Business Report

Table of contents-

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**

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

Answer- Probability = Total number of players injured / Total number of players

(Marginal probability)

Probability = 145/235 = 0.617

Probability that a randomly chosen player would suffer an injury is **0.617**

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

Answer- P (A or B) = P (A) + P (B)

Probability = 94/235 + 29/235 = 0.4 + 0.123 = 0.5234

Probability that player is a forward or a winger **0.5234**

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

Answer- Probability (Striker and Has a Foot injury) = P (Striker) * P (Has a foot injury |Striker)

(Joint Probability)

Probability = 77/235 * 45/77 = 45/235 = 0.1914

Probability that randomly chosen player plays in a striker position and has a foot injury **= 0.191**

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

Answer- P (Striker | Injured) = P (Striker and injury) /P (injured)

(Conditional Probability)

Probability = (45/235)/ (145/235) = 45/145 = 0.310

Probability that randomly chosen injured player is a striker **= 0.310**

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

Answer- P (Forward| Injured) = P (Forward and injury) /P (injured)

Probability = (56/235)/ (145/235) = 56/145

P (Attacking midfielder | Injured) = P (Attacking midfielder and injury) /P (injured)

Probability = (24/235)/ (145/235) = 24/145

Probability that a randomly chosen injured player is either a forward or an attacking midfielder = P (Forward| Injured) + P (Attacking midfielder | Injured)

Probability = 56/145 + 24/145 = 0.5517

Probability that a randomly chosen injured player is either a forward or an attacking midfielder = **0.5517**

Problem 2: 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%.

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

Probability (Event) = Favorable Outcomes/Total Outcomes

The probability of fire = P (Fire)/ (P (Fire) + P (Mechanical) + P (Human Error))

= 20/ (20+50+10) = **0.25%**

The probability of mechanical failure = 50/80 **= 0.625%**

The probability of human error = 10/80 **= 0.125%**

As per the question-

The probability of a radiation leak occurring simultaneously with a fire is given = 0.1%.

The probability of a radiation leak occurring simultaneously with a mechanical failure is given = 0.15%.

The probability of a radiation leak occurring simultaneously with a human error is given = 0.12%

2.2 What is the probability of a radiation leak?

Probability (Radiation leakage) = 0.1+0.15+0.12 **= 0.37%**

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.

P (Fire) = 0.1/0.37 = **0.27%**

P (Mechanical failure) = 0.15/0.37= **0.40%**

P (Human error) = 0.12/0.37 = **0.32%**

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.

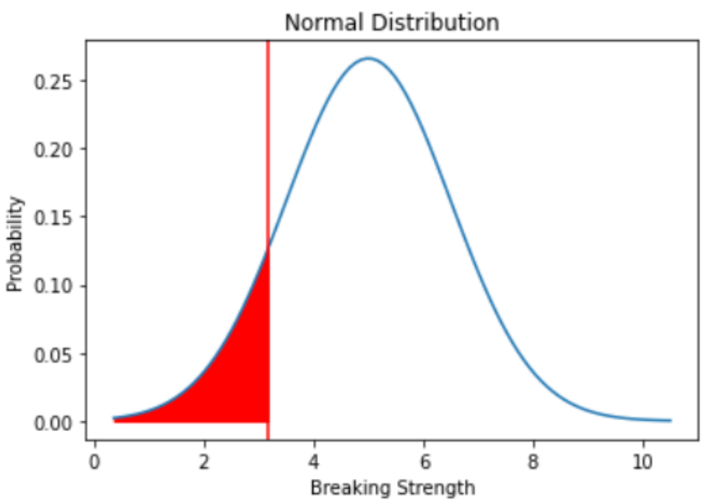
Mean = **5 kg per sq. centimeter**

Standard deviation = **1.5 kg per sq. centimeter**

The breaking strength of gunny bags used for packaging cement is **normally distributed**

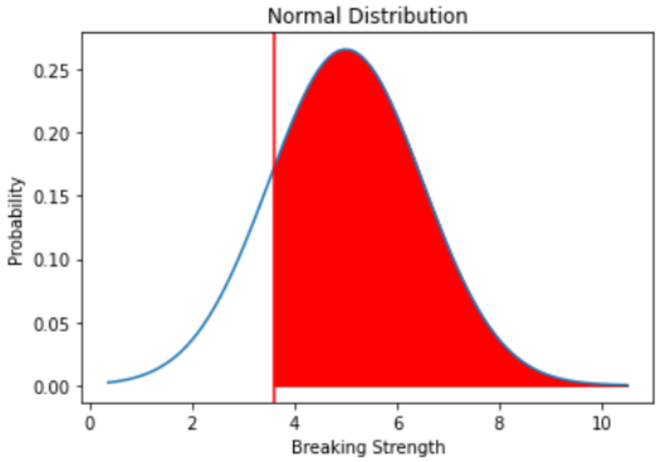
3.1 What proportion of the gunny bags have a breaking strength less than 3.17 kg per sq.cm?

Proportion of the gunny bags have a breaking strength < 3.17 kg per sq. cm = 0.1112



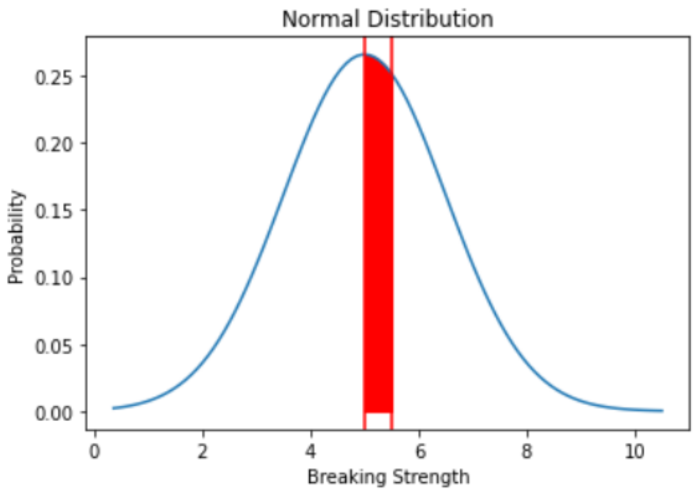
3.2 What proportion of the gunny bags have a breaking strength at least 3.6 kg per sq. cm?

Proportion of the gunny bags have a breaking strength at least 3.6 kg per sq. cm. = 0.8247



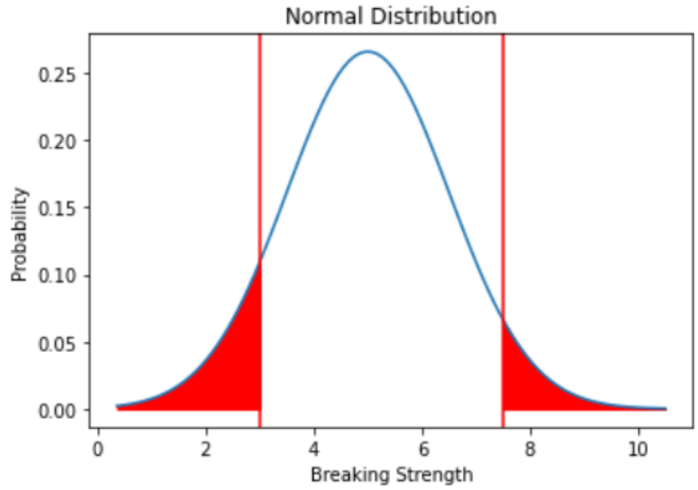
3.3 What proportion of the gunny bags have a breaking strength between 5 and 5.5 kg per sq. cm?

Proportion of the gunny bags have a breaking strength between 5 and 5.5 kg per sq. cm. = 0.1306



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

Proportion of the gunny bags have a breaking strength NOT between 3 and 7.5 kg per sq. cm. = 0.139

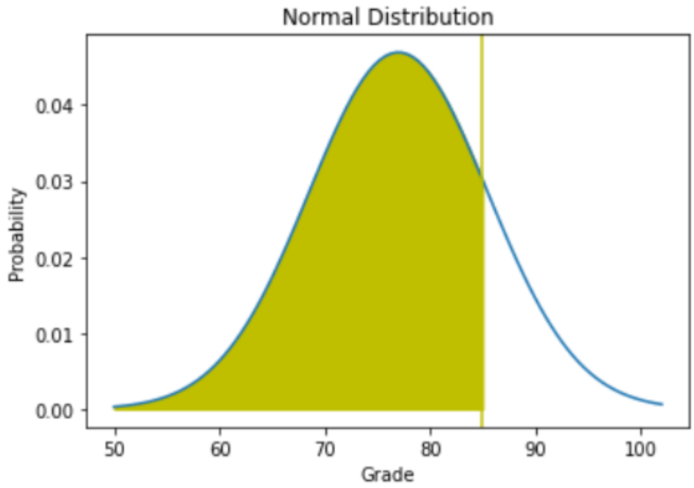


Problem 4:

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.

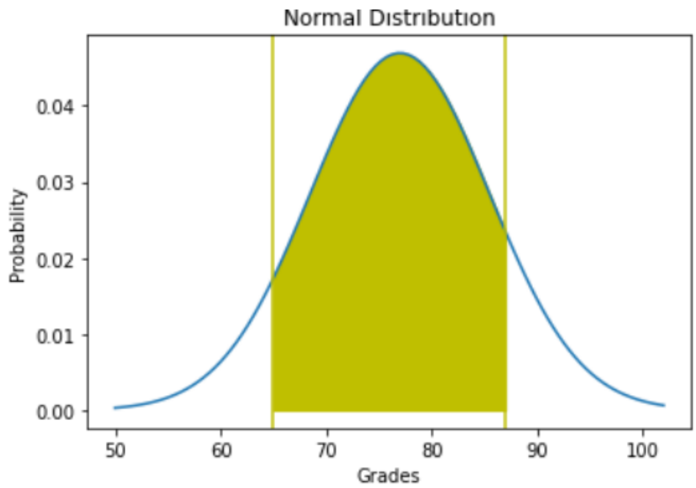
4.1 What is the probability that a randomly chosen student gets a grade below 85 on this exam?

Probability that a randomly chosen student gets a grade below 85 on this exam = 0.8267



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

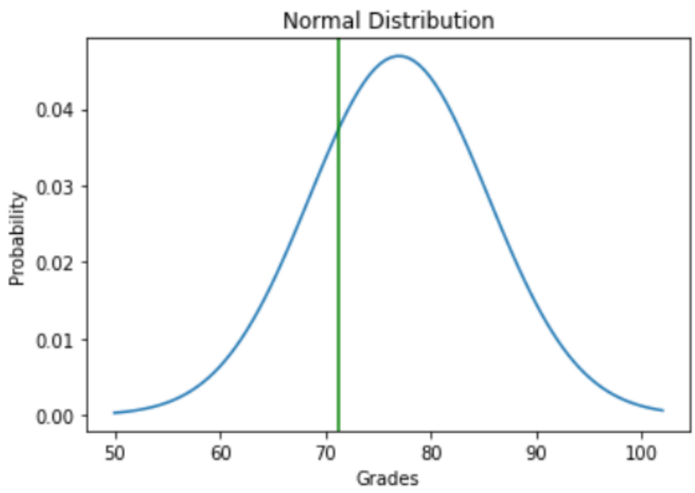
Probability that a randomly selected student scores between 65 and 87= 0.8013



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

Passing cut-off so that 75% of the students clear the exam= 71.26

71.26 should be the cutoff , if we want that only 75% of the students shall 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.

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?

We will check for unpolished stones-

Null Hypothesis : Brinell's hardness index of Unpolished stones = 150

Alternate Hypothesis : Brinell's hardness index of Unpolished stones < 150

Alpha=0.05

We will perform left tailed one sample t-test-

We get result as-

t statistic: -4.164629601426758 p value: 4.1712869974196425e-05

At 5% significance level pvalue (4.1712869974196425e-05) < 0.05

So we reject the null hypothesis.

We do have statistical evidence to say that Brinell's hardness index of Unpolished stones is significantly less than 150

Zingaro is right about the believe that the unpolished stones are not suitable for printing.

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

We will perform 2 sample 2 tailed t-test

Null Hypothesis : mean hardness of the polished = mean hardness of the unpolished stones

Altermnate Hypothesis : mean hardness of the polished is not equal to mean hardness of the unpolished stones

Alpha=0.05

We get results as-

tstat= 3.242232050141406

P Value= 0.001465515019462831

At 5% significance level pvalue (0.001465515019462831) < 0.05

So we reject the null hypothesis.

We do have statistical evidence to say that the mean hardness of the polished stones is significantly different from 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.

We will perform right tailed paired t-test

Difference= Pushups of candidate after program - Pushups of candidate before program

Null Hypothesis : Difference <= 5

Alternate Hypothesis : Difference > 5

Alpha= 0.05

Results that we get are-

tstat -1.915

p-value for one-tail: 0.029198872141011155

At 5% significance level

pvalue (0.029198872141011155) < 0.05

So we reject the null hypothesis.

That is after the program.

We have enough statistical evidence to say candidates are able to do more than 5 push-ups as compared to when he/she enrolled in the program

Program is 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.

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.?

Solution- Alloy 1

We are considering the case when dentists are using **Alloy=1**

The null hypothesis states that the mean implant hardness is equal among all the Dentists.

the mean implant hardness for each group(dentists) is equal

H0 : $\mu1=\mu2=\mu3=\mu4=\mu5$ (the mean implant hardness for each group(dentists) is equal)

HA : Not all the means are equal (At least one group mean is different from the rest)

Alpha=0.05

We will perform ANOVA to test the hypothesis-

We get-

	df	sum_sq	mean_sq	F	PR(>F)
C(Dentist)	4.0	106683.688889	26670.922222	1.977112	0.116567
Residual	40.0	539593.555556	13489.838889	NaN	NaN

As p-value (0.116) > alpha(0.05)

We failed to reject the null hypothesis

This means we don't have sufficient evidence to say that there is a statistically significant difference between the mean implant hardness of the Five groups(Dentists).

Solution- Alloy 2

We are considering the case when dentists are using **Alloy=2**

The null hypothesis states that the mean implant hardness is equal among all the Dentists.

H0 : μ1=μ2=μ3=μ4=μ5 (the mean implant hardness for each group(dentists) is equal)

HA : Not all the means are equal (At least one group mean is different from the rest)

We will perform ANOVA to test the hypothesis-

We get-

	df	sum_sq	mean_sq	F	PR(>F)
C(Dentist)	4.0	5.679791e+04	14199.477778	0.524835	0.718031
Residual	40.0	1.082205e+06	27055.122222	NaN	NaN

As p-value (0.718) > alpha(0.05)

We failed to reject the null hypothesis

This means we don't have sufficient evidence to say that there is a statistically significant difference between the mean implant hardness of the Five groups(Dentists).

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

Assumptions-

1. The sample drawn from different populations are independent and random.
2. The response variable of all the populations are continous and ideally normally distributed.
3. The variance of all the populations are equal atleast approximately.

Assumptions are not full filled-

For Alloy 1 - Anderson-Darling Normality Test gives following results-

```
AndersonResult(statistic=2.561066309273194, critical_values=array([0.535, 0.609, 0.731, 0.853, 1.014]),
significance_level=array([15. , 10. ,  5. ,  2.5,  1. ]))
```

The test statistic is 2.56 We can compare this value to each critical value that corresponds to each significance level to see if the test results are significant.

The critical value for α = 0.025 is 0.853. Because the test statistic (2.56) is greater than this critical value, the results are significant at a significance level of 0.025.

Same is the case with all the other values-

We can see that the test results are significant at every significance level, which means we would reject the null hypothesis of the test no matter which significance level we choose to use. Thus, we have sufficient evidence to say that the sample data is not normally distributed.

For Alloy 2 - Anderson-Darling Normality Test gives following results-

```
AndersonResult(statistic=1.8931311726356412, critical_values=array([0.535, 0.609, 0.731, 0.853, 1.014]),
significance_level=array([15. , 10. ,  5. ,  2.5,  1. ]))
```

The test statistic is 1.89 We can compare this value to each critical value that corresponds to each significance level to see if the test results are significant.

The critical value for α = 0.025 is 0.853. Because the test statistic (1.89) is greater than this critical value, the results are significant at a significance level of 0.025.

Same is the case with all the other values-

We can see that the test results are significant at every significance level, which means we would reject the null hypothesis of the test no matter which significance level we choose to use. Thus, we have sufficient evidence to say that the sample data is not normally distributed.

- Both the distributions are not normally distributed , which voilates the assumptions.
- Considerable presence of out liars could also be seen for both alloys, and ANOVA is sensitive to Outliars.

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

H0 : $\mu1=\mu2=\mu3=\mu4=\mu5$ (the mean implant hardness for each group(dentists) is equal)

HA : Not all the means are equal (At least one group mean is different from the rest)

Alpha= 0.05

Alloy 1- As p-value (0.116) > alpha(0.05)

Alloy 2- As p-value (0.718) > alpha(0.05)

We failed to reject the null hypothesis for both the alloys .

This means we don't have sufficient evidence to say that there is a statistically significant difference between the mean implant hardness of the Five groups(Dentists) for both alloys(1 and 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?

Solution- Alloy=1

We are considering the case when dentists are using **Alloy=1**

H0 : μ1=μ2=μ3 (the mean implant hardness for each group(Method) is equal)

HA : Not all the means are equal (At least one group mean is different from the rest)

Alpha= 0.05

We will perform ANOVA to test the hypothesis-

We get-

	df	sum_sq	mean_sq	F	PR(>F)
C(Method)	2.0	148472.177778	74236.088889	6.263327	0.004163
Residual	42.0	497805.066667	11852.501587	NaN	NaN

As p-value (0.004163) < alpha(0.05)

We reject the null hypothesis.

This means we have sufficient evidence to say that there is a statistically significant difference between the mean implant hardness of the three groups(Methods)

Solution- Alloy=2

We are considering the case when dentists are using **Alloy=2**

H0 : μ1=μ2=μ3 (the mean implant hardness for each group(Method) is equal)

HA : Not all the means are equal (At least one group mean is different from the rest)

Alpha= 0.05

We will perform ANOVA to test the hypothesis-

We get-

	df	sum_sq	mean_sq	F	PR(>F)
C(Method)	2.0	499640.4	249820.200000	16.4108	0.000005
Residual	42.0	639362.4	15222.914286	NaN	NaN

As p-value (0.000005) < alpha(0.05)

We reject the null hypothesis.

This means we have sufficient evidence to say that there is a statistically significant difference between the mean implant hardness of the three groups(Methods)

Since the p value is less than the significance level(0.05), we can reject the null hypothesis and conclude that there is a difference in the mean implant hardness. mean implant hardness is different for at-least one category of methods for both the alloys.

As we have rejected the null hypothesis for both the alloys , lets check for which methods the mean implant hardness is significantly different.

We can use Tukey's multiple comparison test for this-

For Alloy 1 -

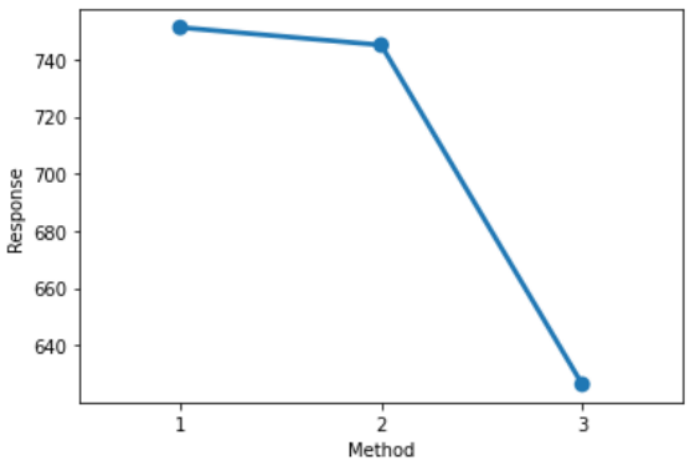
Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
1	2	-6.1333	0.9	-102.7105	90.4438	False
1	3	-124.8	0.0085	-221.3771	-28.2229	True
2	3	-118.6667	0.0128	-215.2438	-22.0895	True

Mean difference for method 3 is quite high when compared with both methods 1 & 2.

Tukey's HSD Test for multiple comparisons found that the mean implant hardness was significantly different between Method 1 and Method 3 it is different for Method 2 and Method 3 as well.

It can also be seen from below plot-



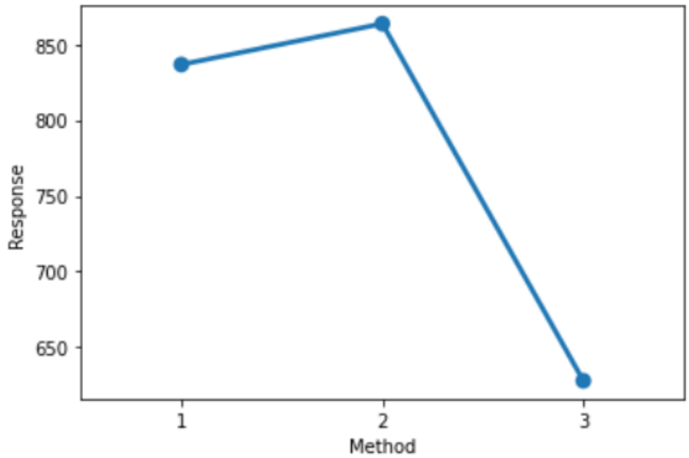
For Alloy 2-

```
 Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====================================================
group1 group2 meandiff p-adj   lower     upper   reject
-----------------------------------------------------
     1      2     27.0 0.8046  -82.4506  136.4506  False
     1      3   -208.8  0.001 -318.2506  -99.3494   True
     2      3   -235.8  0.001 -345.2506 -126.3494   True
-----------------------------------------------------
```

Mean difference for method 3 is quite high when compared with both methods 1 & 2.

Tukey's HSD Test for multiple comparisons found that the mean implant hardness was significantly different between Method 1 and Method 3 it is different for Method 2 and Method 3 as well.

It can also be seen from below plot-



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?

Solution- Alloy=1

We are considering the case when dentists are using **Alloy=1**

H0 : µ1=µ2=µ3 (the mean implant hardness for each group(Temperature) is equal)

HA : Not all the means are equal (At least one group mean is different from the rest)

Alpha= 0.05

We will perform ANOVA to test the hypothesis-

We get-

	df	sum_sq	mean_sq	F	PR(>F)
C(Temp)	2.0	10154.444444	5077.222222	0.335224	0.717074
Residual	42.0	636122.800000	15145.780952	NaN	NaN

As p-value (0.7170) > alpha(0.05)

We failed to reject the null hypothesis

This means we don't have sufficient evidence to say that there is a statistically significant difference between the mean implant hardness of the three groups(Temprature).

Solution- Alloy=2

We are considering the case when dentists are using **Alloy=2**

H0 : $\mu1=\mu2=\mu3$ (the mean implant hardness for each group(Temperature) is equal)

HA : Not all the means are equal (At least one group mean is different from the rest)

Alpha= 0.05

We will perform ANOVA to test the hypothesis-

We get-

	df	sum_sq	mean_sq	F	PR(>F)
C(Temp)	2.0	9.374893e+04	46874.466667	1.883492	0.164678
Residual	42.0	1.045254e+06	24886.996825	NaN	NaN

As p-value (0.164) > alpha(0.05)

We failed to reject the null hypothesis

This means we don't have sufficient evidence to say that there is a statistically significant difference between the mean implant hardness of the three groups(Temprature).

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

Solution Alloy 1

Lets state the hypothesis (**Alloy 1**)

H0- Interection between Dentist and Method does not exist.

HA- Interection between Dentist and Method exists.

Alpha=0.05

We get-

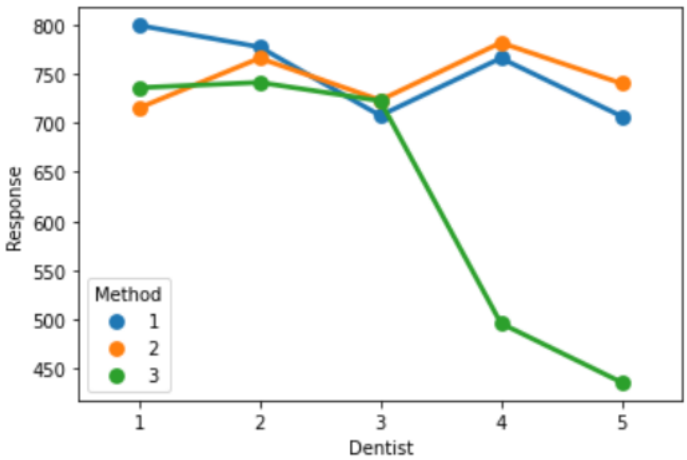
	df	sum_sq	mean_sq	F	PR(>F)
C(Method)	2.0	148472.177778	74236.088889	10.854287	0.000284
C(Dentist)	4.0	106683.688889	26670.922222	3.899638	0.011484
C(Method):C(Dentist)	8.0	185941.377778	23242.672222	3.398383	0.006793
Residual	30.0	205180.000000	6839.333333	NaN	NaN

As p-value (0.0067) < alpha(0.05)

We reject the null hypothesis.

There is statistically significant evedence to say that there is interection between Dentist and Method feature.

We can also see that once we are taking interection of method and dentist features into consideration, this also changes the way denstist column can impact mean implant hardness.



It could be noticed that there is a good interection between dentist and method feature.

Solution Alloy 2

Lets state the hypothesis (**Alloy 2**)

H0- Interection between Dentist and Method does not exist.

HA- Interection between Dentist and Method exists.

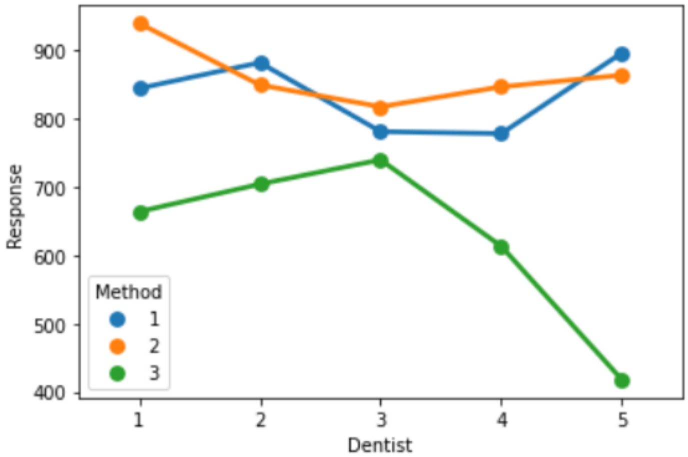
Alpha=0.05

We get-

	df	sum_sq	mean_sq	F	PR(>F)
C(Method)	2.0	499640.400000	249820.200000	19.461218	0.000004
C(Dentist)	4.0	56797.911111	14199.477778	1.106152	0.371833
C(Method):C(Dentist)	8.0	197459.822222	24682.477778	1.922787	0.093234
Residual	30.0	385104.666667	12836.822222	NaN	NaN

As p-value (0.09) > alpha(0.05)

We failed to reject the null hypothesis.There is no significant interection between Dentist and Method considering only alloy 2.



There is no as such significant interection between method and dentist feature for alloy 2.

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?

Solution Alloy 1

We are considering the case when dentists are using **Alloy=1**

We have to consider 2 hypothesis one for method and one for dentist.

H0 : $\mu1=\mu2=\mu3$ (the mean implant hardness for each group(Method) is equal)

HA : Not all the means are equal (At least one group mean is different from the rest)

Alpha= 0.05

H0 : $\mu1=\mu2=\mu3$ (the mean implant hardness for each group(Dentist) is equal)

HA : Not all the means are equal (At least one group mean is different from the rest)

Alpha= 0.05

	df	sum_sq	mean_sq	F	PR(>F)
C(Method)	2.0	148472.177778	74236.088889	7.212522	0.002211
C(Dentist)	4.0	106683.688889	26670.922222	2.591255	0.051875
Residual	38.0	391121.377778	10292.667836	NaN	NaN

As p-value (0.002) < alpha(0.05) for method feature

We reject the null hypothesis

This means we have sufficient evidence to say that there is a statistically significant difference between the mean implant hardness of the three groups(Methods) for alloy 1.

As p-value (0.051) > alpha(0.05) for dentist feature

We fail to reject the null hypothesis

This means we don't have sufficient evidence to say that there is a statistically significant difference between the mean implant hardness of the five groups(Dentist) for alloy 1.

Solution Alloy 2

We are considering the case when dentists are using **Alloy=2**

We have to consider 2 hypothesis one for method and one for dentist.

H0 : $\mu1=\mu2=\mu3$ (the mean implant hardness for each group(Method) is equal)

HA : Not all the means are equal (At least one group mean is different from the rest)

Alpha= 0.05

H0 : $\mu1=\mu2=\mu3$ (the mean implant hardness for each group(Dentist) is equal)

HA : Not all the means are equal (At least one group mean is different from the rest)

Alpha= 0.05

	df	sum_sq	mean_sq	F	PR(>F)
C(Method)	2.0	499640.400000	249820.200000	16.295479	0.000008
C(Dentist)	4.0	56797.911111	14199.477778	0.926215	0.458933
Residual	38.0	582564.488889	15330.644444	NaN	NaN

As p-value (0.000008) < alpha(0.05)

We reject the null hypothesis

This means we have sufficient evidence to say that there is a statistically significant difference between the mean implant hardness of the three groups(Methods) for alloy 2.

As p-value (0.45) > alpha(0.05)

We fail to reject the null hypothesis

This means we don't have sufficient evidence to say that there is a statistically significant difference between the mean implant hardness of the five groups(Dentist) for alloy 2.

Its not possible to identify which dentists are different as mean implant hardness is equal among all the Dentists for both the alloys.

Its possible to identify which methids are different using

We can use Tukey's multiple comparison test for this-

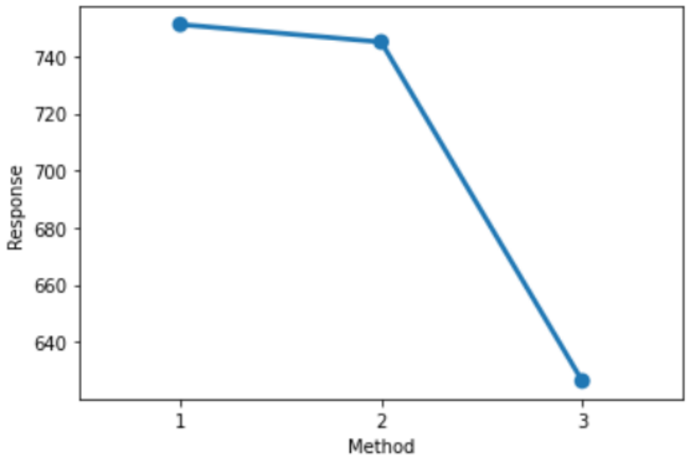
For Alloy 1 –

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
1	2	-6.1333	0.9	-102.7105	90.4438	False
1	3	-124.8	0.0085	-221.3771	-28.2229	True
2	3	-118.6667	0.0128	-215.2438	-22.0895	True

Mean difference for method 3 is quite high when compared with both methods 1 & 2.

It can also be seen from below plot-



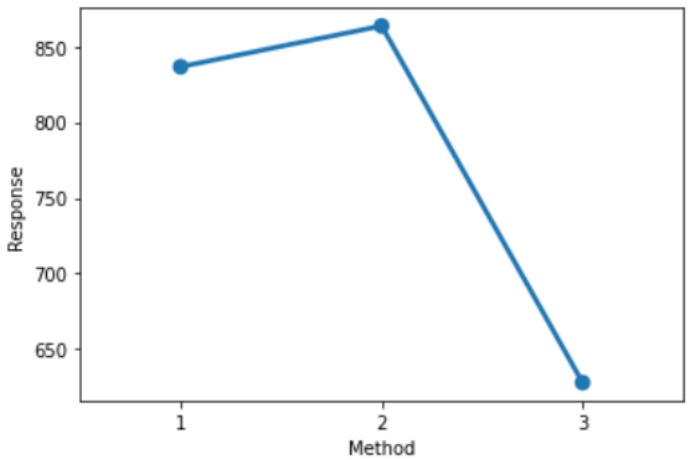
For Alloy 2-

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
1	2	27.0	0.8046	-82.4506	136.4506	False
1	3	-208.8	0.001	-318.2506	-99.3494	True
2	3	-235.8	0.001	-345.2506	-126.3494	True

Mean difference for method 3 is quite high when compared with both methods 1 & 2.

It can also be seen from below plot-



To identify which interection levels are different , we have to check for all the interections possible-

For Alloy 1

Considering alpha= 0.05

Interection	p-value
Dentist- Method	0.006793
Dentist- Temp	0.862862
Method –Temp	0.898357

Only the interection between Dentist and Method is significant. Other then that there is no significant interection between variables for alloy 1.

For Alloy 2

Considering alpha= 0.05

Interection	p-value
Dentist- Method	0.093234
Dentist- Temp	0.825318
Method –Temp	0.675983

There is no significant interection between variables for alloy 2.

END