Advanced Statistics Project

By Vinish Vincent

Great Learnings

PGP-DSBA - 23

# Table of content

Table of Content

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……………….3

* 1. What is the probability that a randomly chosen player would suffer an injury? ………………………….3
  2. What is the probability that a player is a forward or a winger? ……………………………………….………….3
  3. What is the probability that a randomly chosen player plays in a striker position and has a foot injury? ………………………………………………………………………………………………………………………………….…….3
  4. What is the probability that a randomly chosen injured player is a striker? ……………………….……….4
  5. What is the probability that a randomly chosen injured player is either a forward or an attacking midfielder? …………………………………………………………………………………………………………………………..…….4

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. ………………………………………….……….4

* 1. What are the probabilities of a fire, a mechanical failure, and a human error respectively? ………………………………………………………………………………………………………………………………………………..….5
  2. What is the probability of a radiation leak? ……………………………………………………………………….……….5
  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. …………………………………………………….…………….5
* A Mechanical Failure. …………………………………………………………………….5
* A Human Error. ………………………………………………………………………….…………….5

Problem 3: 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; (Provide an appropriate visual representation of your answers, without which marks will be deducted)…………………………………………………………………………………………….…………….5

* 1. What proportion of the gunny bags have a breaking strength less than 3.17 kg per sq cm?….6
  2. What proportion of the gunny bags have a breaking strength at least 3.6 kg per sq cm.?…….6
  3. What proportion of the gunny bags have a breaking strength between 5 and 5.5 kg per sq cm.? ………………………………………………………………………………………………………………………….………….7
  4. What proportion of the gunny bags have a breaking strength NOT between 3 and 7.5 kg per sq cm.? …………………………………………………………………………………………………………………………..…….7

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.

* 1. What is the probability that a randomly chosen student gets a grade below 85 on this exam?.….8
  2. What is the probability that a randomly selected student scores between 65 and 87? ……...…..….8
  3. What should be the passing cut-off so that 75% of the students clear the exam? ………...…..……….9

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 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. 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? ……………………………………………………………………………………………………………………………………….…….9
  2. Is the mean hardness of the polished and unpolished stones the same? ……………………..………….10

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 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….………11

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.

* 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.? …………11
  2. Before the hypotheses may be tested, state the required assumptions. Are the assumptions fulfilled? Comment separately on both alloy types.? ……………………………………………………………….11
  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? ………………………………………………………………………………………………………………………………..…….12
  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? …………………………………………….12
  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? …………..….13
  6. Consider the interaction effect of dentist and method and comment on the interaction plot, separately for the two types of alloys? …………………………………………………………………………………….13
  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?.……………………………………………………………….…………….13

# 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?**

**Solution**: Total no. of injured players = 145

Total no. of players = 235

So, P(Injury) = 145/235 = **0.617**

**There is an ~61% probability that a randomly chosen player would suffer an injury**

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

**Solution**: Total no. of forward players (f) = 94

Total no. of winger players (w) = 29

Total no. of players = 235

So, P(f&w) = (94/235) + (29/235) = (123/235) = **0.523**

**There is an ~52% probability that a player is a forward or a winger**

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

**Solution**: Total no. of injured strikers (s&i) = 45

Total no. of players = 235

So, P(s&i) = 45/235 = **0.191**

**There is an ~19% probability that a randomly chosen player plays in a striker position and has a foot injury**

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

**Solution**: Total no. of injured strikers (si) = 45

Total no. of injured players (i) = 145

Total no. of players = 235

So, P(i&si) = 45/145 = **0.310**

**There is an ~31% probability that a randomly chosen injured player is a striker**

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

**Solution**: Total no. of injured forward (fi) = 56

Total no. of injured midfielder (mi) = 24

Total no. of injured players = 145

So, P(fi&mi) = (56 + 24)/145 = 80/145 = **0.551**

**There is an ~55% probability that a randomly chosen injured player is either a forward or an attacking midfielder**

# 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%.

On the basis of the information available, answer the questions below:

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

**Solution**: Probability of radiation leak in case of a fire = 20% = 0.20

Probability of a radiation leak occurring simultaneously with a fire = 0.1% = 0.001

P(fire) = 0.001/0.20 = **0.005**

Probability of radiation leak in case of a mechanical failure = 50% = 0.50

Probability of a radiation leak occurring simultaneously with  
 a mechanical failure = 0.15% = 0.0015

P(mf) = 0.0015/0.50 = **0.003**

Probability of radiation leak in case of a human error = 10% = 0.10

Probability of a radiation leak occurring simultaneously with  
 a human error = 0.12% = 0.0012

P(he) = 0.0012/0.10 = **0.012**

* 1. **What is the probability of a radiation leak?**

**Solution**: Probability of a radiation leak P(RL) = P(fire + mechanical failure + human error)

**P(RL) = 0.001 + 0.0015 + 0.0013 = 0.0037**

* 1. **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

**Solution**: P(fire/RL) = P(fire&RL) / P(RL) = 0.001/0.0037 = **0.270**

P(mf/RL) = P(mf&RL) / P(RL) = 0.0015/0.0037 = **0.4054**

P(he/RL) = P(he&RL) / P(RL) = 0.0012/0.0037 = **0.3243**

# Problem 3:

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; (Provide an appropriate visual representation of your answers, without which marks will be deducted)

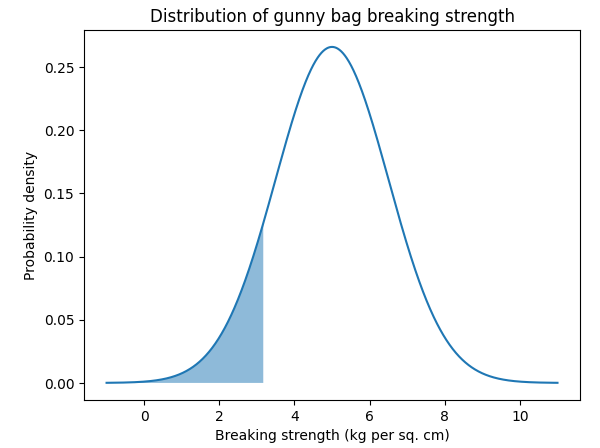
* 1. **What proportion of the gunny bags have a breaking strength less than 3.17 kg per sq. cm?**

**Solution**: Mean = 5

Standard deviation = 1.5

Gunny bag strength (x) = 3.17

cdf value = 0.1112

****

**~11.12% of the gunny bags have a breaking strength less than 3.17 kg per sq. cm**

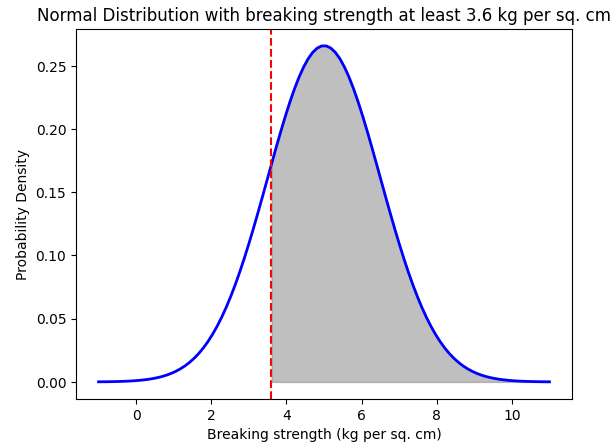
* 1. **What proportion of the gunny bags have a breaking strength at least 3.6 kg per sq. cm?**

**Solution**: Mean = 5

Standard deviation = 1.5

x = 3.6

cdf value = 0.8246

****

**~82.46% of the gunny bags have a breaking strength of at least 3.6 kg per sq. cm**

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

**Solution**: Mean (mu) = 5

Standard deviation (sd)= 1.5

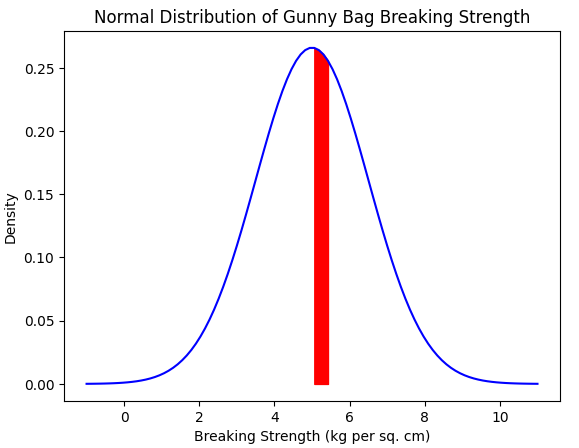
x1 = 5

x2 = 5.5

z1(z-value) = x1-mu/sd = (5-5)/1.5 = 0

z2(z-value) = x2-mu/sd = (5.5-5)/1.5 = 0.33

cdf value = 0.1304

****

**~13.04% of the gunny bags have a breaking strength between 5 and 5.5 kg per sq. cm**

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

**Solution**: Mean (mu) = 5

Standard deviation (sd)= 1.5

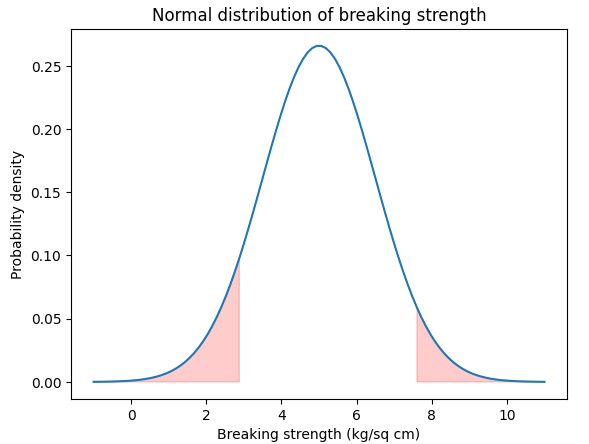
x1 = 3

x2 = 7.5

z1(z-value) = x1-mu/sd = (3-5)/1.5 = -1.333

z2(z-value) = x2-mu/sd = (7.5-5)/1.5 = 1.666

cdf value = z1 + (1 - z2) = **0.1390**



**~13.90% of the gunny bags does not have a breaking strength between 3 and 7.5 kg per sq cm**

# 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.

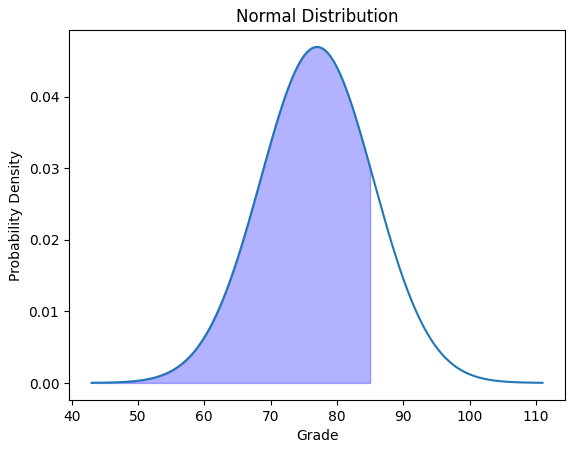
* 1. **What is the probability that a randomly chosen student gets a grade below 85 on this exam?**

**Solution**: Mean = 77

Standard deviation = 8.5

X = 85

cdf value = 0.8266

****

**Probability that a randomly chosen student getting a grade below 85 on the exam is ~82.66%**

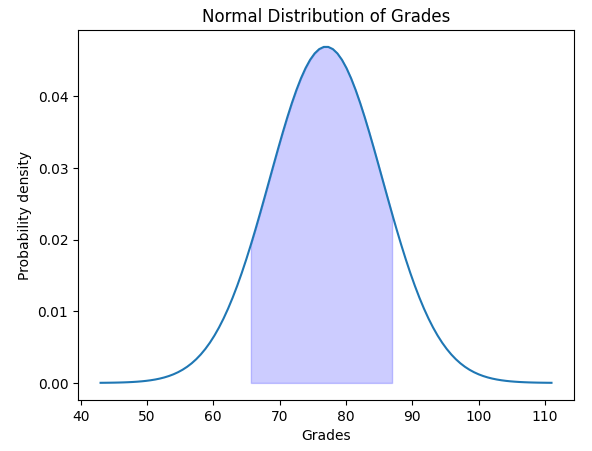
* 1. **What is the probability that a randomly selected student scores between 65 and 87?**

**Solution**: Mean = 77

Standard deviation = 8.5

X1 = 65  
 X2 = 87  
 Calculated z-score for both values

cdf value = 0.8012

****

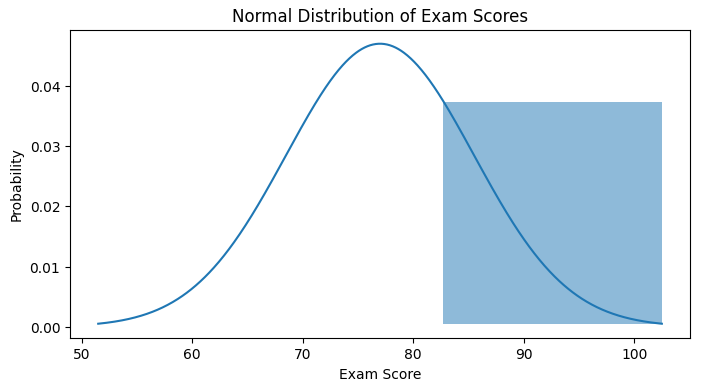
**~80.12% is the probability that a randomly selected student scores between 65 and 87**

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

**Solution**: Mean = 77

Standard deviation = 8.5

Z\_score = 75

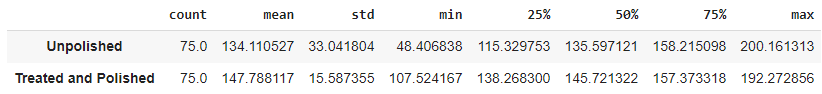
**** cdf value = 0.8273

**~82.73** **should be the passing cut-off so that 75% of the students clear the exam**

# 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 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. **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?**

**Solution**:

Sample value = 75

α , level of significance = 0.05

Defining Null and alternative hypotheses

***Ho :* µ ≥ 150**  
 Null hypothesis : Unpolished stone are suitable for printing

***Ha :* µ < 150**  
 Alternative hypothesis : Unpolished stone are not suitable for printing



**P-Value here is less than α, hence, we reject the null hypothesis. Unpolished stones may not be suitable for printing.**

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

**Solution**: Sample value = 75

α , level of significance = 0.05

Defining Null and alternative hypotheses

***Ho :* µ1 = µ2**  
 Null hypothesis : Average hardness of polished stone = Average hardness of  
 unpolished stone

 ***Ha :* µ1 ≠ µ2**  
 Alternative hypothesis : Hardness of polished and unpolished stone are not equal

We did 2-sample t-test and found the p-value =

**P-Value here is less than α, hence,** **we have enough evidence to reject the null hypothesis in favour of alternative hypothesis. We conclude that the Hardness of polished and unpolished stone are not equal**

# 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 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.

**Solution**: Sample = 100

Degree of Freedom = 99

α , level of significance = 0.05

Defining Null and alternative hypotheses

***Ho :* µa(after) - µb (before) ≤ 5**  
 Null hypothesis : The candidate is not able to do more than 5 push-ups compared  
 to he/she enrolled in program

***Ha :* µa(after) - µb (before) > 5**  
 Alternative hypothesis : The candidate is able to do more than 5 push ups compared  
 to he/she enrolled in program



**P-Value here is less than α, hence,** **we have enough evidence to reject the null hypothesis in favour of alternative hypothesis. The candidate is able to do more than 5 push-ups compared to when he/she enrolled in program, Thus, making program successful.**

# 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.

* 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.?**

**Solution**: Defining Null and alternative hypotheses  
 **Hypothesis for Alloy 1**  
 ***Ho1 :*** The mean implant hardness for dentists are equal for alloy 1  
 ***Ha1 :*** The mean implant hardness for at least one dentist is not equal for alloy 1

**Hypothesis for Alloy 2**  
 ***Ho2 :*** The mean implant hardness for dentists are equal for alloy 2  
 ***Ha2 :*** The mean implant hardness for at least one dentist is not equal for alloy 2

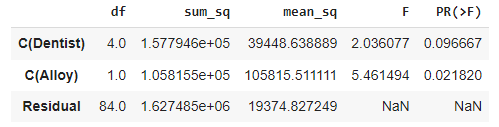
* 1. **Before the hypotheses may be tested, state the required assumptions. Are the assumptions fulfilled? Comment separately on both alloy types?**

**Solution**: Assumptions:

* Dependent variables should be measured at continuous level
* Two independent variables should each consist of 2 or more categorical independent group
* There should be no significant outliers
* Dependent variable should be approx. normally distributed for each combination of the groups of the two independent variables.

* 1. **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?**

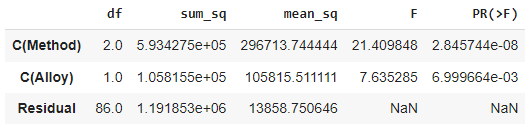
**Solution**:

****

**Since P-value is greater than 0.05 we fail to reject null hypothesis. Thus, we can say the implant hardness is not dependent on dentists**

* 1. **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?**

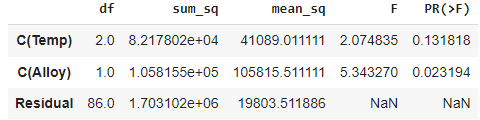
**Solution**:

****

**Since the p- value is less than 0.05 the null hypothesis is rejected and therefore the implant hardness is dependent on the methods used**

* 1. **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?**

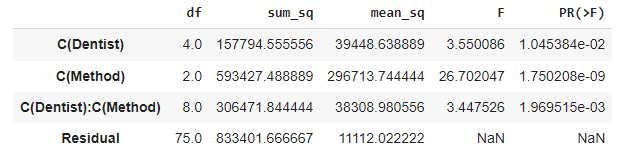
**Solution**:

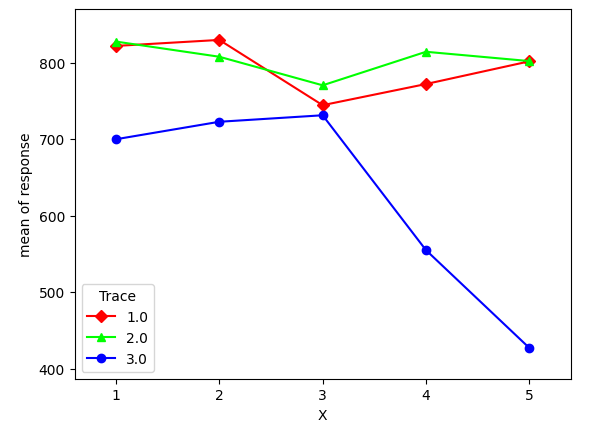
****

**Since P-value is greater than 0.05 we fail to reject null hypothesis. Thus, we can say the implant hardness is not dependent on the temperature**

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

**Solution**:

****

****

**P-value is less than significance value, there are some interactions for Dentist and Method, so making Method as a significant cause in response in both the alloy**

**Thank you**