ADVANCED STATS PROJECT REPORT

DSBA



Table of Contents

Problem 1

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

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

1.3 What is the probability that a randomly chosen player plays in a striker position and has a foot injury?

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

1.5 What is the probability that a randomly chosen injured player is either a forward or an attacking midfielder?

Problem 2

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

2.2 What is the probability of a radiation leak?

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.

**Problem 3**

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

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

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

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

**Problem 4**

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

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

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

**Problem 5**

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?

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

**Problem 6**

6 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 7**

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.?

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

7.3 Irrespective of your conclusion in 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?

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?

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?

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

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?

**Problem 1**

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Striker | Forward | Attacking Midfielder | Winger | **Total** |
| Players Injured | 45 | 56 | 24 | 20 | **145** |
| Players Not Injured | 32 | 38 | 11 | 9 | **90** |
| **Total** | **77** | **94** | **35** | **29** | **235** |

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

Ans.) Probability that a randomly chosen player would suffer an injury = Total no. of players who all are injured / Total no. of players in the data.

Pr.(Player Injury) = n(Injured Players) / n(Total players)

Pr.(Player Injury) = 140 / 235

**Pr.(Player Injury) = 0.59**

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

Ans.) Probability that a player is a forward or a winger = Probability that a player is a forward + Probability that a player is a winger

Pr.(forward or winger) = Pr.(player forward) + Pr.(player winger)

Pr.(forward or winger) = [n(forward players)/n(total players)] + [n(winger players)/n(total players)]

Pr.(forward or winger) = (94 / 235) + (29 / 235)

Pr.(forward or winger) = 123/235

**Pr.(forward or winger) = 0.52**

**1.3 What is the probability that a randomly chosen player plays in a striker position and has a foot injury?**

Ans.) Probability that a randomly chosen player plays in a striker position and has foot injury = Total no. of players who are injured in striker position / Total no. of players in Striker position.

Pr.(Striker player with injury) = n(Injured striker players) / n(Players in striker position)

Pr.(Striker player with injury) = 45 / 77

**Pr.(Striker player with injury) = 0.58**

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

Ans.) Probability that a randomly chosen injured player is a striker = Total no. of players who are injured in striker position / Total no. of Injured players

Pr.(Injury player is from Striker) = n(Injured striker players) / n(Players injured)

Pr.(Injury player is from Striker) = 45 / 145

**Pr.(Injury player is from Striker) = 0.31**

**1.5 What is the probability that a randomly chosen injured player is either a forward or an attacking midfielder?**

Ans.) Probability that a randomly chosen injured player is either a forward or an attacking midfielder = Probability that a chosen injured player is Forward + Probability that a chosen injured player is Attacking midfielder.

Pr.(Injured player is Forward or Attacking midfielder) = Pr.(Injured player is Forward) + Pr.(Injured player is Attacking midfielder)

Pr.(Injured player is Forward or Attacking midfielder) = [n(Forward players)/n(Total injured players)] + [n(Attacking midfielder players)/n(Total injured players)]

Pr.(Injured player is Forward or Attacking midfielder) = (56 / 145) + (24 / 145)

Pr.(Injured player is Forward or Attacking midfielder) = 80 / 145

**Pr.(Injured player is Forward or Attacking midfielder) = 0.55**

**Problem 2**

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.

According to the studies carried out by the organization, the probability of a radiation leak in case of a fire is 20%, the probability of a radiation leak in case of a mechanical 50%, and the probability of a radiation leak in case of a human error is 10%. The studies also showed the following;

* The probability of a radiation leak occurring simultaneously with a fire is 0.1%.
* The probability of a radiation leak occurring simultaneously with a mechanical failure is 0.15%.
* The probability of a radiation leak occurring simultaneously with a human error is 0.12%.

Ans.)

Given Probabilities from the question are,

* Probability of radiation leak in case of a fire is
  + Prob\_RadLeak\_given\_Fire = 0.2
* Probability of radiation leak in case of a Mechanical Failure is
  + Prob\_RadLeak\_given\_MechFail = 0.5
* Probability of radiation leak in case of a Human Error is
  + Prob\_RadLeak\_given\_HumanErr = 0.1

Also given probabilities for,

* Probability of a radiation leak occurring simultaneously with a fire is
  + Prob\_RadLeak\_and\_Fire = 0.001
* Probability of a radiation leak occurring simultaneously with a Mechanical Failure is
  + Prob\_RadLeak\_and\_MechFail = 0.0015
* Probability of a radiation leak occurring simultaneously with a Human Error is
  + Prob\_RadLeak\_and\_HumanErr = 0.0012

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

Ans.)

We can use Conditional Probability to find out probabilities separately using formula,

* **P(A|B) = P(A and B) / P(B) => P(B) = P(A and B) / P(A|B)**
* where, P(A|B) is Probability of event A given event B
* P(A and B) is Probability of event A and event B happening together
* P(B) is Probability of event B occurring.

Using above formula we can modify as,

* Prob\_RadLeak\_given\_Fire = Prob\_RadLeak\_and\_Fire / Prob\_Fire
  + Prob\_Fire = Prob\_RadLeak\_and\_Fire / Prob\_RadLeak\_given\_Fire
* Prob\_RadLeak\_given\_MechFail = Prob\_RadLeak\_and\_MechFail / Prob\_MechFail
  + Prob\_MechFail = Prob\_RadLeak\_and\_MechFail / Prob\_RadLeak\_given\_MechFail
* Prob\_RadLeak\_given\_HumanErr = Prob\_RadLeak\_and\_HumanErr / Prob\_HumanErr
  + Prob\_HumanErr = Prob\_RadLeak\_and\_HumanErr / Prob\_RadLeak\_given\_HumanErr

**OUTPUT:**

* The Probability of **Fire** is **0.005**
* The Probability of **Mechanical Failure** is **0.003**
* The Probability of **Human Error** is **0.012**

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

Ans.)

Probability of radiation leak can be calculated using the law of total probability, which states that the probability of an event can be determined by summing the probabilities of that event occurring under each possible condition.

Probability of Fire given Radiation Leak + Probability of Mechanical Error given Radiation Leak + Probability of Human Error given Radiation Leak = 1

**Prob\_Fire\_given\_RadLeak + Prob\_MechFail\_given\_RadLeak + Prob\_HumanErr\_given\_RadLeak = 1**

We already know that,

* Prob\_Fire\_given\_RadLeak = (Prob\_RadLeak\_given\_Fire \* Prob\_Fire)/Prob\_RadLeak
* Prob\_MechFail\_given\_RadLeak = (Prob\_RadLeak\_given\_MechFail \* Prob\_MechFail)/Prob\_RadLeak
* Prob\_HumanErr\_given\_RadLeak = (Prob\_RadLeak\_given\_HumanErr \* Prob\_HumanErr)/Prob\_RadLeak
* (Prob\_RadLeak\_given\_Fire \* Prob\_Fire)/Prob\_RadLeak + (Prob\_RadLeak\_given\_MechFail \* Prob\_MechFail)/Prob\_RadLeak + (Prob\_RadLeak\_given\_HumanErr \* Prob\_HumanErr)/Prob\_RadLeak = 1
* **Prob\_RadLeak = (Prob\_RadLeak\_given\_Fire \* Prob\_Fire + Prob\_RadLeak\_given\_MechFail \* Prob\_MechFail + Prob\_RadLeak\_given\_HumanErr \* Prob\_HumanErr)**

**OUTPUT**:

The 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.**

Ans.)

Probability of Radiation leak caused by any one of Fire, Mechanical Failure, and Human Error is as:

Probabiliy of Fire given Radiation leak is

* Prob\_Fire\_given\_RadLeak = (Prob\_RadLeak\_given\_Fire \* Prob\_Fire)/Prob\_RadLeak

Probabiliy of Mechanical Failure given Radiation leak is

* Prob\_MechFail\_given\_RadLeak = (Prob\_RadLeak\_given\_MechFail \* Prob\_MechFail)/Prob\_RadLeak

Probabiliy of Human Error given Radiation leak is

* Prob\_HumanErr\_given\_RadLeak = (Prob\_RadLeak\_given\_HumanErr \* Prob\_HumanErr)/Prob\_RadLeak

**OUTPUT**:

The Probability of **Fire given Radiation Leak** is **0.27**

The Probability of **Mechanical Failure given Radiation Leak** is **0.405**

The Probability of **Human Error given Radiation Leak** is **0.324**

**Problem 3**

The breaking strength of gunny bags used for packaging cement is normally distributed with a mean of 5 kg per sq. centi-meter and a standard deviation of 1.5 kg per sq. centi-meter. 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.

Ans.)

From problem, given mean(mu=5) and standard deviation(sigma=1.5)

As, standard deviation of population is given, then we can use **Z-test** for this problem.

Formula for Z-test is,

**Test Statistic = (x\_bar – mu) / sigma**

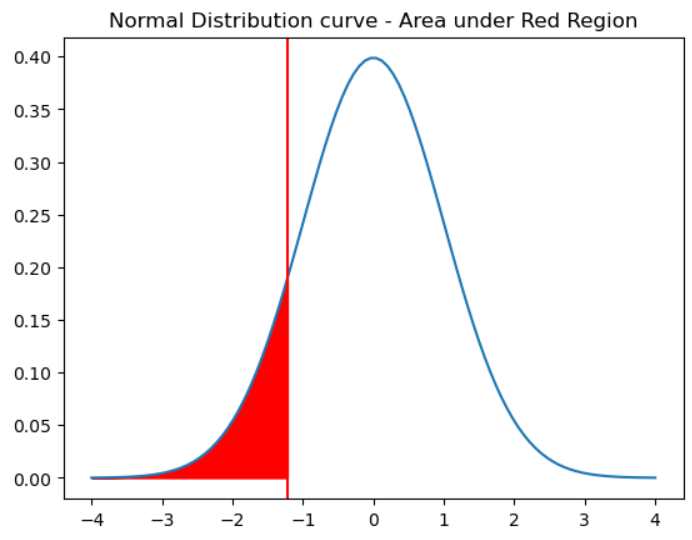
Where, x\_bar is given x value to calculate statistic

Mu is Mean of the Population

Sigma is Standard Deviation of the Population

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

Ans.)

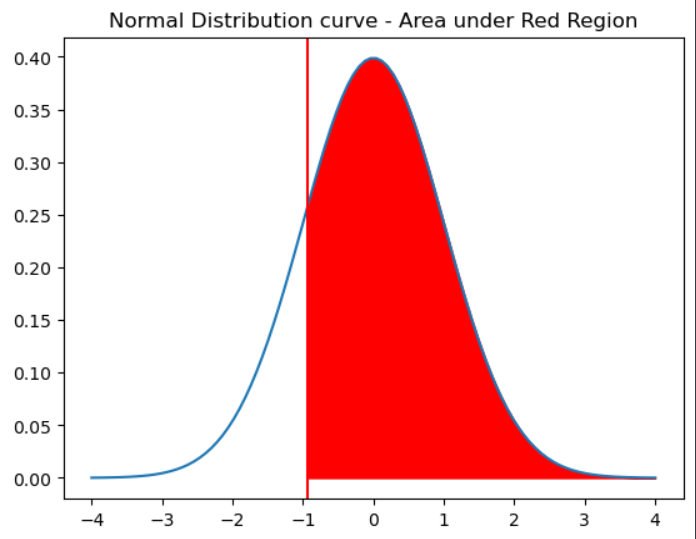


Cumulative Distribution Area under the Red Region is **0.111**

* From above plot, we need to find out the probability or area under the shaded region for the given question that proportion less than 3.17 kg per sq cm
* We can see that the cumulative probability is just 11.1%

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

Ans.)

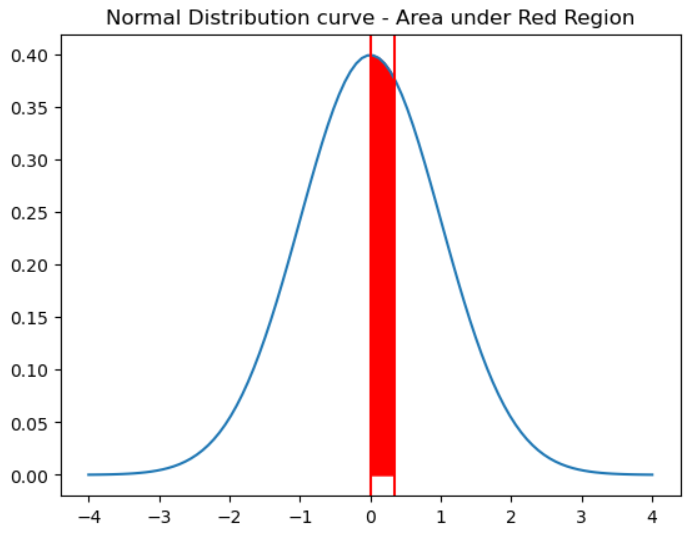


Cumulative Distribution Area under the Red Region is **0.825**

* From above plot, we need to find out the probability or area under the shaded region for the given question that the proportion at least 3.6 kg per sq cm.
* We can see that the cumulative probability is 82.5% which is very high. We can assume that proportion of the gunny bags have a breaking strength at least 3.6 kg per sq cm is more.

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

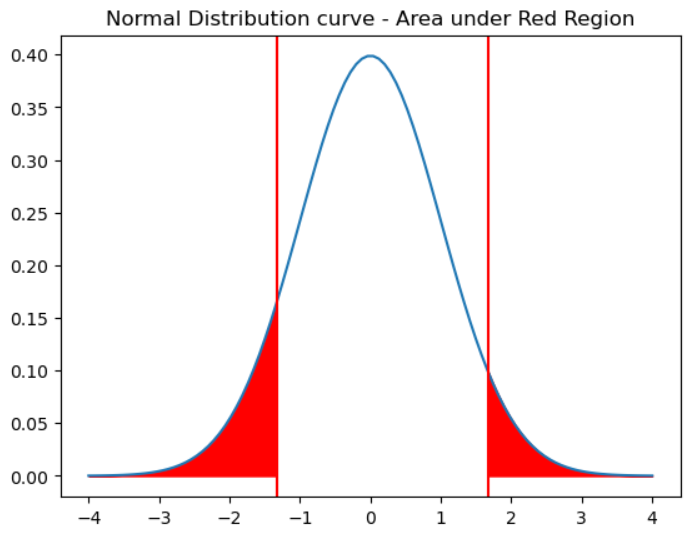
Ans.)



Cumulative Distribution Area under the Red Region is **0.131**

* From above plot, we need to find out the probability or area under the shaded region for the given question that proportion between 5 and 5.5 kg per sq. cm.
* We can see that the cumulative probability is 13.1% which is low because the interval between 5 and 5.5 kg per sq. cm is very less.

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



Cumulative Distribution Area under the Red Region is **0.139**

* From above plot, we need to find out the probability or area under the shaded region for the given question that proportion NOT between 3 and 7.5 kg per sq. cm.
* We can see that the cumulative probability is 13.9% which is low because the interval NOT between 3 and 7.5 kg per sq. cm is very less.

**Problem 4**

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.

Ans.)

From problem, given mean(mu=77) and standard deviation(sigma=8.5)

As, standard deviation of population is given, then we can use Z-test for this problem.

Formula for Z-test is,

**Test Statistic = (x\_bar – mu) / sigma**

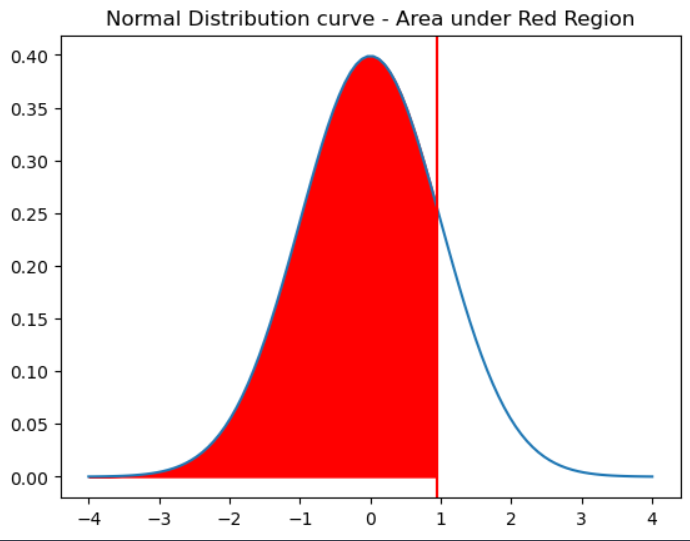
Where, x\_bar is given x value to calculate statistic

Mu is Mean of the Population

Sigma is Standard Deviation of the Population

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

Ans.)

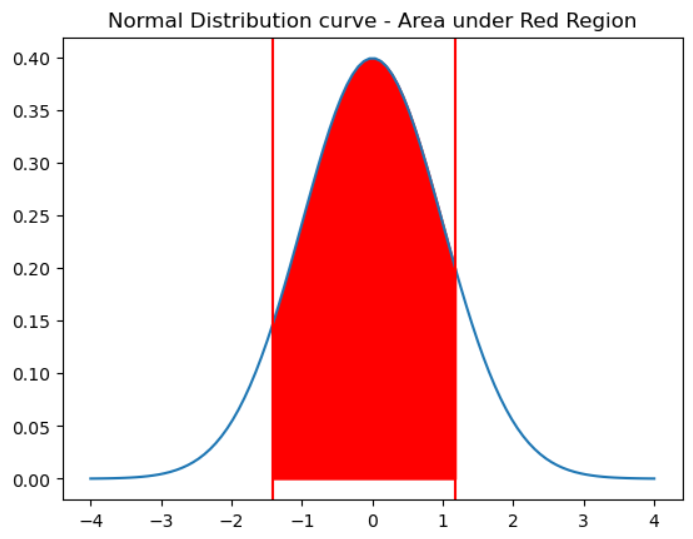


Cumulative Distribution Area under the Red Region is **0.827**

* From above plot, we need to find out the probability or area under the shaded region for the given question that probability of grade less than 85.
* We can see that the cumulative probability is 82.7%

**4.2 What is the probability that a randomly selected student scores between 65 and 87?**

Ans.)



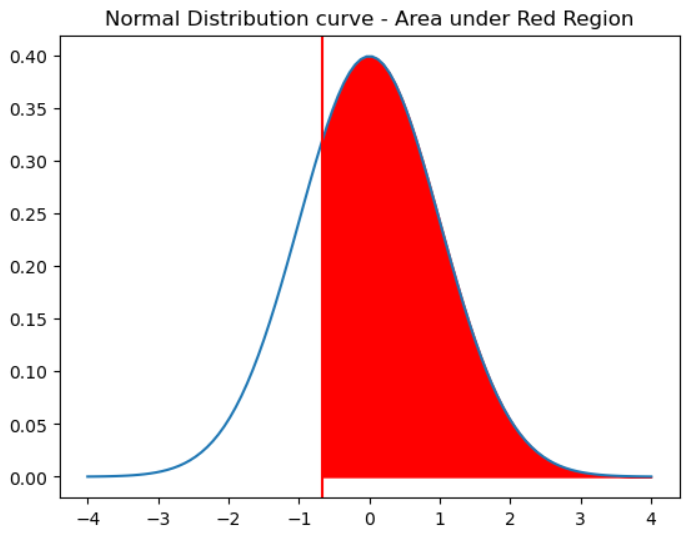
Cumulative Distribution Area under the Red Region is **0.801**

* From above plot, we need to find out the probability or area under the shaded region for the given question that probability of scores between 65 and 87.
* From below, we can see that the cumulative probability is 80.1%

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

Ans.)

Passing cut-off so that 75% of the students clear the exam is **71.267**



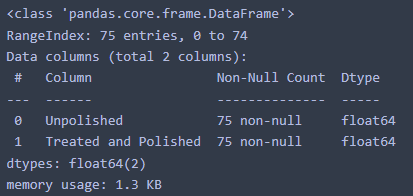
Cumulative Distribution Area under the Red Region is **0.75**

**Problem 5**

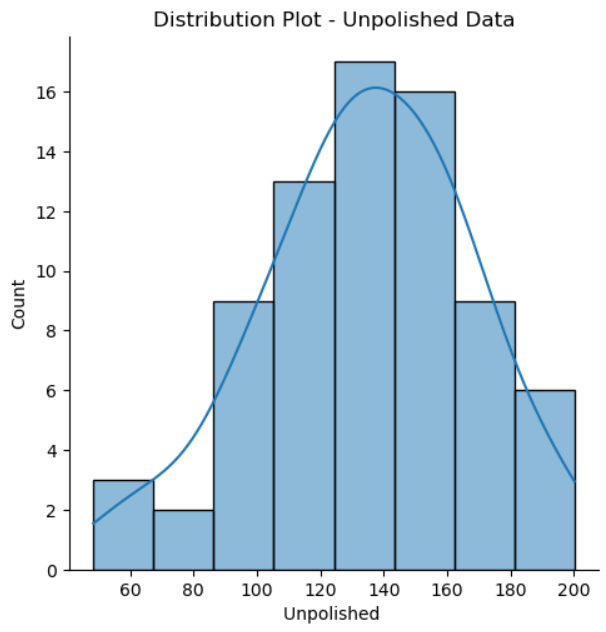
Zingaro stone printing is a company that specializes in printing images or patterns on polished or unpolished stones. However, for the optimum level of printing of the image the stone surface must have a Brinell's hardness index of at least 150. Recently, Zingaro has received a batch of polished and unpolished stones from its clients. Use the data provided to answer the following (assuming a 5% significance level).

Ans.)

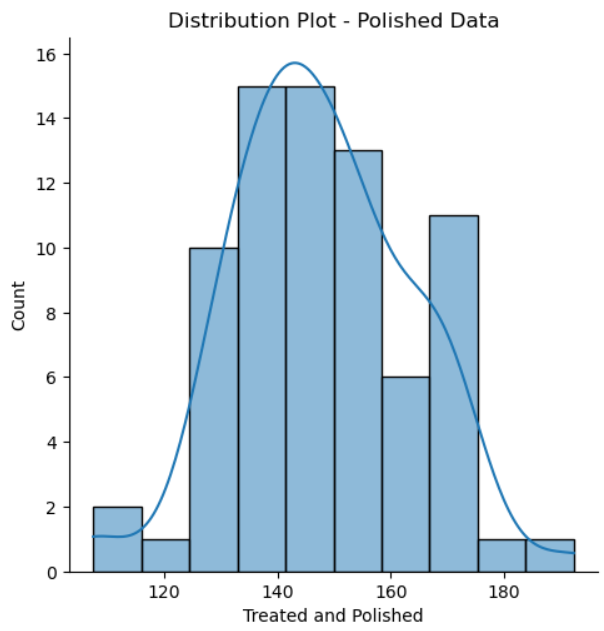
Glimpse of the data,



There are no null values and we have float data in both columns.



Unpolished data is almost normally distributed.



Polished data is slightly normally distributed with some skewness.

**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?**

Ans.)

Hypothesis -

* **H0 (Null Hypothesis)** - Hardness index of unpolished stones >=150
* **H1 (Alternative Hypothesis)** - Hardness index of unpolished stones < 150
* Level of significance is 5% i.e., **alpha=0.05**
* It is a **One-Tailed test**, so we need to divide p-value by 2 for comparing with alpha.

For Unpolished data,

* Test statistic value = -4.165
* P-value = 4e-05

**Conclusion:**

* P value is too less than alpha i.e., 0.000042 (<<<0.05), Hence we have enough evidence to **Reject Null Hypothesis** stating that Hardness index of unpolished stones is less than 150.
* Zingaro has enough reason to believe that the unpolished stones will not be suitable for printing at 95% confidence.

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

Ans.)

Hypothesis -

* **H0 (Null Hypothesis)** - Mean hardness of polished stones = Mean hardness of unpolished stones
* **H1 (Alternative Hypothesis)** - Mean hardness of polished stones != Mean hardness of unpolished stones
* Level of significance is 5% i.e., **alpha=0.05**
* It is a **Two-Tailed test** and 2 sample tests.

From T-Test Independent samples,

* Test statistic value = -3.242
* P-value = 0.001

**Conclusion:**

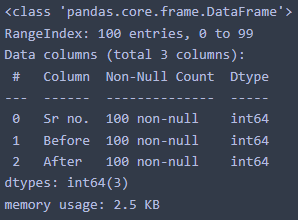
* P value is less than alpha i.e., 0.0015 (<0.05), Hence we have enough evidence to **Reject Null Hypothesis**.
* This concludes that Mean hardness of polished stones is not equal to the Mean hardness of unpolished stones.

**Problem 6**

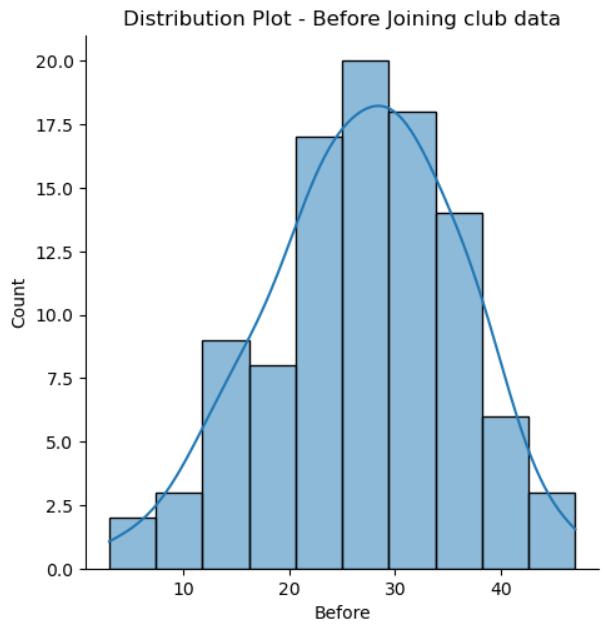
Aquarius health club, one of the largest and most popular cross-fit gyms in the country has been advertising a rigorous program for body conditioning. The program is considered successful if the candidate can do more than 5 push-ups, as compared to when he/she enrolled in the program. Using the sample data provided can you conclude whether the program is successful? (Consider the level of Significance as 5%)

Ans.)

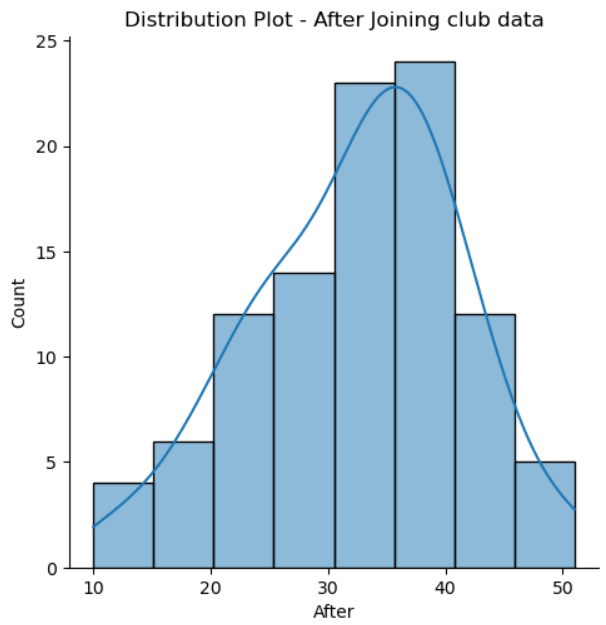
Glimpse of the dataset,



There are no null values and all are Integer data type.



Data is normally distributed for customers Before Joining club.



Data is slightly normally distributed for customers After Joining club.

**Q) - 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.

Hypothesis -

* **H0 (Null Hypothesis):** Push-ups after - Push-ups before <= 5
* **H1 (Alternative Hypothesis):** Push-ups after - Push-ups before > 5
* Level of significance is 5% i.e., **alpha = 0.05**
* It is a **One-Tailed – 2 sample tests**, so we need to divide p-value by 2 for comparing with alpha.

From T-Test 2 sample analysis,

* Test Statistic value is -19.323
* P-value is 1.14e-35 (very close to 0)

Conclusion:

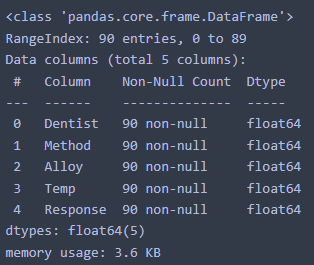
* P value is too less than alpha or close to zero (0) i.e., 1.1 \* 10-35 (<<<<0.05), Hence we have enough evidence to **Reject Null Hypothesis** stating that Push-ups after – Push-ups before >5.
* This concludes that, **Training will make a difference of more than 5 push-ups when compared to before start of the program**.

**Problem 7**

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.

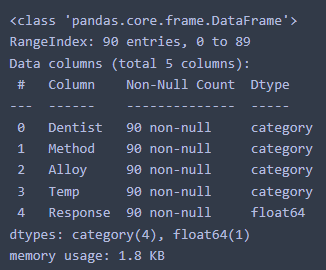
Ans.)

Glimpse of the data,

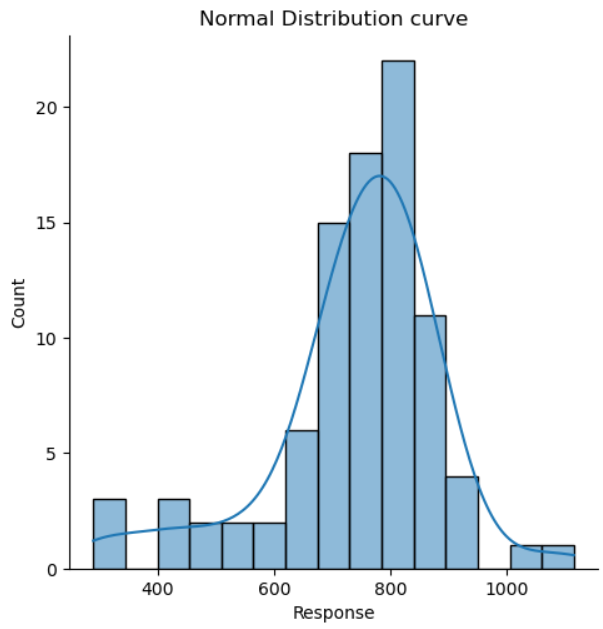


There are no null values and all are float type data.

The columns Dentist, Method, Alloy, and Temp are Categorical data and we can convert them and do analysis.



Columns are casted to respective categories after analysis.



Response or Target Variable which is almost normally distributed with some skewness.

**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?**

Ans.)

**For Alloy 1,**

**Null hypothesis (for Alloy 1):** There is no difference in means among the dentists in terms of implant hardness for Alloy 1.

**Alternative hypothesis (for Alloy 1):** There is a difference in means among the dentists in terms of implant hardness for Alloy 1.

**For Alloy 2,**

**Null hypothesis (for Alloy 2):** There is no difference in means among the dentists in terms of implant hardness for Alloy 2.

**Alternative hypothesis (for Alloy 2):** There is a difference in means among the dentists in terms of implant hardness for Alloy 2.

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

Ans.)

**Hypothesis Test for Shapiro Wilk Test for Alloy 1 -**

* **Null Hypothesis (for Alloy 1)**: The population from which the sample is drawn follows a normal distribution.
* **Alternative Hypothesis (for Alloy 1):** The population from which the sample is drawn does not follow a normal distribution.

From Analysis of Wilk test,

* Test Statistic value = 0.83
* P-value = 1e-05

**Conclusion**:

* We can conclude that P-value is 1e-05 (<<0.05), so we have enough evidence to **Reject Null Hypothesis** and consider that sample drawn for Alloy 1 does not follow a normal distribution.

**Hypothesis Test for Shapiro Wilk Test for Alloy 2 -**

* **Null Hypothesis (for Alloy 2)**: The population from which the sample is drawn follows a normal distribution.
* **Alternative Hypothesis (for Alloy 2):** The population from which the sample is drawn does not follow a normal distribution.

From Analysis of Wilk test,

* Test Statistic value = 0.888
* P-value = 0.0004

**Conclusion**:

* We can conclude that P-value is 0.0004 (<<0.05), so we have enough evidence to **Reject Null Hypothesis** and consider that sample drawn for Alloy 2 does not follow a normal distribution.

**Hypothesis Test for Levene Test for Alloy data -**

* **Null Hypothesis**: The variances of the groups or samples being compared are equal for Alloy 1 and Alloy 2 data.
* **Alternative Hypothesis:** The variances of the groups or samples being compared are not equal for Alloy 1 and Alloy 2 data.

From Analysis of Levene test,

* Test Statistic value = 2.085
* P-value = 0.1523

**Conclusion:**

* We can conclude that P-value is 0.152 (>0.05), so we have enough evidence for **Fail to Reject Null Hypothesis** and consider that variances of the groups or samples being compared are equal for Alloy 1 and Alloy 2 data.

**7.3 Irrespective of your conclusion in 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?**

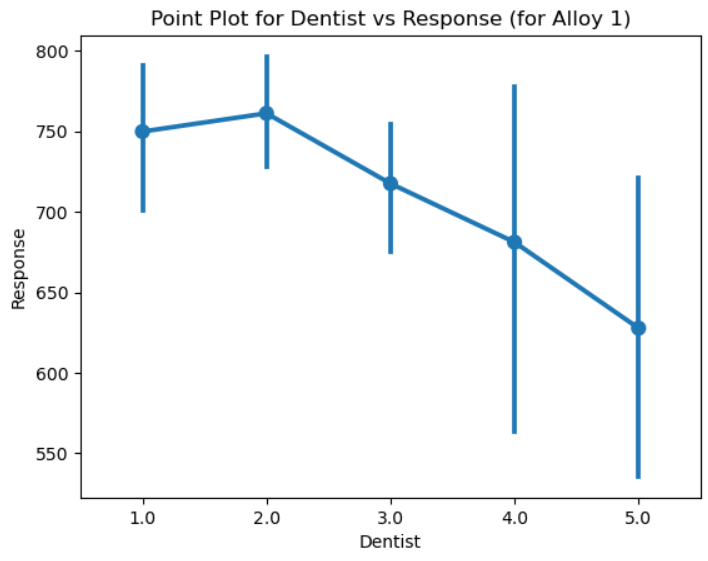
Ans.)

**Hypothesis Test for Alloy 1 data –**

* **Null hypothesis (for Alloy 1):** There is no difference in means among the dentists in terms of implant hardness for Alloy 1.
* **Alternative hypothesis (for Alloy 1):** There is a difference in means among the dentists in terms of implant hardness for Alloy 1.

Analysis –

* **P-value is 0.12** which is greater than alpha i.e., 0.05. Hence, we **Fail to reject null hypothesis** and consider there is no difference in means among the dentists in terms of implant hardness for Alloy 1.



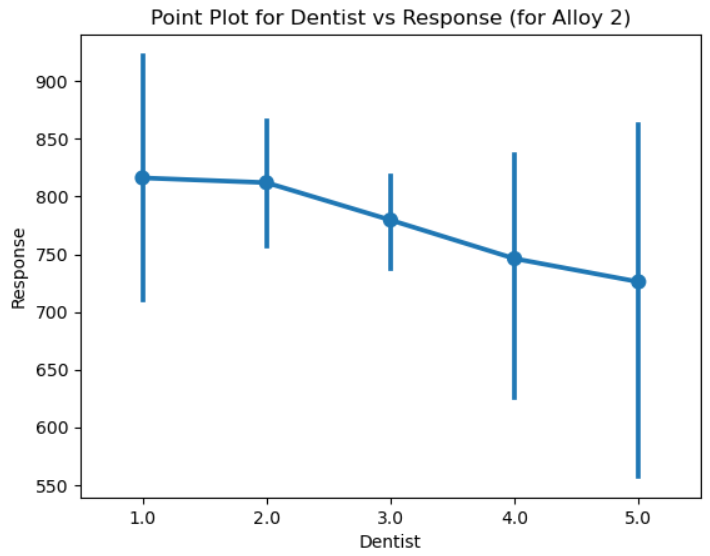
There’s slight difference but not very significant for our analysis.

**Hypothesis Test for Alloy 2 data –**

* **Null hypothesis (for Alloy 2):** There is no difference in means among the dentists in terms of implant hardness for Alloy 2.
* **Alternative hypothesis (for Alloy 2):** There is a difference in means among the dentists in terms of implant hardness for Alloy 2.

Analysis –

* **P-value is 0.72** which is very greater than alpha i.e., 0.05. Hence, we **Fail to reject null hypothesis** and consider there is no difference in means among the dentists in terms of implant hardness for Alloy 2.



There is no much difference in data for Alloy 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?**

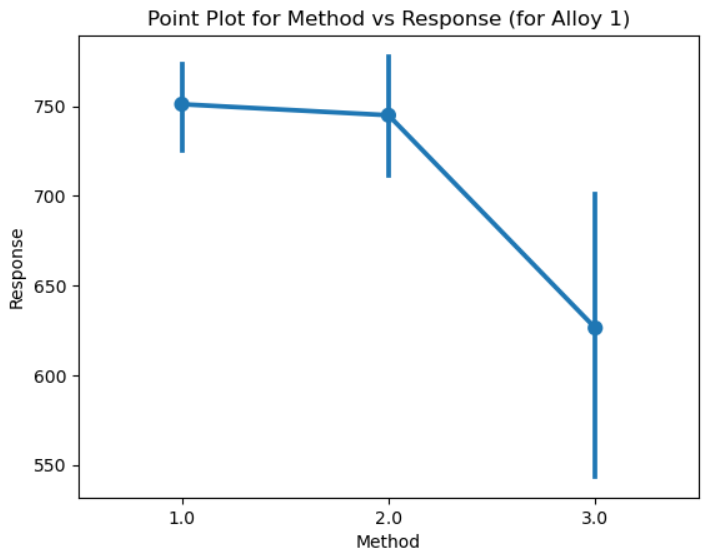
Ans.)

**Hypothesis Test for Alloy 1 data** –

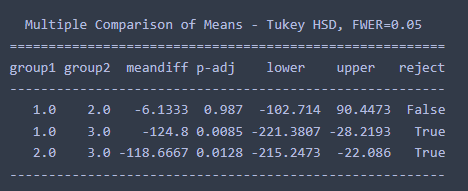
* **Null hypothesis (for Alloy 1):** There is no difference in means among the Methods in terms of implant hardness for Alloy 1.
* **Alternative hypothesis (for Alloy 1):** There is a difference in means among the Methods in terms of implant hardness for Alloy 1.

Analysis –

* **P-value is 0.004** which is lesser than alpha i.e., 0.05. Hence, we have enough evidence to **reject null hypothesis** and consider there is a difference in means among the Methods in terms of implant hardness for Alloy 1.



Method 3 has lot variation when compared with other two.



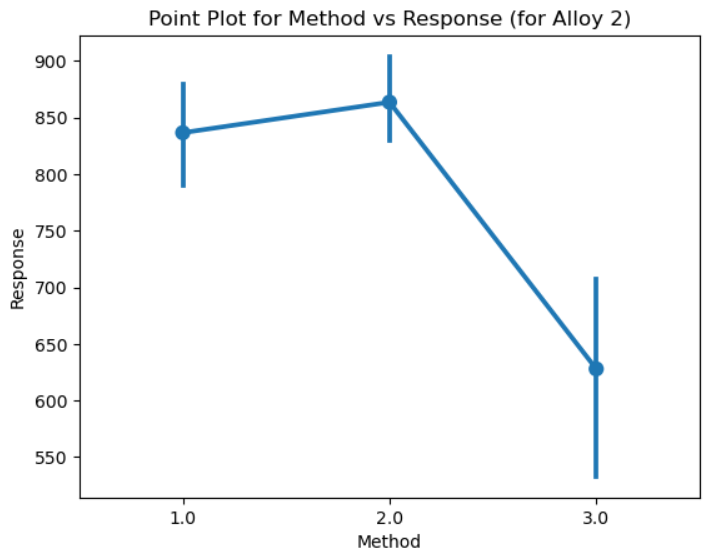
* The pairs which are interlinked here are
  + Method 1 with Method 3
  + Method 2 with Method 3

**Hypothesis Test for Alloy 2 data** –

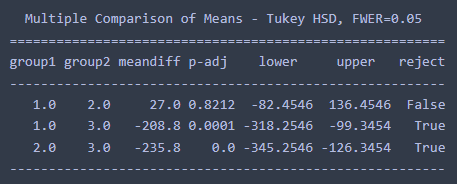
* **Null hypothesis (for Alloy 2):** There is no difference in means among the Methods in terms of implant hardness for Alloy 2.
* **Alternative hypothesis (for Alloy 2):** There is a difference in means among the Methods in terms of implant hardness for Alloy 2.

Analysis –

* **P-value is 0.000005** which is very lesser than alpha i.e., 0.05. Hence, we have enough evidence to **reject null hypothesis** and consider there is a difference in means among the Methods in terms of implant hardness for Alloy 2.



Method 3 has variation when compared with other two.



* The pairs which are interlinked here are,
  + Method 1 with Method 3
  + Method 2 with Method 3

**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?**

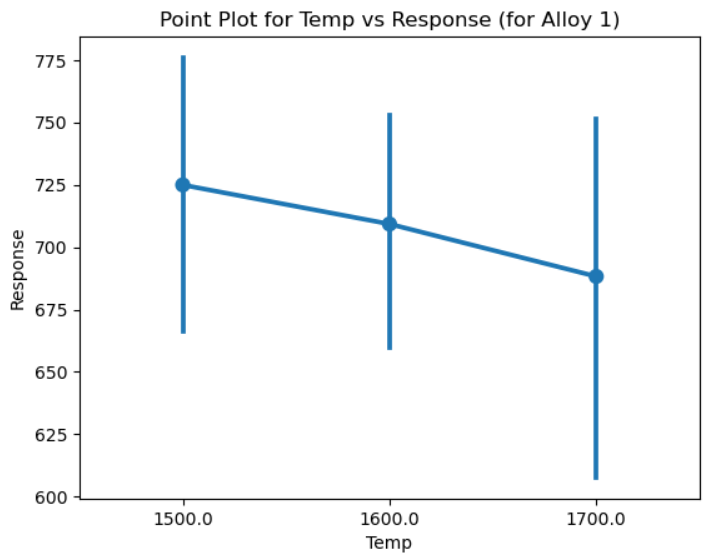
Ans.)

**Hypothesis Test for Alloy 1 data –**

* **Null hypothesis (for Alloy 1):** There is no difference in means among the Temperature levels in terms of implant hardness for Alloy 1.
* **Alternative hypothesis (for Alloy 1):** There is a difference in means among the Temperature levels in terms of implant hardness for Alloy 1.

Analysis –

* **P-value is 0.72** which is very greater than alpha i.e., 0.05. Hence, we **Fail to reject null hypothesis** and consider there is no difference in means among the Temperature levels in terms of implant hardness for Alloy 1.



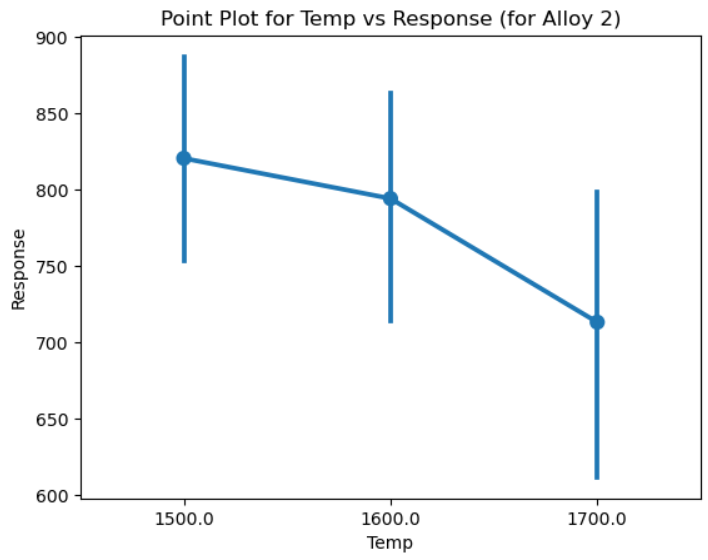
The 3 levels of Temp look with no much difference among them.

**Hypothesis Test for Alloy 2 data –**

* **Null hypothesis (for Alloy 2):** There is no difference in means among the Temperature levels in terms of implant hardness for Alloy 2.
* **Alternative hypothesis (for Alloy 2):** There is a difference in means among the Temperature levels in terms of implant hardness for Alloy 2.

Analysis –

* **P-value is 0.17** which is very greater than alpha i.e., 0.05. Hence, we **Fail to reject null hypothesis** and consider there is no difference in means among the Temperature levels in terms of implant hardness for Alloy 2.



There’s slight difference but not very significant for Temp.

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

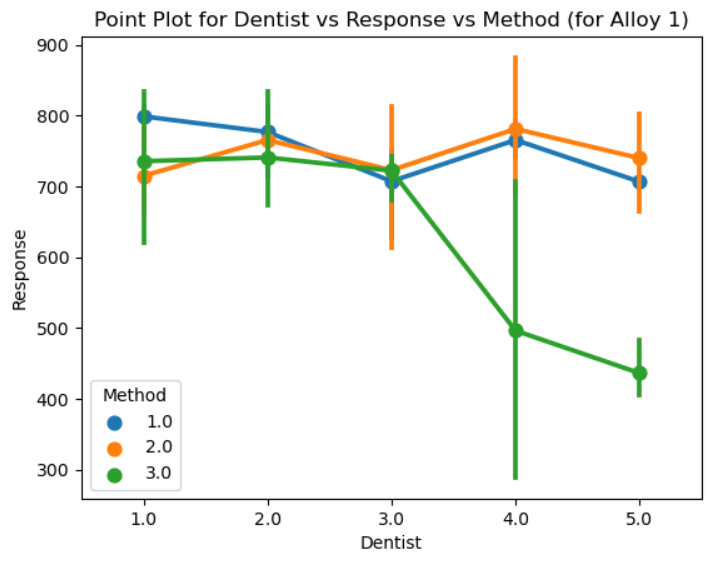
Ans.)

**Hypothesis Test for Alloy 1 data –**

* **Null hypothesis (for Alloy 1):** There is no difference in means among the Interaction effect between Dentist and Method levels in terms of implant hardness for Alloy 1.
* **Alternative hypothesis (for Alloy 1):** There is a difference in means among the Interaction effect between Dentist and Method levels in terms of implant hardness for Alloy 1.

Analysis –

* **P-value is 0.000221** which is lesser than alpha i.e., 0.05. Hence, we have enough evidence to **reject null hypothesis** and consider there is a difference in means among the Interaction effect between Dentist and Method levels in terms of implant hardness for Alloy 1.



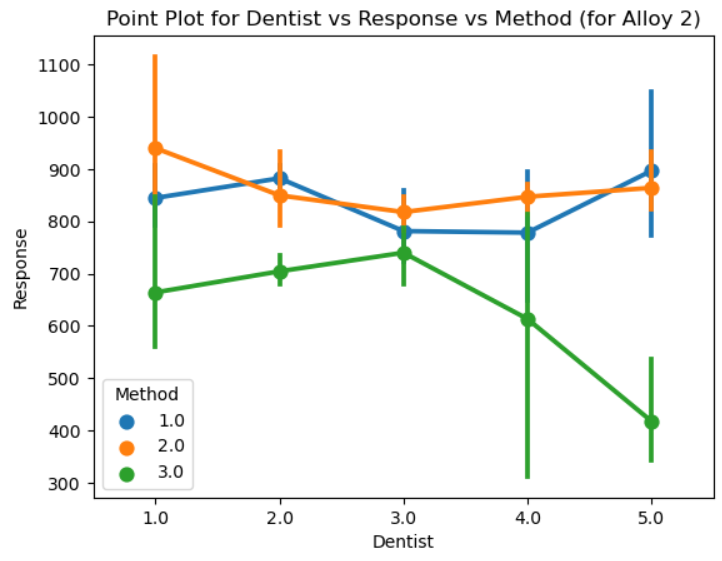
There is lots of Interaction between Dentist and Method variables

**Hypothesis Test for Alloy 2 data –**

* **Null hypothesis (for Alloy 2):** There is no difference in means among the Interaction effect between Dentist and Method levels in terms of implant hardness for Alloy 2.
* **Alternative hypothesis (for Alloy 2):** There is a difference in means among the Interaction effect between Dentist and Method levels in terms of implant hardness for Alloy 2.

Analysis –

* **P-value is 0.000482** which is lesser than alpha i.e., 0.05. Hence, we have enough evidence to **reject null hypothesis** and consider there is significant difference in means among the Interaction effect between Dentist and Method levels in terms of implant hardness for Alloy 2.



There is Interaction between Dentist and Method in Alloy 2.

* The pairs which are interlinked here are (Dentist : Method),
  + 1.0:1.0 with 5.0:3.0
  + 1.0:2.0 with 5.0:3.0
  + 2.0:1.0 with 5.0:3.0
  + 2.0:2.0 with 5.0:3.0
  + 3.0:1.0 with 5.0:3.0
  + 3.0:2.0 with 5.0:3.0
  + 4.0:1.0 with 5.0:3.0
  + 4.0:2.0 with 5.0:3.0
  + 5.0:1.0 with 5.0:3.0
  + 5.0:2.0 with 5.0:3.0

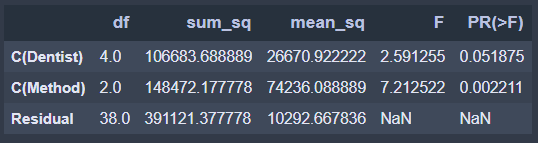
**7.7 Now consider the effect of both factors, dentist, and method, separately on each alloy. What do you conclude? Is it possible to identify which dentists are different, which methods are different, and which interaction levels are different?**

Ans.)

**Hypothesis Test for Alloy 1 data –**

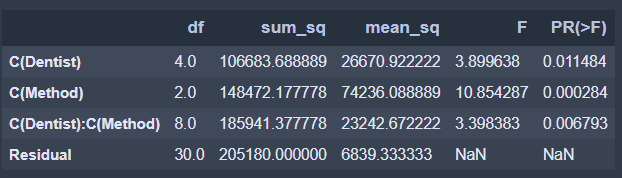
* **Null hypothesis (for Alloy 1):** There is no difference in means among the factors Dentist and Method levels in terms of implant hardness for Alloy 1.
* **Alternative hypothesis (for Alloy 1):** There is a difference in means among the factors Dentist and Method levels in terms of implant hardness for Alloy 1.

Analysis –



Addition of 2 factors but without Interaction variable.

* **P value for Method is 0.0022** which is less than 0.05. Hence, we have enough evidence to **reject null hypothesis** and consider there is at least one pair of Method means are different for Alloy 1.
* The pairs which are interlinked here are
  + Method 1 with Method 3
  + Method 2 with Method 3



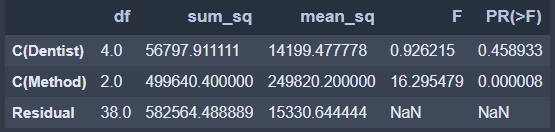
Addition of 2 factors with Interaction variable.

* P value for Dentist, Method, and Interaction variable (Dentist : Method) are all less than 0.05. Hence, we can conclude that to **reject null hypothesis** and consider there is at least one pair of variables means are different for Alloy 1.
* The pairs which are interlinked here are (Dentist : Method),
  + 1.0:1.0 with 4.0:3.0
  + 1.0:1.0 with 5.0:3.0
  + 1.0:2.0 with 5.0:3.0
  + 1.0:3.0 with 5.0:3.0
  + 2.0:1.0 with 4.0:3.0
  + 2.0:1.0 with 5.0:3.0
  + 2.0:2.0 with 4.0:3.0
  + 2.0:2.0 with 5.0:3.0
  + 2.0:3.0 with 5.0:3.0
  + 3.0:1.0 with 5.0:3.0
  + 3.0:2.0 with 5.0:3.0
  + 3.0:3.0 with 5.0:3.0
  + 4.0:1.0 with 4.0:3.0
  + 4.0:1.0 with 5.0:3.0
  + 4.0:2.0 with 4.0:3.0
  + 4.0:2.0 with 5.0:3.0
  + 5.0:1.0 with 5.0:3.0
  + 5.0:2.0 with 5.0:3.0

**Hypothesis Test for Alloy 2 data –**

* **Null hypothesis (for Alloy 2):** There is no difference in means among the factors Dentist and Method levels in terms of implant hardness for Alloy 2.
* **Alternative hypothesis (for Alloy 2):** There is a difference in means among the factors Dentist and Method levels in terms of implant hardness for Alloy 2.

Analysis –



Addition of 2 factors but without Interaction variable.

* **P value for Method is 0.000008** which is too less than 0.05. Hence, we have enough evidence to **reject null hypothesis** and consider there is at least one pair of Method means are different for Alloy 2.
* The pairs which are interlinked here are,
  + Method 1 with Method 3
  + Method 2 with Method 3



Addition of 2 factors with Interaction variable.

* **P value for Method is 0.000004**, which is only less than 0.05 when compared to Dentist (0.371833) and Interaction variable - Dentist : Method (0.093234). So, we have enough evidence to **reject null hypothesis** stating that at least one of the pair means is different for Alloy 2.