**Advance Statistics Project**

Student Name : Suneel Kumar Pentapalli

Date : 29/04/2023



Contents

[***Problem 1*** 3](#_Toc134690879)

[***Problem 2*** 6](#_Toc134690880)

[***Problem 3:*** 9](#_Toc134690881)

[***Problem 4:*** 14](#_Toc134690882)

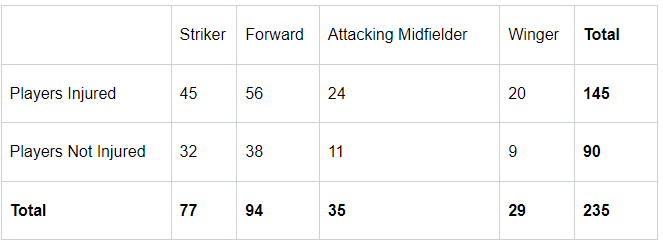
[***Problem 5:*** 16](#_Toc134690883)

[***Problem 6:*** 19](#_Toc134690884)

[***Problem 7:*** 20](#_Toc134690885)

# 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**

****

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

The probability of any event to happen can be defined as ratio between m(number of possible favorable outcomes) and n(total number of possible outcomes)

P=m/n

From the given sample,

Total number of possible outcomes is nothing but Total number of players(n) and total favorable outcomes is the number of players injured(m)

Total number of players(n) = 235

Total number of players injured(m) = 145

Hence, the probability that a randomly chosen player to suffer injury =

number of players injured(m)/total number of players

P=145/235= 0.617

**i.e., there is 61.7% chance that the randomly chosen player would suffer injury**

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

Similar to above, lets find out the m and n for this scenario.

Number of players playing in forward position = 94

Number of players playing at winger position = 29

And total number of players(n)=235

For the given condition i.e., the selected player can only be forward or winger at a time but not both at once. This event is called as mutually exclusive.

Hence we will apply addition rule for two mutually exclusive events which is

P(A U B)= P(A) + P(B)

where P(A) is the probability that a randomly picked player plays at forward position=94/235

P(B) is the probability that a randomly picked player plays at winger position = 29/235

Hence, the probability that a randomly picked player is either a forward or a winger =(94/235)+(29/235) =52.3%

**i.e., there is 52.3% chance that the randomly chosen player is either 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?**

P(player in striker position has foot injury)=

From given sample,

Total number of players in striker position with injury(m) = 45

Total number of players(n) = 235

P(player in striker position has foot injury)=45/235 = 0.1914

**i.e., there is 19.14% chance that the randomly chosen player is in a striker position and has a foot injury**

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

P(chosen injured player is a striker)=

From given sample,

Total number of striker players who are injured(m) = 45

Total number of injured players(n) = 145

P(chosen inured player is a striker)=45/145 = 0.3103

**i.e., there is 31.03% chance that the randomly chosen injured player is in a striker position**

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

Again we are applying addition rule between two mutually exclusive events here i.,e P(A U B)= P(A) + P(B)

P(chosen injured player is either a forward or an attacking midfielder)=

++

=+=0.5517

**i.e., there is 55.17% chance that the 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:**

The given problem statement can be solved by using conditional probability, which says

“The probability of an event A based on the occurrence of another event B is termed conditional Probability. It is denoted as P(A|B) and represents the probability of A when event B has already happened.”

P(A|B) = P(A∩B) / P(B)

Where, P(A ∣ B) is the conditional probability of event A occurring, given that B is true,

P(A∩B) is the probability of P(A) and P(B) occurring together,

P(A) and P(B) are the probabilities of A and B occurring independently of one another.

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

Let, F, M, HE, RL be the events as follows,

F = event of fire

M = event of mechanical failure

HE = event of human error and

RL = event of radiation leak

In the given statement,

Probability of Radiation Leak due to Fire(F) is P(RL/F)=20%

Probability of Radiation Leak due to Mechanical Failure(MF) is P(RL/M)=50%

Probability of Radiation Leak due to Human Error(HE) is P(RL/HE)=10%

Also given,

Probability of radiation occurring simultaneously with fire is P(RL ∩ F)=0.1%

Probability of radiation occurring simultaneously with mechanical failure is P(RL ∩ M)=0.15%

Probability of radiation occurring simultaneously with human error is P(RL ∩ HE)=0.12%

Now according to conditional probability, we have

P(RL/F) = P(RL ∩ F)/P(F)

P(RL/M) = P(RL ∩ M)/P(M)

P(RL/HE) = P(RL ∩ HE)/P(HE)

Therefore, Probability of file, P(F)= P(RL ∩ F)/P(RL/F)=(0.1/100)/(20/100)\*100=0.5 %

Similarly,

Probability of mechanical failure= P(M)= P(RL ∩ M)/P(RL/M)=(0.15/100)/(50/100)\*100=0.3 %

Probability of human error= P(HE)= P(RL ∩ HE)/P(RL/HE)=(0.12/100)/(10/100)\*100=1.2 %

**Hence, Probabilities of fire, mechanical failure and human error are 0.5%, 0.3% and 1.2% respectively.**

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

Let, F, M, HE, RL be the events as follows,

F = event of fire

M = event of mechanical failure

HE = event of human error and

RL = event of radiation leak

There are only three reasons that cause radiation leak in this scenario, they are fire, mechanical failure and Human error.

So the probability of radiation leak P(R)= P(RL ∩ F) + P(RL ∩ M) + P(RL ∩ HE) for which we have values as below,

P(RL ∩ F)=0.1/100

P(RL ∩ M)=0.15/100

P(RL ∩ HE)=0.12/100

P(RL)=0.001+0.0015+0.0012=0.0037

**Hence, Probability of radiation leak is 0.37%, which is pretty low.**

**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:**

We have Probability of radiation leak P(RL)=0.0037 and

P(RL ∩ F)=0.1/100

P(RL ∩ M)=0.15/100

P(RL ∩ HE)=0.12/100

**A Fire.**

P(F/RL)=P(F ∩ RL)/P(RL) = 0.001/0.0037 = 0.270

**Hence probability that given radiation leak is caused by fire is 27%**

**A Mechanical Failure.**

P(M/RL)=P(M ∩ RL)/P(RL) = 0.0015/0.0037 = 0.405

**Hence probability that given radiation leak is caused by Mechanical Failure is 40.5%**

**A Human Error.**

P(HE/RL)=P(HE ∩ RL)/P(RL) = 0.0012/0.0037 = 0.324

**Hence probability that given radiation leak is caused by Human Error is 32.4%**

# 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)

Given,

Population mean(mu)=5

Population standard deviation(sigma)=1.5

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

Population mean(mu)=5

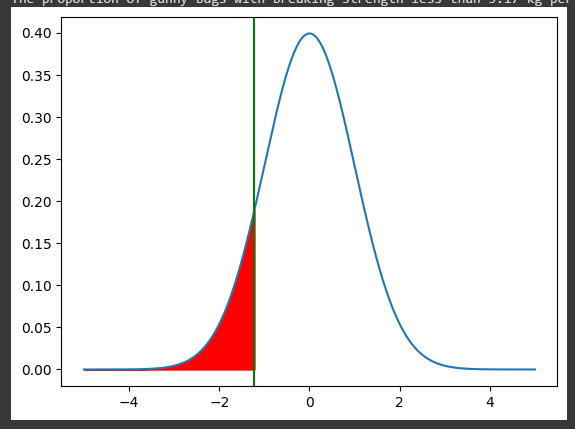
Population standard deviation(sigma)=1.5

The proportion of gunny bags having breaking strength less than 3.17 kg per sq cm can be calculated using cdf

Round(Stats.norm.cdf(3.17,mu,sigma)\*100,2)=11.12

**i.e., 11.12 % of gunny bags have breaking strength less than 3.17 kg per sq cm**

here is the graph representing the same,



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

Population mean(mu)=5

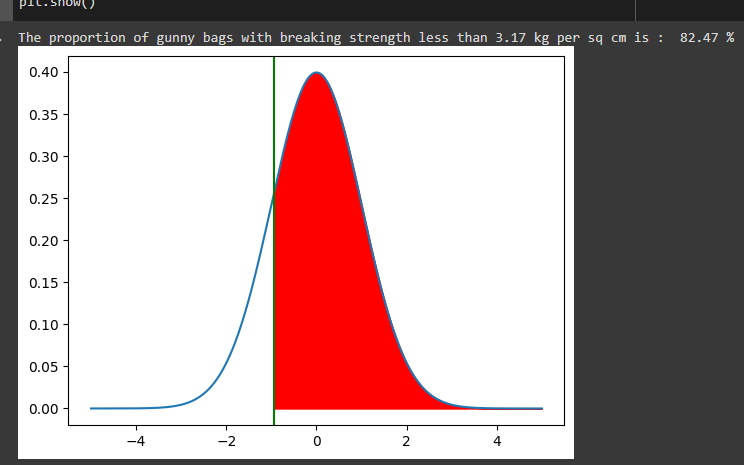
Population standard deviation(sigma)=1.5

The proportion of gunny bags having breaking strength atleast 3.6 kg per sq cm can be calculated using cdf

1 - Round(Stats.norm.cdf(3.6,mu,sigma)\*100,2)=82.47

**i.e., 82.47 % of gunny bags have breaking strength of atleast 3.6 kg per sq cm**

Here is the graph representing the same,



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

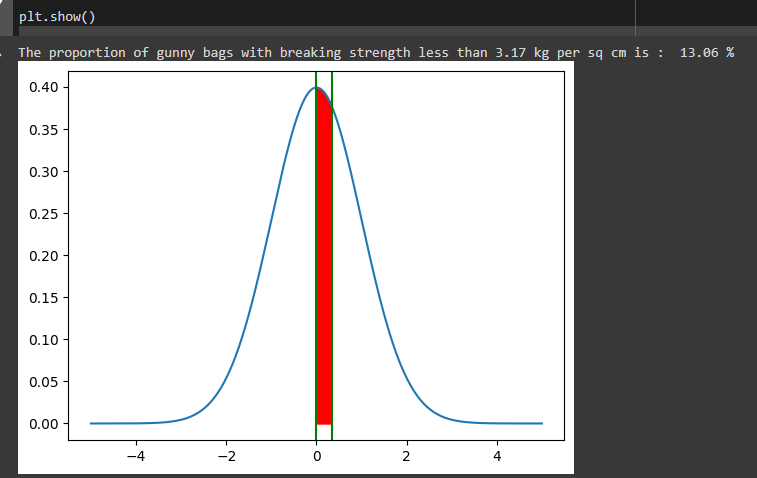
The proportion of gunny bags having breaking strength between 5 and 5.5 kg per sq cm can also be calculated using cdf as below

(Proportion of gunny bags with strength 5.5 kg per sq cm) **minus (**proportion of gunny bags with strength 5 kg per sq cm)

round(((stats.norm.cdf(5.5,mu,sigma))-(stats.norm.cdf(5,mu,sigma)))\*100,2) = 13.06

**i.e., 13.06 % of gunny bags have breaking strength between 5 and 5.5 kg per sq cm**

Here is the graph representing the same,



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

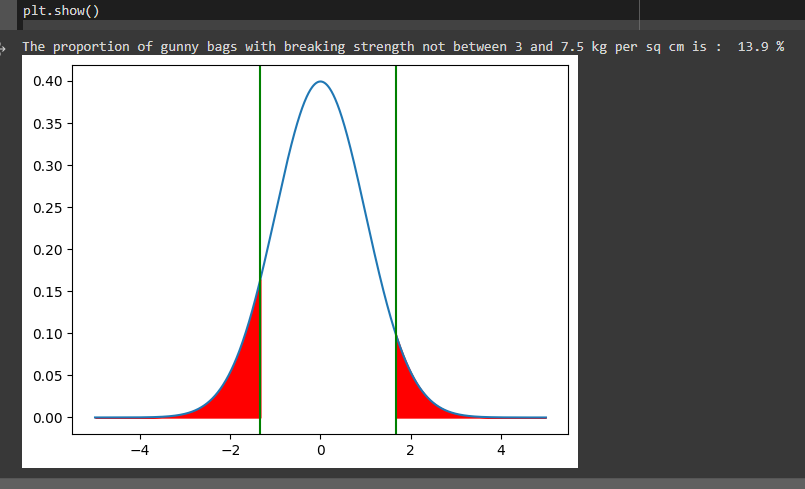
The proportion of gunny bags having breaking strength not between 3 and 7.5 kg per sq cm can also be calculated using cdf as below

(Proportion of gunny bags with strength 3 kg per sq cm) **plus (**proportion of gunny bags with strength of atleast 7.5 kg per sq cm)

round(((1-stats.norm.cdf(7.5,mu,sigma))+(stats.norm.cdf(3,mu,sigma)))\*100,2) = 13.9

**i.e., 13.9 % of gunny bags have breaking strength not between 3 and 7.5 kg per sq cm**

Here is the graph representing the same,



# 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.**

(Proportion of gunny bags with strength 3 kg per sq cm) **plus (**proportion of gunny bags with strength of atleast 7.5 kg per sq cm)

Population mean(mu) = 77

Population standard deviation(sigma) = 8.5

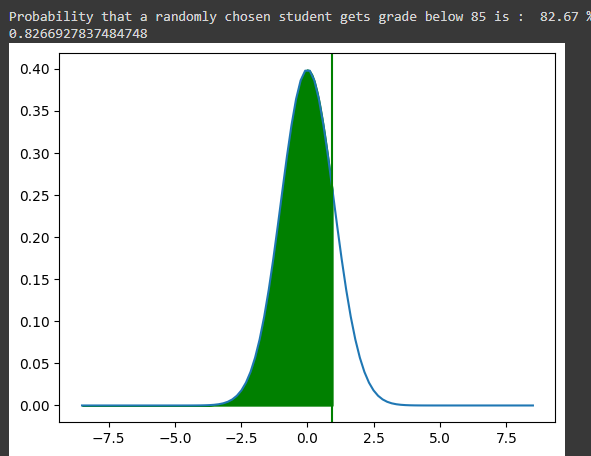
**4.1 What is the probability that a randomly chosen student gets a grade below 85 on this exam?**

Probability that a randomly chosen student gets a grade below 85 =

Stats.norm.cdf(85,loc=mu,scale=sigma)= 82.67 %

**i.e., there is 82.67 % chance that a randomly chosen student gets a grade below 85.**

Here is the graphical representation of the same.



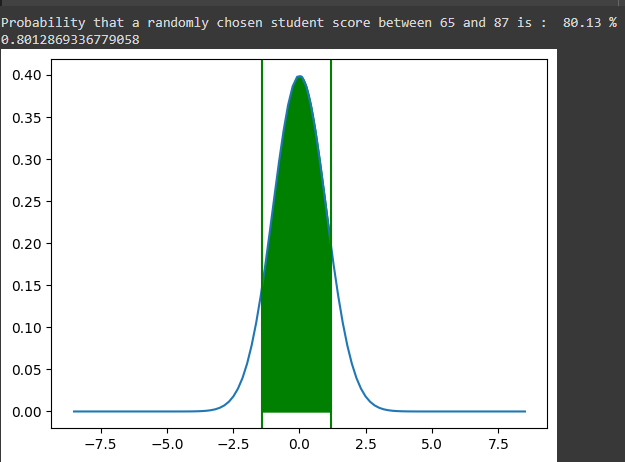
**4.2 What is the probability that a randomly selected student scores between 65 and 87?**

Probability that a randomly chosen student gets a grade between 65 and 87 =

stats.norm.cdf(87,loc=mu,scale=sigma)-stats.norm.cdf(65,loc=mu,scale=sigma) = 80.13 %

**i.e., there is 80.13 % chance that a randomly selected student scores between 65 and 87.**

Here is the graphical representation of the same.



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

#75 % students should clear the exam, i.e., the distribution above 25%,

Passing cut-off so that 75 % of the students clear the exam is = stats.norm.ppf(.25,loc=mu,scale=sigma) = 71.26

**i.e., to get 75% students passed in the exam, the cut off should be 71.26.**

# 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);**

Given values from above statement,

Alpha = 0.05

Required Brinell hardness = 150

**5.1 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?**

For the stone to be eligible/favorable for printing, minimum brinell’s hardness should be 150.

Hence, lets consider this as our null hypothesis,

H0= Brinell’s hardness >= 150

Ha= Brinell’s hardness < 150

Lets test H0 using given data,

Here, we consider sample of unpolished stones alone for our hypothesis test.

**Step 1:** Define null and alternative hypotheses

Let null hypothesis(H0) be mean hardness >= 150

and alternate hypothesis(Ha) be mean hardness < 150

**Step 2:** Decide the significance level

This is already given in problem statement, alpha = 0.05

**Step 3:** Identify the test statistic

We have one samples and we do not know the population standard deviation.

Hence we use t distribution and tstat test statistic.

**Step 4:** Calculate p\_value and t\_statistic

UsingSuneel\_Kumar\_30\_April\_2023\_AS.ipynb to calculate p\_value and t\_statistic

t\_statistic = -4.164629601426757

p\_value = 8.342573994839304e-05

**As p\_value is less than our significance level, we reject the null hypothesis which says mean hardness >= 150. i.e., Zingaro’s assumption is correct that the unpolished stones are not suitable for printing.**

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

Here we have two samples i.e., sample of polished stones and sample of unpolished ones.

**Step 1:** Define null and alternative hypotheses

Let H0 be Mean hardness of Unpolished = polished

Let Ha be Mean hardness of Unpolished <> polished

**Step 2:** Decide the significance level

Given, alpha = 0.05

**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.

Hence we find t stat using two sample paired test.



Where, t is test statistic,

X1 bar = sample1(polished stones) mean

X2 bar = sample2(unpolished stones) mean

S1= sample variance of polished stones

S2 = sample variance of unpolished stones

N1 = sample size of polished stones

N2 = sample size of unpolished stones

But we can find out t\_statistic and p\_value directly using ttest\_ind in python(please check Suneel\_Kumar\_30\_April\_2023\_AS.ipynb for reference)

Upon executing ttest\_ind on given two samples i.e., “Treated and Polished” and “Unpolished” stones

We get, t\_statistic = 3.2422320501414053

p\_value= 0.0014655150194628353

**As p\_value(0.0015) is less than the given significance level(0.05), hence we reject null hypothesis which means, the mean hardness of “Treated and Polished” stones is different from 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 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.**

**Step 1:** Define null and alternative hypotheses

Let Null hypothesis(H0) be µafter- µbefore >= 6 (more than 5 days)

Let Alternate hypothesis(Ha) be µafter- µbefore < 6

**Step 2:** Define significance level,

From the given problem statement,

alpha = 0.05

**Step 3:** Identify the test statistic

As the sample size of both the pairs is same and also we don’t know the population standard deviation, hence lets consider using t distribution and calculate tstat using two sample paired test

**Step 4:** Calculate t\_statistic and p\_value

UsingSuneel\_Kumar\_30\_April\_2023\_AS.ipynb to calculate p\_value and t\_statistic

t\_statistic = -19.322619811082458

p\_value = 2.2920419252511966e-35

**We can reject null hypothesis as p\_value is less than our given significance level(0.05). i.e., The program conducted by Aquarius health club is unsuccessful as the candidate is not able to perform more than 5 push-ups when compared to number of push-ups before joining 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.**

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.?**

UsingSuneel\_Kumar\_30\_April\_2023\_AS.ipynb to calculate p\_value and t\_statistic

Lets consider two datasets corresponding to each alloy. i.e., one dataset for alloy type 1 and another dataset with alloy type 2.

**For Alloy1:**

**Step 1:** Define null and alternate hypothesis

Null hypothesis (H0) = mean implant hardness is **same** for the patients treated by different dentists.

Alternate hypothesis (Ha) = mean implant hardness is **not same** for the patients treated by different dentists.

**Step 2:** Assume significance level to be 0.05

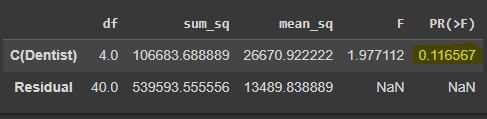
**Step 3:** Define formula and fit out dataframe for alloy\_1(in .ipynb file, I have created two df from the given dataset based on alloy types)

Formula = ‘Response ~ C(Dentist)’

model1=ols(formula,alloy\_1).fit()

anova\_tbl\_for\_alloy\_1=anova\_lm(model1)

anova\_tbl\_for\_alloy\_1



From the above output, p value is 0.116567, its clearly greater than given significance level(0.05).

**Hence we cannot reject null hypothesis which says “mean hardness is same for the patients treated by different dentists”**

**For Alloy2:**

**Step 1:** Define null and alternate hypothesis

Null hypothesis (H0) = mean implant hardness is **same** for the patients treated by different dentists.

Alternate hypothesis (Ha) = mean implant hardness is **not same** for the patients treated by different dentists.

**Step 2:** Assume significance level to be 0.05

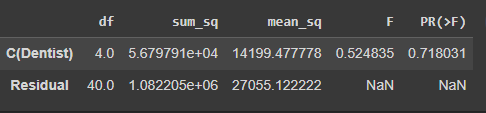
**Step 3:** Define formula and fit out dataframe for alloy\_2(in .ipynb file, I have created two df from the given dataset based on alloy types)

Formula = ‘Response ~ C(Dentist)’

Model2=ols(formula,alloy\_2).fit()

anova\_tbl\_for\_alloy\_2=anova\_lm(model2)

anova\_tbl\_for\_alloy\_2



From the above output, p value is 0.718031, its clearly greater than given significance level(0.05).

**Hence we cannot reject null hypothesis which says “mean hardness is same for the patients treated by different dentists”**

**From the above, its clear that, irrespective of alloy types, mean hardness has no impact when a patient is treated by different dentists.**

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

Assumptions of one way ANOVA are as follows:

**The dependent variable needs to be continuous.**

For both alloy 1 and alloy 2, the dependent variable is continuous.

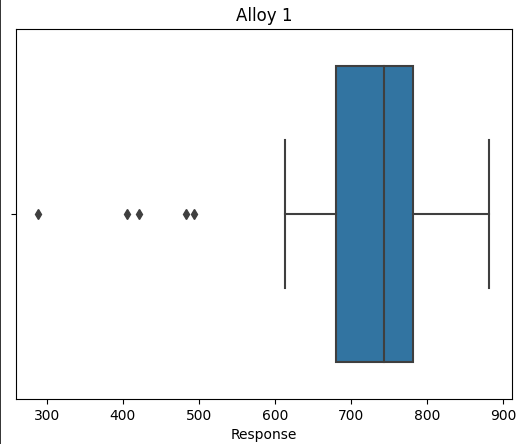
**The dependent variable should be approximately normally distributed.**

For both alloy 1 and alloy2, the dependent variable when checked using Anderson darling method, it is observed that the “random” variable is not distributed normally.

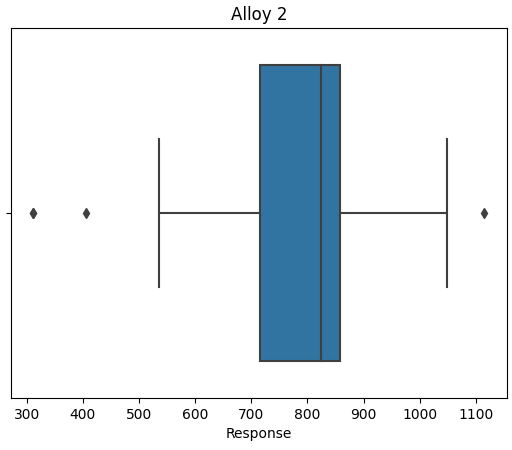
The statistic value from Anderson test for both alloy types is greater than all the values in critical\_value list.

**There should be no outliers in the dependent variable.**

**Alloy 1**



**Alloy 2**



As shown above, there are outliers in the dependent variable(Response) in both the alloy types.

**The independent variable should have atleast 2 or more categorical groups.**

The independent variable ‘Dentist’ has 5 categories hence this assumption holds true.

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?**

From the test performed in step 1, it is clear that the p value is higher than significance level, hence we cannot reject null hypothesis.

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?**

**For Alloy1:**

**Step 1:** Define null and alternate hypothesis

Null hypothesis (H0) = mean implant hardness is **same** across all methods.

Alternate hypothesis (Ha) = mean implant hardness is not **same** across all methods.

**Step 2:** Assume significance level to be 0.05

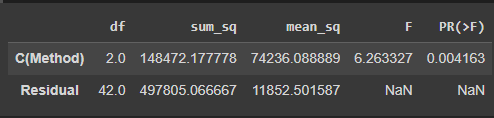
**Step 3:** Define formula and fit out dataframe for alloy\_1(in .ipynb file, I have created two df from the given dataset based on alloy types)

formula='Response ~ C(Method)'

alloy1\_method\_model=ols(formula,alloy\_1).fit()

anova\_method\_response=anova\_lm(alloy1\_method\_model)

anova\_method\_response:

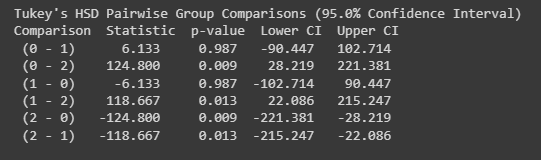


From the above output, p value is 0.004163 which is clearly less than the given significance level(0.05)

The average amount of variation between groups is greater than that within groups

**Hence we can reject null hypothesis i.e., “For implants made with alloy 1, mean hardness is not same for the patients treated using different methods”**

To check which group has different means, we can perform tukey\_hsd tst from scipy.stats



**As per output from tukey\_hsd test, for methods 1-3 and 2-3 the means are different for alloy 1**

**For Alloy2:**

**Step 1:** Define null and alternate hypothesis

Null hypothesis (H0) = mean implant hardness is **same** across all methods.

Alternate hypothesis (Ha) = mean implant hardness is not **same** across all methods.

**Step 2:** Assume significance level to be 0.05

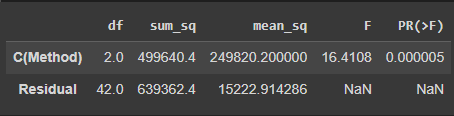
**Step 3:** Define formula and fit out dataframe for alloy\_1(in .ipynb file, I have created two df from the given dataset based on alloy types)

formula='Response ~ C(Method)'

alloy2\_method\_model=ols(formula,alloy\_2).fit()

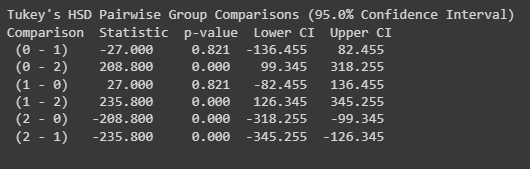
anova\_method\_response=anova\_lm(alloy2\_method\_model)

anova\_method\_response:



From the above output, p value is 0.000005 which is clearly less than the given significance level(0.05)

**Hence we can reject null hypothesis i.e., “For implants made with alloy 2, mean hardness is not same for the patients treated using different methods”**



**As per output from tukey\_hsd test, for methods 1-3 and 2-3 the means are different for alloy 2**

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?**

**For Alloy1:**

**Step 1:** Define null and alternate hypothesis

Null hypothesis (H0) = mean implant hardness is **same** across all temperature levels.

Alternate hypothesis (Ha) = mean implant hardness is **not same** across all temperature levels

**Step 2:** Assume significance level to be 0.05

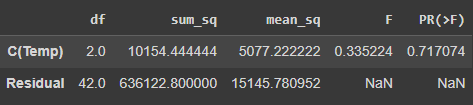
**Step 3:** Define formula and fit out dataframe for alloy\_1(in .ipynb file, I have created two df from the given dataset based on alloy types)

formula = 'Response ~ C(Temp)'

model\_alloy1\_temp=ols(formula,alloy\_1).fit()

anova\_alloy1\_temp= anova\_lm(model\_alloy1\_temp)

anova\_alloy1\_temp



From the above output, p value is 0.717074 which is clearly greater than the given significance level(0.05)

**Hence we cannot reject null hypothesis i.e., “For implants made with alloy 1, mean hardness is same for the patients treated at different temperature levels”**

**For Alloy2:**

**Step 1:** Define null and alternate hypothesis

Null hypothesis (H0) = mean implant hardness is **same** across all temperature levels.

Alternate hypothesis (Ha) = mean implant hardness is **not same** across all temperature levels

**Step 2:** Assume significance level to be 0.05

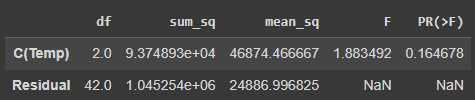
**Step 3:** Define formula and fit out dataframe for alloy\_1(in .ipynb file, I have created two df from the given dataset based on alloy types)

formula = 'Response ~ C(Temp)'

model\_alloy2\_temp=ols(formula,alloy\_2).fit()

anova\_alloy2\_temp= anova\_lm(model\_alloy2\_temp)

anova\_alloy2\_temp

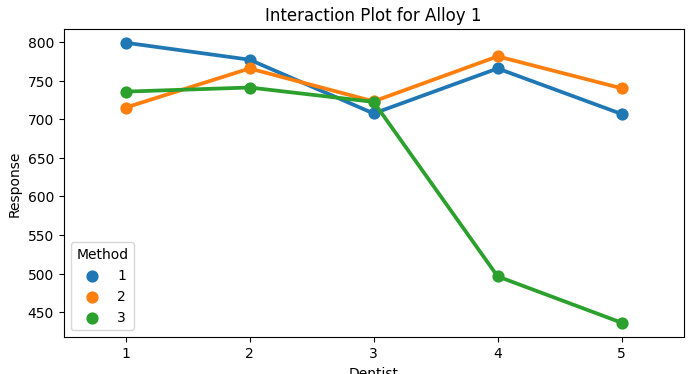


From the above output, p value is 0.164678 which is clearly greater than the given significance level(0.05)

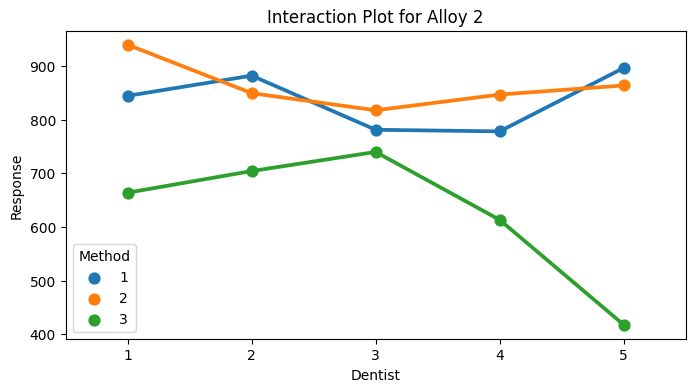
**Hence we cannot reject null hypothesis i.e., “For implants made with alloy 2, mean hardness is same for the patients treated at different temperature levels”**

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

Interaction plot for alloy 1:

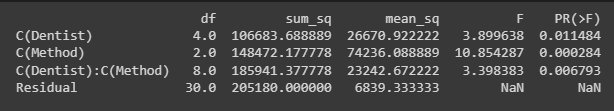


Interaction plot for alloy 2:



1. **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?**

**For Alloy 1:**



As p value of interaction items is less than our significance level(0.05), we not reject null hypothesis. i.e., for alloy 1 , there is significant interaction between Dentist and Method terms.

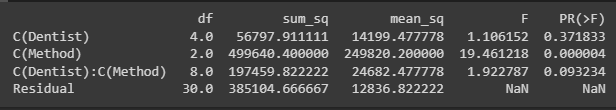
Due to the inclusion of the interaction effect term, we can see a slight change in the p-value of the first two treatments as compared to the Two-Way ANOVA without the interaction effect terms. And we see that the p-value of the interaction effect term of 'Dentist and 'Method' suggests that the Null Hypothesis is rejected in this case

Same can be seen in interaction graph,

For methods 1 and 2 for dentists(3,4,5), there is no much interaction between dentist and method as the interactions lines are pretty much parallel.

For rest of the methods and dentists there is significant interaction as the lines are intersecting as well.

**For Alloy 2:**



As p value of interaction items is greater than our significance level(0.05), we cannot reject null hypothesis. i.e., for alloy 2 , there is no significant interaction between Dentist and Method terms.