# List of Figures

Figure 1 : Zingaro dataset 4

Figure 2 : Zingaro data information 5

Figure 3 : Zingaro null values 5

Figure 4 : Histogram distribution of polished and unpolished stones 6

Figure 5 : Boxplot distribution of polished and unpolished stones 6

Figure 6 : Aquarius data set 6

Figure 7 : Aquarius data information 7

Figure 8 :Aquarius null values 7

Figure 9 : Histogram distribution before and after enrolling in the program 7

Figure 10 : Boxplot distribution before and after enrolling in the program 8

Figure 11 : Data set with difference column 8

Figure 12 : Dental implant dataset 9

Figure 13 : Dental implant data information 10

Figure 14 : Dental implant null values 10

Figure 15 : Distribution of alloy1 and alloy2 11

Figure 16 : Response distribution of alloy1 with respect to dentists 11

Figure 17 : Response distribution of alloy2 with respect to dentists 12

Figure 18 :Dentist Vs Response for alloy1 and alloy2 12

Figure 19 : ANOVA table for alloy1 dentists 13

Figure 20 : ANOVA table for alloy2 dentists 13

Figure 21 : Visualization for EducationResponse distribution of alloy1 with respect to methods 14

Figure 22 : Response distribution of alloy2 with respect to methods 14

Figure 23 : Response Vs Method for alloy1 and alloy2 15

Figure 24 : ANOVA table for alloy1 methods 16

Figure 25 : Tucky-HSD table for alloy1 16

Figure 26 : ANOVA table for alloy2 methods 17

Figure 27 : Tucky-HSD table for alloy2 18

Figure 28 : Response distribution of alloy1 with respect to Temperature 19

Figure 29 : Response distribution of alloy2 with respect to Temperature 19

Figure 30 : Temp Vs Response for alloy1 and alloy2 20

Figure 31 : ANOVA table for alloy1 temperatures 21

Figure 32 : ANOVA table for alloy2 temperatures 21

#### **PROBLEM 1:**

**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. Use the data provided to answer the following (assuming a 5% significance level).**

**1.1 Earlier experience of Zingaro with this particular client is favorable 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?**

**Problem Definition:**

* A Zingaro Stone is a printing company that specializes in printing images or patterns on polished or unpolished stones, the optimum level of printing of the image on the stone surface has to have a Brinell’s hardness index of at least 150. Now, they have received a batch of polished and unpolished stones from their clients.
* The significance level is set as 5%(alpha)
* As per earlier experience Zingaro believes that polished stones have adequate hardness compared to unpolished stones.
* Now we are supposed to find out whether Zingaro is justified as per the earlier experience.

**Dataset:**

* As per the dataset we have two columns **“Unpolished”** and **“Treated and Polished”**
* **Unpolished:** A batch of unpolished stones from its clients for printing
* **Treated and Polished:** A batch of treated and polished stones from its clients for printing.

A screenshot of a computer

Description automatically generated

Figure 1 : Zingaro dataset

* The dataset has 75 rows and 2 columns of float data type.

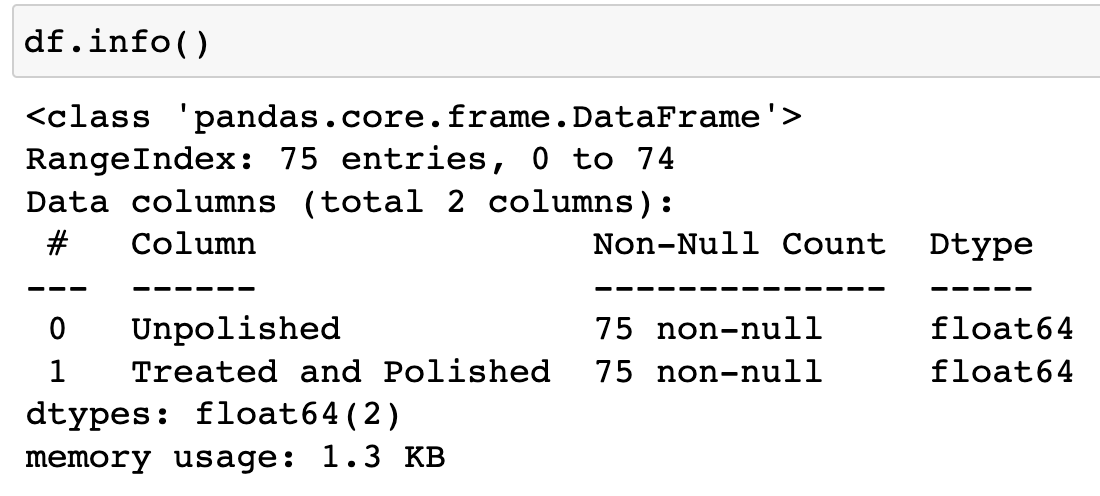


Figure 2: Zingaro Data information

* There are no null values present in the data.

A screenshot of a computer code

Description automatically generated

Figure 3: Zingaro Null values

**Distribution of dataset:**

**A comparison of a graph

Description automatically generated**

Figure 4: Histogram distribution of polished and unpolished stones

**A comparison of a graph

Description automatically generated**

Figure 5: Boxplot distribution of polished and unpolished stones

**Hypothesis Testing Assumptions:**

* **Null Hypothesis(H0):** Assuming the mean Brinell’s hardness index of the unpolished stones is equal to or greater than 150 **(mu>=150)**
* **Alternate Hypothesis(H1):** The mean Brinell’s hardness index of the unpolished stones is less than 150 **(mu<150)**

**Steps to be considered in Hypothesis Testing:**

* Calculate the t-statistic and p-value for Treated and Polished Stones
* Calculate the t-statistic and p-value for Unpolished
* **T\_statistic** for Unpolished Stone is **-4.164**
* **P-value** for Unpolished Stone is **4.171\*10-05**

**Result:**

* For Unpolished, P-value(4.171e-05**)** is less than alpha (0.05) ,hence we reject the null hypothesis
* So, the Zingaro Company is right that the unpolished stones are not suitable for printing.

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

**Problem Definition:**

We will use the existing dataset to check the whether the mean hardness of the polished and unpolished stones are same using **“independent sample t-test”**

**Hypothesis Testing Assumptions:**

* **Null Hypothesis(H0):** The Mean hardness of the polished stones **(mu1)** is equal to the

Mean hardness of the unpolished stones **(mu2)**

* **Alternate Hypothesis(H1):** The Mean hardness of the polished stones **(mu1)** is not equal to the Mean hardness of the unpolished stones (**mu2)**

**Steps to be considered in Hypothesis Testing:**

* Calculate t-statistic and p-value

**T\_statistic** value **is -3.2422**

**P-Value** is **0.001465**

**Result:**

* P value (**0.001465**) is less than alpha(0.05). Hence we reject the null hypothesis
* So both means are not the same at 90% confidence interval

#### **PROBLEM 2:**

**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. Using the sample data provided can you conclude whether the program is successful? (Consider the level of Significance as 5%)**

**Note that this is a problem of the paired-t-test. Since the claim is that the training will make a difference of more than 5, the null and alternative hypotheses must be formed accordingly.**

**Problem Definition:**

Aquarius health club has a program for body conditioning, and success is defined as being able to do more than 5 push-ups compared to when the candidate enrolled in the program. We need to determine if the program is successful based on the provided sample data. The significance level is set at 5%. Now, it is tasked to find out whether the program is a success or a failure.

**Dataset:**

* The data set consists of Two columns “Before” and “After”.
* Before: The number of Pushups a candidate makes before enrolling in the program.
* After: The number of pushups a candidate does after the program.

A screenshot of a number

Description automatically generated

Figure 6: Aquarius dataset

* Data set contains 100 rows and 3 columns.

A screenshot of a computer code

Description automatically generated

Figure 7: Aquarius data information

* There are no null values in the data set.

A black and white text

Description automatically generated

Figure 8: Aquarius null values

**Assumptions of T-Test:**

1. Each observation should be independent of the other observation.
2. The Difference between the pair is normally distributed.
3. There should be no extreme outlier with the Difference.

A comparison of a diagram

Description automatically generated

Figure 9: Histogram distribution before and after enrolling in the program

A two lines with text on them

Description automatically generated

Figure 10: Boxplot distribution before and after enrolling in the program

For this approach, we shall check the validity of the claim of the program's success with Paired T-Test.

**Hypothesis Definition:**

**Null Hypothesis (H0):** Mean of push-ups before program is equal to mean of push-ups after enrolling into the program.

**Alternate Hypothesis (H1):** Mean of push-ups before program is not equal to mean of push-ups after enrolling into the program.

**Paired T-Test:**

Paired T-Test is conducted with mean 1 (as before and mean 2 ( as after. We get the P value and T statistic as below:

**T- Statistic** is **19.3226**

**P – Value** is **2.292\*10-35**

**RESULT:**

* P value (**2.292\*10-35**) is less than alpha(0.05). Hence we reject the null hypothesis
* So there is some difference before and after the program.
* So the program is effective.

**Verification of Push – Up increase rate of 5.**

For this part of the verification, we use the column “Difference” and perform 1 Sample T-test and verify whether the push-up rate has really gone up more than 5.

A screenshot of a number grid

Description automatically generated

Figure 11: Data set with difference column

**Hypothesis Definition:**

**Null Hypothesis (H0):** The Push-up rates have not gone up and remain less than or equal to 5 ( =< 5)

**Alternate Hypothesis (H1):** The Push-up rate has gone up After the program and is more than 5 ( > 5)

**1 Sample T-test:**

We'll use a one-sample t-test to compare the sample mean increase in the number of push-ups to 5.

1 – Sample T Test is conducted to validate whether the push-up rates After the program has gone up more than 5 numbers or not.

**T – Statistic** is **1.9148**

**P – Value** is **0.02919**

Result:

* P – value is((0.029188 ) lower than alpha(0.05). Hence we reject the Null Hypothesis.
* We can conclude that the Program is successful and the pushup rates after the program have increased more than 5 numbers.

#### **PROBLEM 3:**

**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 favour one method above another and may work better in his/her favourite method. The response is the variable of interest.**

1. **Is there any difference among the dentists on the implant hardness?**
2. **Is there any difference among the methods on the implant hardness?**
3. **Is there any difference among the temperature levels on the implant hardness?**

# **DATASET**:

* The dataset consists of 5 different columns.
* Dentist, Method, Alloy, Temp,Response

A table of numbers and a number

Description automatically generated

Figure 12: Dental implant dataset

* The data consists of 90 rows and 5 columns.

A screenshot of a computer program

Description automatically generated

Figure 13: Dental implant data information

* The implant hardness depends on Dentist, Method, Alloy, and Time.
* The dataset has 5 types of dentists, 3 different types of Methods and 2 types of alloys.
* There are no null values present.

A white background with black text

Description automatically generated

Figure 14: Dental implant null values

The question demands us to conduct tests separately for two different types of alloy that are present. So, we begin the analysis by first splitting the data as df\_1 for alloy1 and df\_2 for alloy\_2.

**Shapriro-Wilk test:**

The analysis is begun by conducting a Shapiro-Wilk test which is performed to test if the given data is normally distributed.

**Hypothesis Definition:**

**Null Hypothesis (H0):** Data is normally distributed.

**Alternate Hypothesis (H1):** Data is not normally distributed.

For alloy1: P\_value =1.1945e-05

For alloy2: P\_value =0.0004029

Both the values are less than 0.05, thus we reject null hypothesis. i.e-alloy1 and alloy2 are not normally distributed.

A comparison of a graph

Description automatically generated

Figure 15: Distribution of alloy1 and alloy2

**The Levene test:**

The Levene test tests the null hypothesis that all input samples are from populations with equal variances

**Null Hypothesis (H0):** - The distribution has equal variances

**Alternate Hypothesis (H1):** - The distribution does not have equal variances

* For ‘Method’: There are 3 different types of methods in both alloys.

Alloy1: P\_value =0.00341

Alloy2: P\_value =0.0446

Both the values are less than 0.05, thus we reject null hypothesis. i.e- One of the methods in alloy1 and alloy2 have different variance.

* For ‘Dentist’: There are 5 different types of dentists in both alloys.

Alloy1: P\_value =0.25655

Alloy2: P\_value =0. 23686

Both the values are greater than 0.05, thus we fail to reject null hypothesis. i.e- All variances of dentists in alloy1 and alloy2 are same.

* For ‘Temp’: There are 3 different types of temperatured in both alloys.

Alloy1: P\_value =0.76869

Alloy2: P\_value =0.51719

Both the values are greater than 0.05, thus we fail to reject null hypothesis. i.e- All variances of temps in alloy1 and alloy2 are same.

* 1. **Is there any difference among the dentists on the implant hardness?**

The answer begins by plotting the distribution of alloy1 and alloy2 with respect to dentists.

A graph of a distribution of response

Description automatically generated

Figure 16: Response distribution of alloy1 with respect to dentists

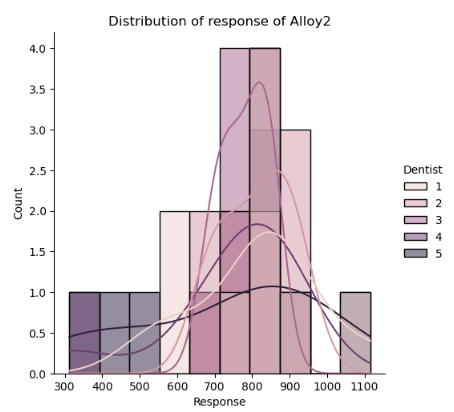


Figure 17: Response distribution of alloy2 with respect to dentists

A graph of different colored squares

Description automatically generated

Figure 18: Dentist Vs Response for alloy1 and alloy2

Figure 15: Response distribution of alloy1 with respect to dentists

The boxplot of Dentist vs Response shows the presence of a few outliers.

**ALLOY1:**

**Hypothesis Definition:**

**Null Hypothesis (H0):** The implant hardness of alloy1 is the same for all dentists.

**Alternate Hypothesis (H1):** The implant hardness of alloy1 is not the same for all dentists.

ANOVA test is conducted, and the following is obtained:



Figure 19: ANOVA table for alloy1 dentists

**Result:**

* The P-value (0.116567)>0.05, hence we fail to reject null hypothesis.
* So, the implant hardness of alloy1 is same for all dentists.

**ALLOY2:**

**Hypothesis Definition:**

**Null Hypothesis (H0):** The implant hardness of alloy2 is same for all dentists

**Alternate Hypothesis (H1):** The implant hardness of alloy2 is not same for all dentists

ANOVA test is conducted, and the following is obtained:

A number on a white background

Description automatically generated

Figure 20: ANOVA table for alloy2 dentists

**Result:**

* The P-value (0.718031)>0.05, hence we fail to reject null hypothesis.
* So, the implant hardness of alloy2 is same for all dentists.
  1. **Is there any difference among the methods on the implant hardness?**

Plot was drawn for the distribution of alloy1 and alloy2 with respect to methods.

A graph of a distribution of response

Description automatically generated

Figure 21: Response distribution of alloy1 with respect to methods

A graph of a distribution of response

Description automatically generated

Figure 22: Response distribution of alloy2 with respect to methods

A chart with blue and orange squares

Description automatically generated

Figure 23: Response Vs Method for alloy1 and alloy2

The boxplot of Method vs Response shows the presence of a few outliers.

**ALLOY1:**

**Hypothesis Definition:**

**Null Hypothesis (H0):** The implant hardness of alloy1 is the same for all methods

**Alternate Hypothesis (H1):**The implant hardness of alloy1 is not the same for all methods

ANOVA test is conducted, and the following is obtained:

A number on a white background

Description automatically generated

Figure 24: ANOVA table for alloy1 methods

**Result:**

* The P-value (0.004163) <0.05, hence we reject the null hypothesis.
* So, the implant hardness of alloy1 is not same for all methods.

To check which means are different, we perform the Tucky-HSD test:

A black and white text on a white background

Description automatically generated

Figure 25: Tucky-HSD table for alloy1

**ALLOY2:**

**Hypothesis Definition:**

**Null Hypothesis (H0):** The implant hardness of alloy2 is same for all methods

**Alternate Hypothesis (H1):** The implant hardness of alloy2 is not same for all methods

ANOVA test is conducted, and the following is obtained:



Figure 26: ANOVA table for alloy2 methods

**Result:**

* The P-value (0.000005) <0.05, hence we reject null hypothesis.
* So, the implant hardness of alloy2 is not same for all methods.

**Tucky-HSD test:**

A black and white text on a white background

Description automatically generated

Figure 27: Tucky-HSD table for alloy2

* 1. **Is there any difference among the temperature levels on the implant hardness?**

Plot was drawn for the distribution of alloy1 and alloy2 with respect to methods.

A graph of a distribution of response

Description automatically generated

Figure 28: Response distribution of alloy1 with respect to Temperature

A graph of a distribution of response

Description automatically generated

Figure 29: Response distribution of alloy2 with respect to Temperature

A graph of different colored squares

Description automatically generated

Figure 30: Temp Vs Response for alloy1 and alloy2

The boxplot of Temperature vs Response shows the presence of a few outliers.

**ALLOY1:**

**Hypothesis Definition:**

**Null Hypothesis (H0):** The implant hardness of alloy1 is same for at temperatures

**Alternate Hypothesis (H1)**: The implant hardness of alloy1 is not same for at temperatures

ANOVA test is conducted, and the following is obtained:

A number with numbers on it

Description automatically generated

Figure 31: ANOVA table for alloy1 temperatures

**Result:**

* The P-value (0.717074)>0.05, hence we fail to reject null hypothesis.
* So, the implant hardness of alloy1 is same at all temperatures.

**ALLOY2:**

**Hypothesis Definition:**

**Null Hypothesis (H0):** The implant hardness of alloy2 is same at all temperatures

**Alternate Hypothesis (H1):** The implant hardness of alloy2 is not same at all temperatures

ANOVA test is conducted, and the following is obtained:

A number with black text

Description automatically generated

Figure 32: ANOVA table for alloy2 temperatures

**Result:**

* The P-value (0.164678)>0.05, hence we fail to reject null hypothesis.
* So, the implant hardness of alloy2 is same for all dentists.