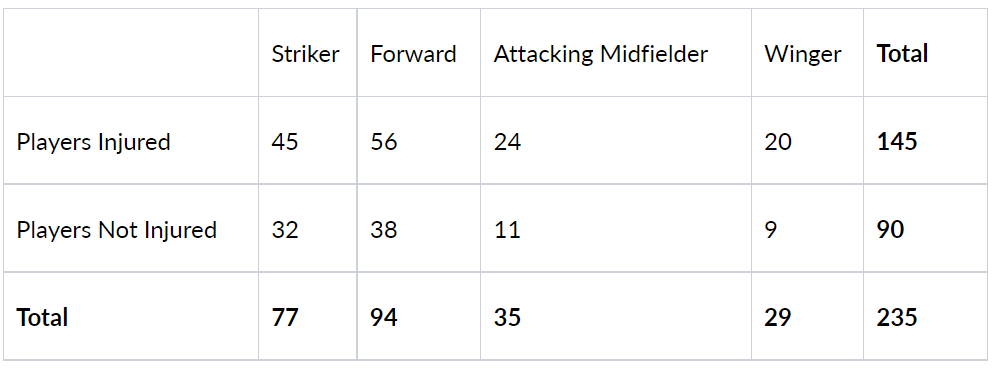
[Figure 41. Plot for Method & Dentist together -7.7 44](#_bookmark25)

# PROBLEM – 1

# Probability calculations

A physiotherapist with a male football team is interested in studying the relationship between foot injuries and the positions at which the players play from the data collected.

Table 1 - Male football Team Data



**Probability refers to chance or likelihood of a particular event taking place.**

**Therefore**

**P(A) = m/n**

**Equation 1**

**where m = number of ways that are favourable to the occurrence of A**

**n = total number of outcomes of the experiment**

**0<= P(A) => 1**

**Impossibility = 0% chance**

**Certainty = 100% chance,**

**Hence the probability lies between 0 to 1**

**Some of the probabilities are naturally expressed in terms of tables with rows and columns which has margins called contingency tables. The term marginal is used to indicate that the probabilities are calculated using a contingency table (also called joint probability table). Hence it is also termed as Marginal probability.**

**1.1 What is the probability that a randomly chosen player would suffer an injury?**

#### Here Total injured players(m) = 145

#### Total no of players present(n) = 235,

#### hence, the probability that a randomly chosen player is injured can be calculated as

#### **P(I) = m/n = 145/235,** which comes out to be **0.617 or 61.7%**

**1.2 What is the probability that a player is a forward or a winger?**

#### If the two events are **mutually exclusive** but are **mutually disjoint**, which means there is no commonality between them, then we can calculate the probability using the addition rule, because the intersection between the two events is 0 in that case.

#### Hence the formula can be given as

#### **P(A∪B) = P(A) + P(B) - P(A∩B),**

**If, P(A∩B) = 0**

**Then, P(A∪B) = P(A) + P(B)**

**Equation 2**

#### where P(A) is Probability of event A happening, P(B) is Probability of event B happening, and P(A∩B) is Probability of happening of both A and B. In this case P(A∩B) = 0, since the events are mutually disjoint

### Therefore, we can say that probability of players either forward or winger

### P (Either Forward or Winger) = P(F) + P(W)

P (F) = 94/235 = 0.4

P (W) = 29/235 = 0.123

Hence, the Probability that a player is either a forward or a winger is **0.5234 or 52.3%**

**1.3 What is the probability that a randomly chosen player plays in a striker**

**position and has a foot injury?**

### This is a case of Joint probability, where the player we are choosing randomly has to be in a striker position and then we have to calculate the probability that he has a foot injury. so, it will be calculated directly by dividing 45 to 235.

### Therefore, P (striker position and has foot injury) = 45/235

Probability of player who is in a striker position and has a foot injury is **19.1%**

**1.4 What is the probability that a randomly chosen injured player is a striker?**

This is a case of **conditional probability** which can be denoted as

**P (B|A) = P(A∩B) / P (A)**

**Where P (B|A) = probability of B given A**

**And, P(A∩B) = joint probability or intersection of B and A**

* **Equation 3**

P (striker and injured) = (45/235) = 0.191

P (injured) = (145/235) = 0.617

Therefore, P (injured player is striker) = P (striker and injured) / P (injured)

Probability of injured player who is striker is **0.3103 or 31.0%**

**1.5 What is the probability that a randomly chosen injured player is either a**

**forward or an attacking midfielder?**

Total injured = 145

P (injured who is forward) = (56/145) = 0.386

P (injured who is attacking midfielder) = (24/145) = 0.166

P (injured who is either forward or attacking midfielder) = P (injured who is forward) + P (injured who is attacking midfielder) = 0.386 + 0.166 = 0.5520

Proportion that randomly chosen injured player is either a forward or an attacking midfielder is **55.2%**

# PROBLEM – 2

# Probability calculations based on Baye’s Theorem

# An independent research organization is trying to estimate the probability that an accident at a nuclear power plant will result in radiation leakage. The types of accidents possible at the plant are, fire hazards, mechanical failure, or human error. The research organization also knows that two or more types of accidents cannot occur simultaneously.

# Let’s first define the events which are given to us along with their probabilities –

# **F = Fire**

# **M = Mechanical Failure**

# **H = Human Error**

# **R = Radiation Leak**

# **N = No Accident**

# P (R | F) = Probability of R given F = 20% = 0.2

# P (R | M) = Probability of R given M = 50% = 0.5

# P (R | H) = Probability of R given H = 10% = 0.1,

# P (R ∩ F) = Probability of R and F = 0.1% = 0.001

# P (R ∩ M) = Probability of R and M = 0.15% = 0.0015

# P (R ∩ H) = Probability of R and H = 0.12% = 0.0012,

Hence, the formula for **BAYE’S THEOREM** can be written as:

**P (A | B) = P(B|A) \* P(A) / P(B)**

**Where P (A|B) = Probability of A given B,**

**P (B|A) = Probability of B given A, P (A) = Probability of A**

**And P (B) = Probability of B**

**-Equation- 4**

The formula for Conditional Probability can be given as:

**P (A ∩ B) = P(B) \* P (A | B) -> given the fact that (If B Has ALREADY HAPPENED)**

**2.1 What are the probabilities of a fire, a mechanical failure, and a human error respectively?**

# First, we will calculate the probability of Fire

# P (R ∩ F) = P(F) \* P (R | F)

# Hence the probability of FIRE = P (R ∩ F)/ P (R | F)

# P (R and F together) = 0.001

# P (R given F) = 0.2

# P (F) = (0.001/0.2) = 0.005

Probability of FIRE is **0.5%**

**Now let’s calculate the probability of mechanical failure -**

P (R ∩ M) = P (M) \* P (R | M)

Hence the probability of MECHANICAL FAILURE = P (R ∩ M)/ P (R | M)

P (R and M together) = 0.0015

P (R given M) = 0.5

P (M) = (0.0015/0.5) = 0.003

Probability of MECHANICAL FAILURE is **0.3%**

**And lastly let’s calculate the probability of human error –**

P (R ∩ H) = P (H) \* P (R | H)

Hence the probability of MECHANICAL FAILURE = P (R ∩ H)/ P (R | H)

P (R and H together) = 0.0012

P (R given H) = 0.1

P (H) = (0.0012/0.1) = 0.012

Probability of HUMAN ERROR is **1.2%**

**2.2 What is the probability of a radiation leak?**

#### Since the probability of "No ACCIDENT" is not given, we can calculate it by subtracting the TOTAL ACCIDENT probabilities from 1, because the total probability is always equal to 1. Here in our case accident can occur by 3 ways the probabilities of which we have already calculated in the above question.

**P (N) = 1 – (P (H) + P (F) + P (M))**

Probability of “No Accident” is **0.98**

#### Also, since in case of "No ACCIDENT" no radiation will Leak so it can be termed as 0, hence P (R | N) = 0

P (R and N together) = 0, Now using the formula of total probability we can calculate the radiation leak probability

**P (R) = (R ∩ H) + P (R ∩ F) + P (R ∩ M) + P (R ∩ N)**

**– Equation 5**

P (R) = 0.001 + 0.0015 + 0.0012 + 0

Probability of Radiation Leak is **0.0037**

**2.3 Suppose there has been a radiation leak in the reactor for which the definite cause is not known. What is the probability that it has been caused by:**

* **A Fire.**
* **A Mechanical Failure.**
* **A Human Error.**

### A. Probability that the radiation leak in the reactor is caused by FIRE which implies P (F | R), probability of FIRE given the Radiation Leak can be written as:

#### P (F | R) = P (R | F) \* P(F) / P(R) - according to BAYES Theorem

P(R | F) = 0.2

P (F) = (0.001/0.2) = 0.005

P (R) = 0.0037

P (F | R) = ((0.2 \* 0.005)/0.0037) = 0.270270

The Probability of FIRE when there is already a Radiation leakage is **0.270270**

### B. Probability that the radiation leak in the reactor is caused by MECHANICAL FAILURE which implies P (M| R), probability of MECHANICAL FAILURE given the Radiation Leak can be written as:

#### P (M | R) = P (R | M) \* P(M) / P(R) - according to BAYES Theorem

P(M | F) = 0.5

P (M) = (0.0015/0.5) = 0.003

P (R) = 0.0037

P (M | R) = ((0.5 \* 0.003)/0.0037) = 0.4054

The Probability of FIRE when there is already a Radiation leakage is **0.4054**

### C. Probability that the radiation leak in the reactor is caused by HUMAN ERROR which implies P (H | R), probability of HUMAN ERROR given the Radiation Leak can be written as:

#### P (H | R) = P (R | H) \* P(H) / P(R) - according to BAYES Theorem

P(H | F) = 0.1

P (H) = (0.0012/0.1) = 0.012

P (R) = 0.0037

P (H | R) = ((0.1 \* 0.012)/0.0037) = 0.3243

The Probability of FIRE when there is already a Radiation leakage is **0.3243**

# PROBLEM – 3

### Breaking strength of gunny bags

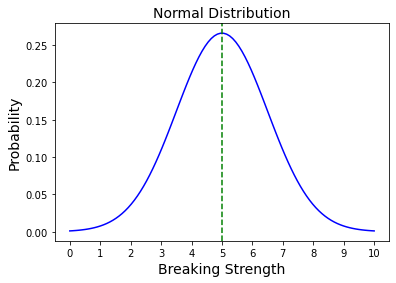
# The breaking strength of gunny bags used for packaging cement is normally distributed with a mean of 5 kg per sq. centimeter and a standard deviation of 1.5 kg per sq. centimeter. The quality team of the cement company wants to know the following about the packaging material to better understand wastage or pilferage within the supply chain; Answer the questions below based on the given information.

# Since cement is normally distributed so we will use “norm” feature from Scipy package of python

The mean (mu) is 5

The standard deviation (sigma) is 1.5

Then we will calculate the PDF (Probability Density Function) of breaking strength of gunny bags and plot the probability distribution function of normal distribution.

fig -1

### 3.1 What proportion of the gunny bags have a breaking strength less than 3.17 kg per sq. cm?

### The required probability can be written as P (X <= 3.17)

### We will set the values of population mean and population standard deviation to 5 and 1.5 respectively

mu (µ), sigma (σ) = 5, 1.5

### Then we will set the value sample mean that is x-bar to 3.17

### x-bar = 3.17

### Now let’s calculate the test statistic for the given values

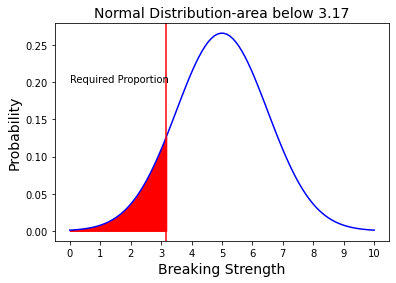
**Test statistic = (x-bar - mu) / (sigma) – Equation 6**

= (3.17- 5)/1.5 = -1.22

Hence, we can calculate the probability using “norm.cdf” function where CDF stands for cumulative distribution function in the python. We can either put the value of test statistic directly into the “norm.cdf” function or we can put the values of mu, sigma, and x-bar and calculate the probability which comes out as follows:

**Cumulative distributive function (x-bar, location=µ, scale=σ)**

**The probability of the gunny bags having a breaking strength less than 3.17 kg per sq. cm is 0.1112**

fig-2

In other words, 0.1112 can interpreted as that there are **11.1%** chance that the gunny bags have breaking

strength less than 3.17Kg per sq. cm when randomly chosen from the company.

**3.2 What proportion of the gunny bags have a breaking strength at least 3.6 kg per sq. cm.?**

The required probability can be written as P (X >= 3.6)

### We will set the values of population mean and population standard deviation to 5 and 1.5 respectively

mu (µ), sigma (σ) = 5, 1.5

### Then we will set the value sample mean that is x-bar to 3.17

### x-bar = 3.6

### Now let’s calculate the test statistic for the given values

**Test statistic = (x-bar - mu) / (sigma) – Equation 7**

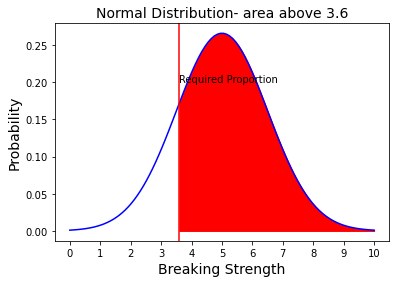
= (3.6- 5)/1.5 = -0.933

Hence, we can calculate the probability using “norm.cdf” function where CDF stands for cumulative distribution function in the python. We can either put the value of test statistic directly into the “(1- norm.cdf)” function or we can put the values of mu, sigma, and x-bar and calculate the probability which comes out as follows:

**Cumulative distributive function (x-bar, location=µ, scale=σ)**

**The probability of the gunny bags having a breaking strength at least 3.6 kg per sq. cm is 0.8247**

Now let’s see that using the graph of probability distribution function and calculate the required area under the probability.

fig-3

In other words, **0.8247** can interpreted as that there are **82.4%** chance that the gunny bags have breaking

strength at least 3.6Kg per sq. cm when randomly chosen from the company. The area under the red colour

shows the required probability in the above normal distribution graph.

**3.3 What proportion of the gunny bags have a breaking strength between 5 and 5.5 kg per sq. cm.?**

The required probability is P (5 < X < 5.5)

For this let’s calculate the corresponding z scores for 5kg and 5.5 kg

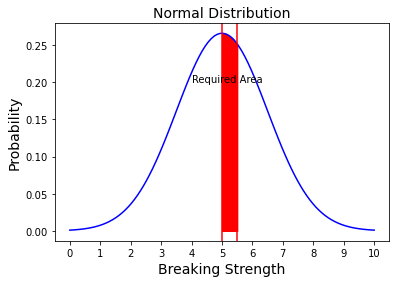
z1= (5-5)/1.55 = 0

z2= (5.5-5)/1.5 = 0.333

Probability between5 and 5.5 kg per sq. cm = cumulative distributive function(z2) – cumulative distributive function(z1) (“using stats. Norm.cdf” in python)

The proportion of the gunny bags have a breaking strength between 5 and 5.5 kg per sq. cm is **0.1306**

Let’s plot this graphically and see the required area under the curve.

fig-4

So, the probability that the breaking strength of bags will fall between 5kg per sq. cm and 5.5 kg per sq. cm is almost equal to 13% that is **13.06%** (when rounded of). This can be calculated using either test statistics or by directly the two values in the norm.cdf function. the difference between two functions will give us the desired probability

**3.4 What proportion of the gunny bags have a breaking strength NOT between 3 and 7.5 kg per sq. cm.?**

Probability here required is = 1 – P (3 <X <7.5)

For this let’s calculate the corresponding z scores for 3kg and 7.5 kg and calculate the probability for P (3 <X <7.5)

z3= (3-5)/1.5 = -1.33

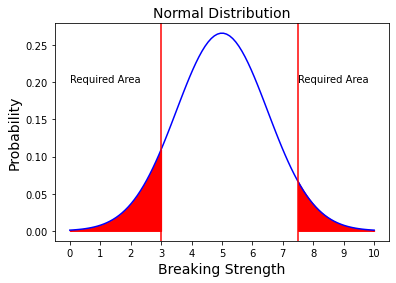
z4 = (7.5 – 5)/1.5 = 1.67

P (Not between 3 and 7.5 kg per sq. cm) = 1 – P (between 3 and 7.5 kg per sq. cm) = 1 – 0.860 = 0.1390

Probability that the breaking strength is below 3 kgs or above 7.5 kgs is **0.1390**

Hence, we can calculate the breaking strength NOT between 3 and 7.5 kg per sq. cm by subtracting the proportion in between these two points from 1, or we can directly add **the area below 3kgs** and **above 7.5 kgs** to get the area which is not between 3 and 7.5 kgs and which is actually the required proportion which in this case comes out to be 0.1390 or we can say **13.9%**

**Now let’s plot the area under the curve for probability below 3 kg per sq. cm and area above 7.5 kg per sq. cm. after adding these areas we will get our desired output for the probability which is not in between 3 to 7.5 kg per sq. cm.**

****fig-5

**So, we can see the area which is not between 3Kgs and 7.5Kgs is filled with red and represents the required breaking strength proportion. This gives us the idea that almost 13.09% of the bags have strength below 3kg per sq. cm and above 7.6kg per sq. cm**

# PROBLEM – 4

# Grades of the final examination in a training course

# Grades of the final examination in a training course are found to be normally distributed, with a mean of 77 and a standard deviation of 8.5. Based on the given information answer the questions below.

# Since cement is normally distributed, so we can say that:

# The mean of the population is (mu1) = 77

# The standard deviation of the population given is (sigma1) = 8.5

# Now if we plot the probability distribution function of the marks scored by the students, we will get the normally distributed curve with mean at 77

# fig-6

# fig-7

# 4.1 What is the probability that a randomly chosen student gets a grade below 85 on this exam?

# The required probability is P (X < 85)

# For the required probability we will calculate the test statistics first and then based on that statistics we will calculate the probability using the “norm.cdf” function in the python. So, lets calculate the test statistic first –

# 

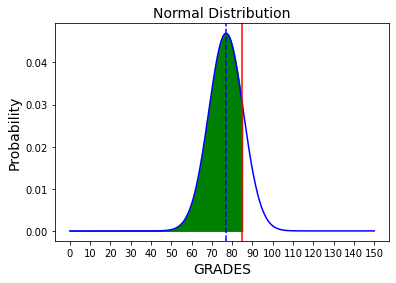
# Test statistic = (x-bar – mu1)/ sigma1

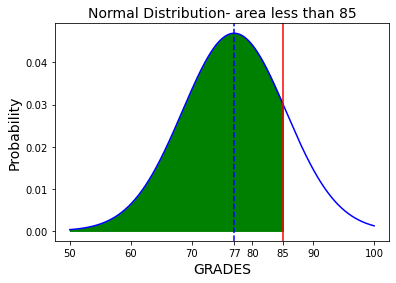
# Z5 = (85-77)/8.5 = 0.94

# Now, using the formula norm.cdf (z5) in python we can calculate P (x < 85)

The probability that a randomly chosen student gets a grade below 85 on this exam is **0.8267**

**We can also plot the probability distribution function of the normally distributed curve using matplotlib, seaborn, and Plotly packages. So, lets plot this graph and find out the area under the curve for the given point of X less than 85**

fig -8

fig-9

# So, the above plot shows the area under the curve for normally distributed grades where 77 is the mean and the line at 85 bifurcates the area required from the total area. hence the green color shows our required probability

# 4.2 What is the probability that a randomly selected student score between 65 and 87?

# The required probability is P (65<X < 87)

# For the required probability we will calculate the test statistics first and then based on that statistics we will calculate the probability using the “norm.cdf” function in the python. So, let’s calculate the test statistic first –

# 

# Test statistic = (x-bar – mu1)/ sigma1

# Z6 = (87-77)/8.5 = 1.18

# Z7 = (65-77)/8.5 = -1.41

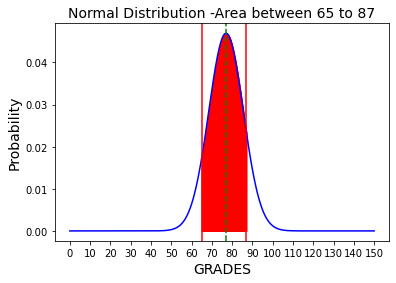
# Now, using the formula norm.cdf (Z6)- norm.cdf (Z7) in python we can calculate P (65< x < 87)

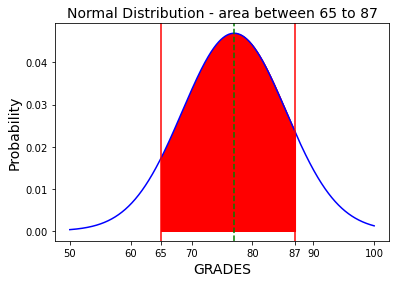
The probability that a randomly chosen student gets a grade between 65 and 87 on this exam is **0.8013**

Hence, we can calculate the probability by calculating the respective z scores and subtracting the values using

norm.cdf function or we can directly put the values of mean and standard deviation in norm.cdf function.

This will give us the required probability which comes out to be 0.8013 in our case, this mean the area between 65 and 87 **marks is almost around 80.13%** which shows that almost 80 percent of students **have got the grades between 65 to 87. Let’s see this graphically as well using normal curve of probability distribution function.**

****fig-10

****fig-11

We can also plot the probability distribution function of the normally distributed curve using matplotlib, seaborn, and Plotly packages. So, lets plot this graph and find out the area under the curve for the grades between 65 and 87.

Hence, we can calculate the probability by calculating the respective z scores and subtracting the values using norm.cdf function or we can directly put the values of mean and standard deviation in norm.cdf function. this will give us the required probability which comes out to be 0.8013 in our case. this mean the area between 65 and 87 marks is **almost around 80.13% which shows that almost 80 percent of students have got the grades between 65 to 87**

# 4.3 What should be the passing cut-off so that 75% of the students clear the exam?

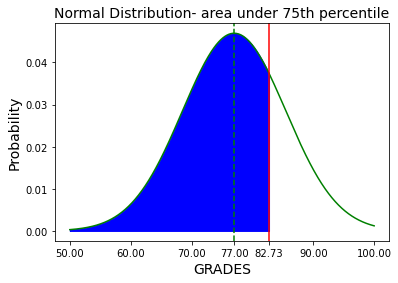
We can calculate this using the PPF (Percent Point Function) function in python

For the score of 75th percentile we will put the point as 0.75 under “norm.ppf” function and this will give us the desired result

The passing cut-off so that 75% of the students clear the exam comes out to be **82.73**

The norm.ppf () function is the inverse of the norm.cdf () function. It takes a percentage p and returns a point such that the probability of the normal random variable being less than or equal to that number is p%. Thus, it just does the opposite work of norm.cdf ().

For example, if the percentage p is equal to 0.92, you will get the point below which 92% of data falls. This also means that 8% of data falls above that point. Let’s plot this graphically and see the area which falls under this point which actually turns out to be **75%**

fig-12

# PROBLEM – 5

# Zingaro stone printing is a company and its analysis

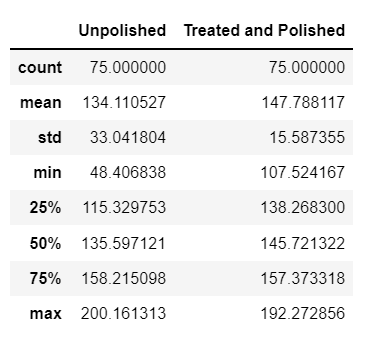
# Zingaro stone printing is a company that specializes in printing images or patterns on polished or unpolished stones. However, for the optimum level of printing of the image the stone surface has to have a Brinell's hardness index of at least 150. Recently, Zingaro has received a batch of polished and unpolished stones from its clients.

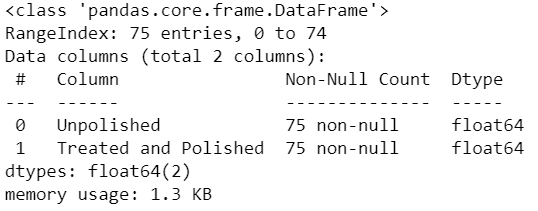
# Now lets first go through the dataset and analyze the sample and its variables. After completing the exploratory data analysis, it will be easy for us to do the hypothesis testing. We will check the variables their datatypes, their null and duplicate values, the outliers, the summary of the data and analyze the boxplots of the continuous variables and then finally after treatment of the data we will further move towards the hypothesis testing.

# Going through the dataset – Reading the top 8 rows

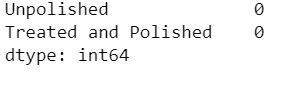
# Table-2-Zingaro Company Data Set

Summary & Information of the data set – before treatment

Table-3, Summary of Zingaro- before treatment

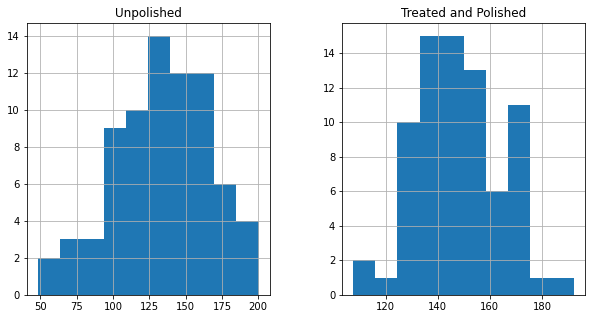
Table -4, Info

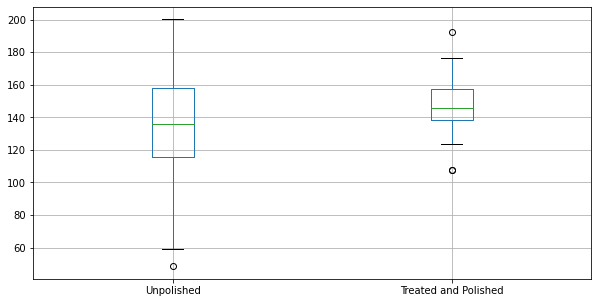
Now let’s check the null and duplicate values of the given sample. If null values or duplicate data are present in the sample data set, we will treat them and if not, we will further proceed with the outlier’s detection and their treatment.

Table-5, null value counts

Also, the duplicate value count in the data set is also zero. So, the next step for us to do is checking the outliers in the given variables of “Unpolished” and “Treated and polished”. This will be done using the analysis of histograms and boxplots of the variables.

**Plotting the histogram and boxplots of “UNPOLISHED AND TREATED & POLISHED”**

fig-13 & 14

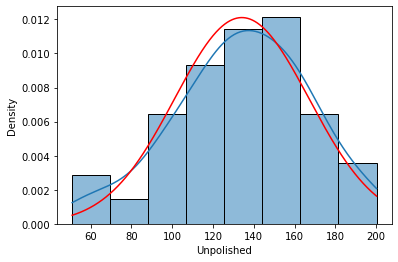
fig-15

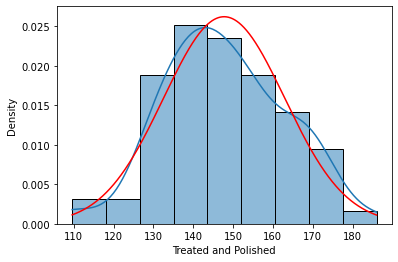
We can see the outliers are present in both the cases so it will be good to treat these outliers first and then move ahead, because outliers are something which can affect the central tendency like mean & median, and mean being the critical points in analysis of various kinds of hypothesis testing, if not treated it may deflect the actual central tendencies and hence the test results.

**5.1 Earlier experience of Zingaro with this particular client is favourable as the stone surface was found to be of adequate hardness. However, Zingaro has reason to believe now that the unpolished stones may not be suitable for printing. Do you think Zingaro is justified in thinking so?**

Plotting the Distribution

It will help us analyse the shape of data distribution and PDF of normal distribution using the calculated statistics (mu and sigma) from data and then we will be able to do further calculations.

 fig-16

 fig-17

The probability distribution function seems to be almost normal hence we can perform the required T-test in this case

**ONE SAMPLED T-TEST**

Step 1: Given: n = 75, σ = not known (variance unknown)

Step 2: Let us formulate the hypothesis.

* H0 (Null hypothesis): Average hardness of unpolished stone <= 150
* μ (unpolished stone) <= 150, which means the unpolished stones are suitable for printing
* Ha (alternate hypothesis): Average hardness of unpolished stone > 150
* μ (unpolished stone) > 150, which means the unpolished stones are not suitable for printing

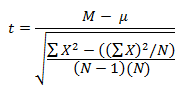
[this is an example of a One-tailed test]

Step 3: Define the test statistic based on the information in the question. Here, we are going to use the T-stat, since the population variances not given

Step 4:

* Identify the test statistic
* We have two samples and we do not know the population standard deviation.
* Sample sizes for both samples are same.
* Although The sample size is large sample, n =75. So, you use the t distribution and the test statistic for one sample (unpolished stone) T- test.

Let us calculate the value of the test statistic.

Equation-8

One sample t test

t statistic: -4.166875533846615 & p value: 0.999958618136

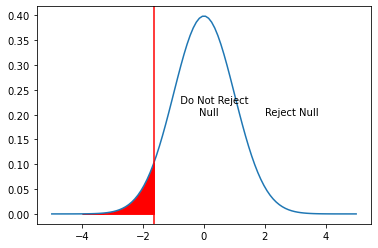
Level of significance(alpha): 0.05

The *t*-value is -4.16463. The value of p value: **4.138186355162632e**-05. The result is significant at *p* < .05.

We have no evidence to reject the null hypothesis since p value > Level of significance(alpha)

We conclude that the unpolished stones may be suitable for printing. Our one-sample t-test p-value= [4.138186355162632e-05]

Zingaro believes that the unpolished stones may not be suitable for printing and their claim seems to be justified based on the values as p value= **4.138186355162e-05** is less than alpha (0.5) and the result is significant**.**

 fig-18

**Insight**

**As our test statistic (~ -4.166875533846615) does lie in the rejection region, we can reject the null hypothesis. Thus, we do have statistical evidence to say that unpolished stones are not suitable for printing**

**5.2 Is the mean hardness of the polished and unpolished stones the same?**

**TWO SAMPLED T-TEST –**

Step 1: Given: n = 75, σ = not known (variance unknown)

Step 2: Let us formulate the hypothesis.

* H0 (null hypothesis): Average hardness of unpolished stone is equals to Average hardness of polished stone.
* μ (unpolished stone) = μ (polished stone)
* Ha (alternate hypothesis): Average hardness of unpolished stone is not equal to Average hardness of polished stone
* μ (unpolished stone) ≠ μ (polished stone)

This is an example of a Two-tailed test

In testing whether the mean hardness of stones is same in both of "unpolished" and "treated and polished stones", the null hypothesis states that the mean hardness of the stones are the same, 𝜇p equals 𝜇u. The alternative hypothesis states that the mean hardness of the stones is different, 𝜇p is not equal to 𝜇u.

𝐻0: 𝜇p - 𝜇u = 0 i.e. 𝜇p = 𝜇u

𝐻𝐴: 𝜇p - 𝜇u ≠ 0 i.e. 𝜇p ≠ 𝜇u

Step 3:

* Identify the test statistic
* We have two samples and we do not know the population standard deviation.
* Sample sizes for both samples are same.
* Although The sample size is large sample, n > 30. So, you use the t distribution and the 𝑡-𝑆𝑇𝐴𝑇 test statistic for two sample unpaired test.

Step 4: Calculate the p - value and test statistic

Difference Scores Calculations

Treatment 1

N1: 75

df1 = N - 1 = 75 - 1 = 74 (degree of freedom)

M1: 134.11

SS1: 80790.3

s21 = SS1/ (N - 1) = 80790.3/ (75-1) = 1091.76

Treatment 2

N2: 75

df2 = N - 1 = 75 - 1 = 74

M2: 147.79

SS2: 17979.46

s22 = SS2/ (N - 1) = 17979.46/ (75-1) = 242.97

For T-value Calculation we can use the formula

s2p = ((df1/ (df1 + df2)) \* s21) + ((df2/ (df2 + df2)) \* s22) = ((74/148) \* 1091.76) + ((74/148) \* 242.97) = 667.36

s2M1 = s2p/N1 = 667.36/75 = 8.9

s2M2 = s2p/N2 = 667.36/75 = 8.9

t = (M1 - M2)/√ (s2M1 + s2M2) = -13.68/√17.8 = -3.24 – Equation-9

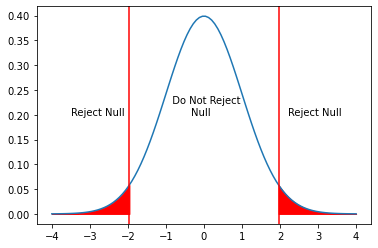
t-stat = -3.2483931503833814

P Value = 0.001436230005012731

Step 5: Decide to reject or accept null hypothesis –

* Two-sample t-test p-value= 0.001436230005012731
* We have enough evidence to reject the null hypothesis in favour of alternative hypothesis
* We conclude that the Average hardness of unpolished stone is not equals to Average hardness of polished stone.

**Hence, we conclude that the hardness of polished and unpolished stones is not equal and rather the hardness of polished stone is greater than that of unpolished. The figure below shows our approach.**

****fig-19

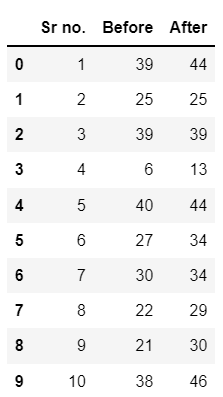
# PROBLEM – 6

# Aquarius gym and their new rigorous program –

# Aquarius health club, one of the largest and most popular cross-fit gyms in the country has been advertising a rigorous program for body conditioning. The program is considered successful if the candidate is able to do more than 5 push-ups, as compared to when he/she enrolled in the program.

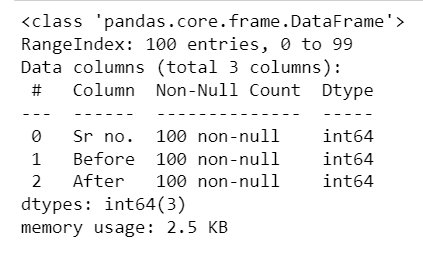
**6.1 Using the sample data provided can you conclude whether the program is successful?**

Loading the dataset and getting to know the data

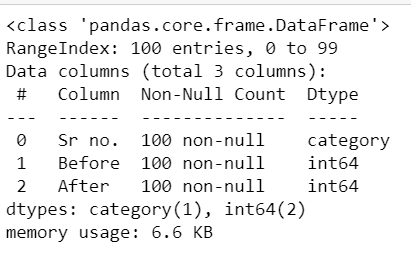
Table-6-Aquarius Health club

# Description of the data, information and summary is as follows:

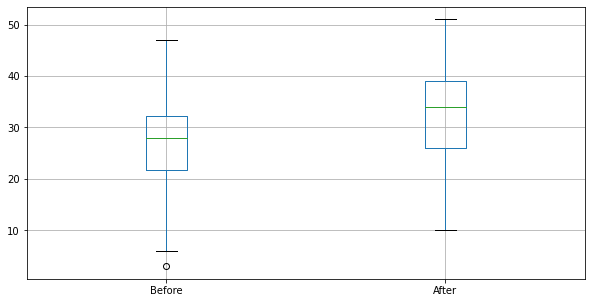
# Table-7

Table-8

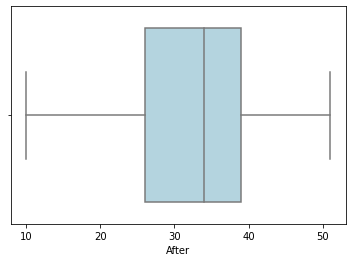
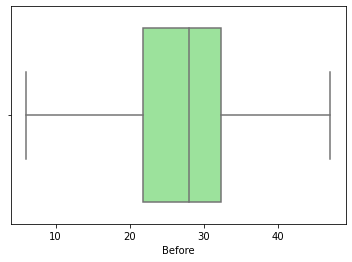
After changing the categorical variable-

 Table-9

Since the null values are also equals to 0, now, we will plot the boxplots and treat the outliers if any.

fig-20

After the treatment of outliers our variables are shown in the figure below-

 fig-21fig-22

* Mean and Median Values of each sample are not much different.
* Both "Before" data and "after" data looks left skewed and has skewness as -0.219 and -0.3189.

Now we will Describe the five per cent significance test to these data to determine whether rigorous program for body conditioning of Aquarius gym has significantly raised outputs

* The level of significance (Alpha) = 0.05.
* But since the population standard deviation (Sigma) is unknown, we have to use a paired t-Stat test.
* Degree of Freedom: Since the sample is the same for both Sampling tests, we have N-1 degrees of freedom: 100-1 =99

The t-test for dependent means (also called a repeated-measures t-test, paired samples t-test, matched pairs t-test and matched samples t-test) is used to compare the means of two sets of scores that are directly related to each other.

* Let’s perform the paired t- test since it is the case where the samples are dependent on each other
* The steps we will follow are as follows
* 1. Calculate the difference (difference (d) = BEFORE - AFTER) between the two observations on each pair, such that we can distinguish between positive and negative differences.
* 2. then we will calculate the mean difference, d.
* 3. Calculate the standard deviation of the differences, standard deviation(sd), and use this to calculate the standard error of the mean difference

**SE(d) = sd/√n -**Equation 10

* 4. after this we will Calculate the t-statistic, which is given by T =d/SE(d), Under the null hypothesis, this statistic follows a t-distribution with n − 1 degrees of freedom.
* 5.Then take out the p value for the given test stat value. This will give the p-value for the paired t-test.

Since the sole purpose of the test is to check whether the rigorous program for body conditioning, if the candidate is able to do more than 5 push-ups, as compared to when he/she enrolled in the program is successful compared to before, we would prefer a One-sided paired T-test.

* Hypothesis Formulation - here in case of we will use Paired two Sample for Means:
* Null Hypothesis**:** Ho = 𝜇after−𝜇before <= 5 (After rigorous program for body conditioning did not raise the output significantly)
* Alternate Hypothesis: Ha = 𝜇after−𝜇before > 5 (After rigorous program for body conditioning, raised the output significantly)

Now for calculating the t-stat and the standard error we will use the mathematical equations

### 𝑡𝑠𝑡𝑎𝑡=(𝑑bar−𝜇𝐷) 𝑆𝑑/𝑛√𝑡𝑠𝑡𝑎𝑡=(d¯−μD) 𝑆\_𝑑/𝑛 -Equation 11

where 𝑆\_𝑑/sqrt(n) is the standard - error

Then we calculate, T =d/SE(d), using this formula, the t- statistic for the given value

Sd = 2.87265055517047

Standard Error = 0.28726505551704

t-stat = 1.8101749238662324

t-critical value = 1.6603911559963902

We fail to reject the null hypothesis if we get a T-statistic less than 1.66039 for the sample size of 100. so basically, it says

Accept it if: Absolute value of T-score < critical value t

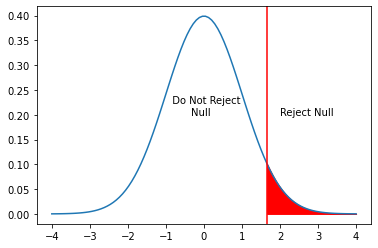
Reject it if: Absolute value of T-score > critical value t

The p-value is .030405 (according to the difference method approach).

The result is significant at p < .05. Hence, we will reject the Null Hypothesis and this states that the random sample of 10-program participants given.

**Paired two-sample t-test p-value= 0.03513433250317666**

**We have enough evidence to reject the null hypothesis in favour of alternative hypothesis.**

**** fig-23

Insight

As our test statistic **(1.810174923866232)** lies in the rejection region, reject the null hypothesis. Thus, have enough statistical evidence to conclude that the claim of the gym is true and after the program its evident that participants are doing more than 5 push-ups.

**Hence the value is significant with respect to alpha = 5% or at 95% confidence interval, we reject null hypothesis and this means that the claim of Aquarius gym is true and their rigorous program for body conditioning is successful and bought a significant change and after the program people are doing 5 push-ups more.**

# PROBLEM – 7

# Dental Implant Data set and the Response –

# Dental implant data: The hardness of metal implant in dental cavities depends on multiple factors, such as the method of implant, the temperature at which the metal is treated, the alloy used as well as on the dentists who may favor one method above another and may work better in his/her favorite method. The response is the variable of interest.

# Going through the dataset and checking and treating

# Table-10

# Table-11

# Table-12 Table-13

# Changing the integer variables data types into categorical data types we obtain the following table and then we plot the Boxplot of the Response

# Table-14

# fig-24

# Here we can see that outliers are present in the dataset but for this case we are ignoring the outliers and then doing the hypothesis testing.

# 7.1 Test whether there is any difference among the dentists on the implant hardness. State the null and alternative hypotheses. Note that both types of alloys cannot be considered together. You must state the null and alternative hypotheses separately for the two types of alloys?

# The Hypothesis for the One-way ANOVA & Two-way ANOVA are:

# For variable A(Dentist) & alloy =1

# There is no difference in the means of variable A categories

# 𝐻0: The means of response (implant Hardness) is the same for the 5 different categories of dentists.

# μ1 = μ2 = μ3 = μ4 = μ5 = μa (for alloy 1)

# 𝐻𝑎: For at least one pair of Dentist, mean number of RESPONSE (implant HARDNESS) is different

# For variable A(Dentist) & alloy =2

# There is no difference in the means of variable A categories

# 𝐻0: The means of response (implant Hardness) is the same for the 5 different categories of dentists.

# μ1 = μ2 = μ3 = μ4 = μ5 = μb (for alloy 2)

# 𝐻𝑎: For at least one pair of Dentist, mean number of RESPONSE (implant HARDNESS) is different

# For variable A and B interaction, that is Both Dentist and Alloy

# 𝐻0: Interactions are not significant means there is no interaction between variable A and variable B

# μ(AiBj) (for alloy 1&2)

# (AiBj) = 0 (such as i = 1 to a, j = 1 to b)

# 𝐻𝑎: Interactions are significant means There is significant interaction between variable A and variable B

# Table-15

# Table-15

# Here we are first checking difference among the dentists on the implant hardness without differentiating the alloys, which comes out to be as p value > 0.05(alpha), which means we 𝐟𝐚𝐢𝐥 𝐭𝐨 𝐫𝐞𝐣𝐞𝐜𝐭 the 𝐍𝐮𝐥𝐥 𝐇𝐲𝐩𝐨𝐭𝐡𝐞𝐬𝐢𝐬 (𝐻0) and hence the mean implant hardness of all the dentists are equal. But now we will consider the alloy 1 and alloy 2 differently and then again check the implant hardness. Now we will perform the analysis on two different Alloys separately.

# For Alloy -1

# Table-16

# Table-17

# Fig-26

# Fig-25

# Now, we see that the corresponding p-value is GREATER (0.116567) than alpha (0.05) in all the cases. Thus, we fail to reject the Null Hypothesis (H0).

# Hence, we can say that means of implant hardness is same in case of 5 categories of Dentists When we take the Alloy 1. Also, we see that even the interactions categorical variables "Dentists" and "Alloys" are not much significant also since the value of p in case of interaction is also greater than alpha.

# For Alloy -2

# Table-18

# Table-19

# fig-27 fig-28

# In case of Alloy -2 as well we fail to reject the null hypothesis as the value of p (0.718031) is > than alpha in case of categories Dentists or even in case of interaction of alloy2 and dentists, the interactions are not much significant.

# 7.2 Before the hypotheses may be tested, state the required assumptions. Are the assumptions fulfilled? Comment separately on both alloy types?

# Assumptions of ANOVA

# 1. Continuity of Response, that means the dependent variable should be continuous in nature.

# 2. Samples drawn from different population should be randomized

# 3. Independent variables should each consist of 2 or more categorical, independent groups

# 4. There should be no significant outliers

# 5. Dependent variables should be approximately normally distributed for each combination of the groups of two independent variables

# 6. Normally distributed and Equal variance

# Let’s perform The Shapiro Test for “Normality”, Anderson test, The Levene for “Homogeneity” for checking the null hypothesis that all input samples are from populations with equal variances.

# Table-20

# 

# So, we can say that assumptions of hypotheses are "not" fulfilled completely in the case of ALLOY 2 where p- value is < 0.05 as the variable is not approximately normally distributed here, which have obtained from the Shapiro and Anderson test.

# We are accepting the null hypothesis in case of ALLOY 1 as The Shapiro-Wilk test tests the null hypothesis that the data was drawn from a normal distribution. here p value is > alpha, so we accept the hypothesis and hence the assumption holds true here

# Also, we can see significant no of outliers present in case of Response as well which is again opposite that of the assumptions required in case of ANOVA testing.

# 7.3 Irrespective of your conclusion in 7.2, we will continue with the testing procedure. What do you conclude regarding whether implant hardness depends on dentists? Clearly state your conclusion. If the null hypothesis is rejected, is it possible to identify which pairs of dentists differ?

# The Hypothesis for ANOVA are:

# For variable A(Dentist)

# There is no difference in the means of variable A categories

# H0: The means of response (implant Hardness) is the same for the 5 different categories of dentists.

# μ1 = μ2 = μ3 = μ4 = μ5 = μa

# Ha: For at least one pair of Dentist, mean number of RESPONSE (implant HARDNESS) is different

# Table-21

# For Alloy -1

# Table-22

# For Alloy -2

# Table-23

# fig-29

# Since in our p-value is greater than 0.05 (alpha), hence we Fail to Reject the Null Hypothesis. This concludes that the means of implant hardness of dentists does not differ in any case. It’s almost same for all the categories of Dentists in case of Alloy -1 and in case of Alloy-2.

# Fig-30 Fig-31

# Irrespective of our conclusion in 7.2 that implant hardness does not depend on Dentists. But if we reject the null hypothesis, we could say that at least one pair of mean is different for different categories of dentists that is the case when alternate hypothesis will be true.

# But we could say possibly which pairs of dentists differ here in this case, for that we have to do further analysis and draw point plots and Box plots to check where the mean and median is highest in case of each alloy separately and boxplot of dentists as a whole without differentiating the alloys.

# So, it shows the mean of RESPONSE of Dentist 2 is highest in case of Alloy-1 and it’s highest for Dentist 1 in case of Alloy-2

# Also, in case of boxplot as a whole it shows that the median value for Dentist 1 is slightly higher than in Dentist 2

# 7.4 Now test whether there is any difference among the methods on the hardness of dental implant, separately for the two types of alloys. What are your conclusions? If the null hypothesis is rejected, is it possible to identify which pairs of methods differ?

# The Hypothesis for ANOVA are:

# For variable A(Method) & alloy=1

# There is no difference in the means of variable A categories

# 𝐻0: The means of response (implant Hardness) is the same for the 3 different categories of Methods.

# μ1 = μ2 = μ3 = μ4 = μ5 = μa (for alloy 1)

# 𝐻𝑎: For at least one pair of Method, mean number of RESPONSE (implant HARDNESS) is different

# For variable A(Method) & alloy=2

# There is no difference in the means of variable A categories

# 𝐻0: The means of response (implant Hardness) is the same for the 3 different categories of Methods.

# μ1 = μ2 = μ3 = μ4 = μ5 = μb (for alloy 2)

# 𝐻𝑎: For at least one pair of Method, mean number of RESPONSE (implant HARDNESS) is different

# For variable A and B interaction, that is Both Method and Alloy

# 𝐻0: Interactions are not significant means There is no interaction between variable A and variable B

# μ(AiBj) (for alloy 1&2)

# (AiBj) = 0 (such as i = 1 to a, j = 1 to b)

# 𝐻𝑎: Interactions are significant means There is significant interaction between variable A and variable B

# Table-24

# Since the p-value < α, H0 is rejected. Some of the groups' averages consider to be not equal. In other words, the difference between the averages of some groups is big enough to be statistically significant.

# For Alloy-1

# Table-25

# fig-32

# Here the p-value (0.00416) is less than alpha (0.005), hence we reject the null hypothesis and can consider that the means of Responses are different for at least one pair of the Methods out of the 3 different methods in case of Alloy-1

# For Alloy-2

# Table-26

# fig-33

# Here the p-value (0.000005) is less than alpha (0.005), hence we reject the null hypothesis and can consider that the means of Responses are different for at least one pair of the Methods out of the 3 different methods for Alloy-2

### Interaction hypothesis:

### Table-27

### Null hypothesis is rejected in case of “Method” category which means that means of Responses that is implant hardness differs for three different methods. It differs in case of alloy-1, alloy-2 and overall, also when alloys are not differentiated.

# 7.5 Now test whether there is any difference among the temperature levels on the hardness of dental implant, separately for the two types of alloys. What are your conclusions? If the null hypothesis is rejected, is it possible to identify which levels of temperatures differ?

## **For Alloy-1**

## Table-28

### P- value is greater than alpha, so we fail to reject the null hypothesis.

## **For Alloy-2**

## Table-29

### P- value greater than alpha so we fail to reject the null hypothesis which means the means of Responses that is the implant hardness are same in case all the three categories of temperatures and hence temperature doesn’t have any impact on implant hardness.

### Table-30

### Together Alloy and temperature doesn’t show much interaction but in case of alloy alone p value is less than alpha (0.05) so we can reject the null which means different alloys has different means for Implant hardness

# 7.6 Consider the interaction effect of dentist and method and comment on the interaction plot, separately for the two types of alloys?

# Fig-34

# For Alloy -1

### Table-31

### Fig-35

# Still, we can see that there is some sort of interaction between the Dentists and Methods specially in case of Dentist 3 and the 3 Methods. So, we will introduce a new term while performing the Two Way ANOVA.

# For Alloy-2

# Table-32

# Fig-36

# Here, we can see that there is not much of interaction between the Dentists and Methods as in case of Alloy-1, since, the value of p (0.09) for interaction part is greater than alpha (0.05). So, we will reject the null hypothesis which says that there.

# 7.7 Now consider the effect of both factors, dentist, and method, separately on each alloy. What do you conclude? Is it possible to identify which dentists are different, which methods are different, and which interaction levels are different?

# Here in case of both Alloy-1 From the Table in 7.6 we can say that

# Since the p-value < α, H0 can be rejected. The averages of all groups cannot assume to be equal. In other words, the difference between the averages of all groups is big enough to be statistically significant. A significant result cannot prove that H0 is not correct, only that the null assumption can be rejected.

# Test statistic

# The test statistic FA equals 3.3983, which is almost in the 95% region of acceptance: [-∞: 3.8056].

# Here in case of both Alloy-2 From the Table in 7.6 we can say that

# Since the p-value > α, H0 cannot be rejected. The averages of all groups can be assumed to be equal. In other words, the difference between the averages of all groups is not big enough to be statistically significant. A non-significant result cannot prove that H0 is correct, only that the null assumption can be rejected.

# Test statistic

# The test statistic FA equals 1.9227, which is almost in the 95% region of acceptance: [-∞: 3.8056].

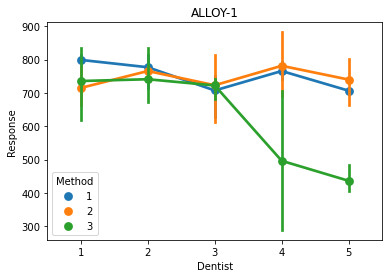
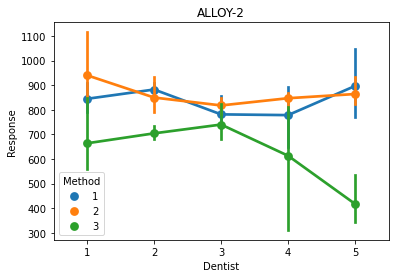
# Now let’s conclude based on the boxplots – of Dentist and Methods Together for Overall Alloy 1 and Alloy 2 and then separately for both the Alloys.

# Fig-37 Fig-38

# Fig-39

# Fig-40

# 

  Fig-41

All these figures in Fig - 41 shows that:

* From answer in 7.2 and its Graph we can consider that mean response of Dentist 2 differs mostly from the other means.
* From answer in 7.4 and its Graph we can consider that mean response of Method 3 differs mostly from the other Method means. Means of Method 1 and 2 are almost similar
* For Dentist 4 & significantly for Dentist 5, Method 3 differs from that of Method 1 & 2 in case of Alloy 1.
* For Dentist 4, Dentist 1 & significantly for Dentist 5, Method 3 differs from Method 2 and a bit for Method 1

*\*\*\*Reference – Great Learning lecture videos and Mentors*