AS Project Report

Preetam Sarmah

DSBA

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**

Probability = Sum of favourable outcomes / Total outcomes

1.1 Probability that a randomly chosen player would suffer an injury?

P = Number of injured players/ Total number of Players

P = 0.617

1.2 Probability that a player is a forward or a winger?

P = (Number forwards + Number of wingers) / Total number of Players

P = 0.5234

1.3 Probability that a player plays in a striker position and has a foot injury??

P = Number of Injured Strikers / Total number of Players

P = 0.1915

1.4 Probability that a randomly chosen injured player is a striker?

P = (Number of Injured Strikers) / Total number of Injured Players

P = 0.3103

1.5 Probability that a injured player is either a forward or an attacking midfielder?

P = (Number Injured forwards + Number Injured attacking midfielder) / Total number of Injured Players

P = 0.5517

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%.

Probability = Sum of favourable outcomes / Total outcomes

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

P_{fire} = 0.250

$P_{\text{mechanical failure}}$ = 0.625

$P_{\text{human error}}$ = 0.125

2.2 What is the probability of a radiation leak?

$P_{\text{radiation leak}} = 0.0037$

2.3 What is the probability that radiation leak has been caused by fire ,mechanical failure, and a human error Respectively?

$P_{\text{radiation leak} \because \text{fire}} = 0.2703$

$P_{\text{radiation leak} \because \text{mechanical failure}} = 0.4054$

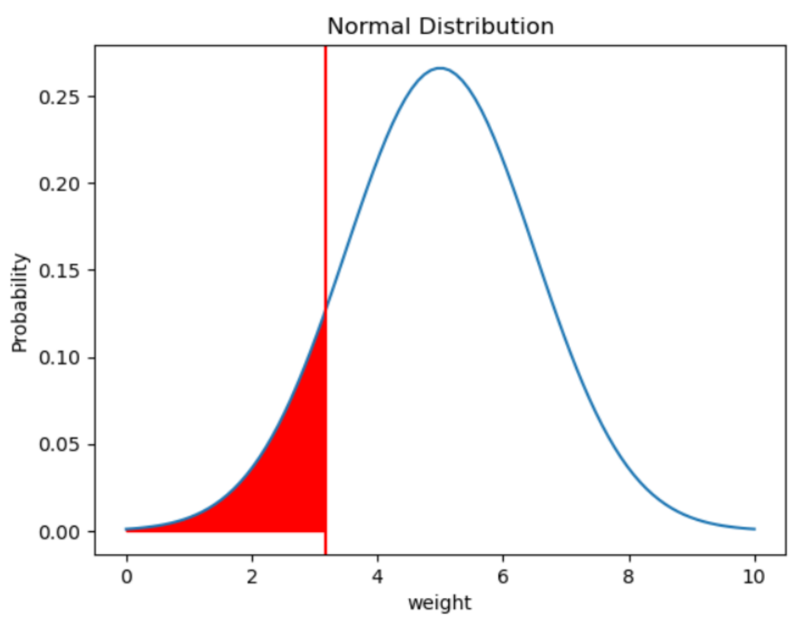
$P_{\text{radiation leak} \because \text{human error}} = 0.3243$

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.

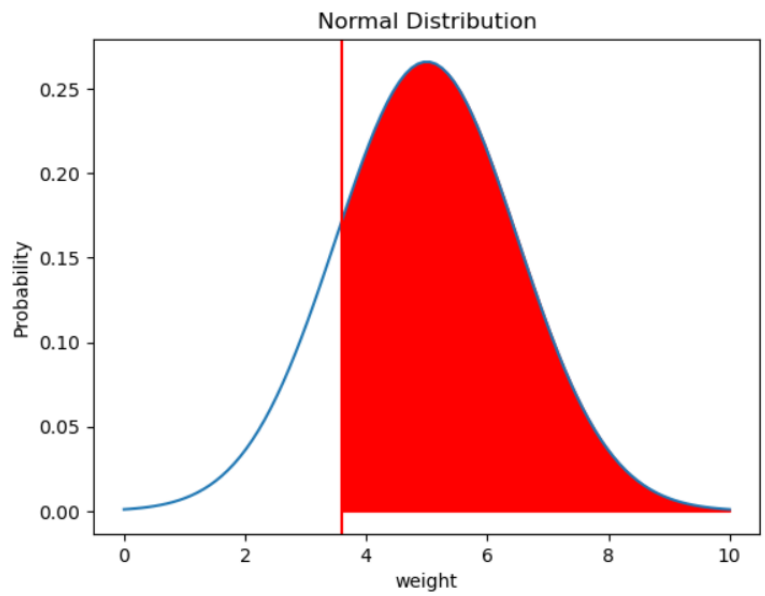
3.1 What proportion of the gunny bags have a breaking strength less than 3.17 kg per sq cm?

A) p = 0.1112



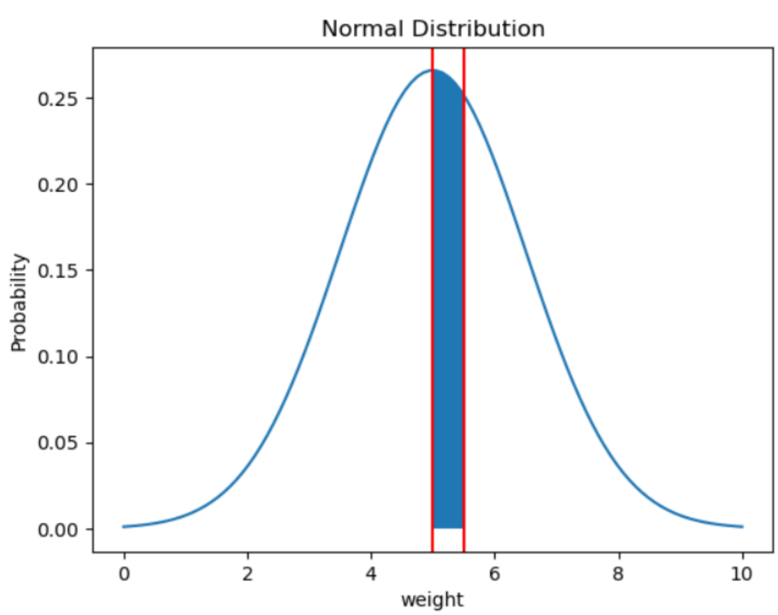
3.2 What proportion of the gunny bags have a breaking strength at least 3.6 kg per sq cm.?

A) p = 0.8247



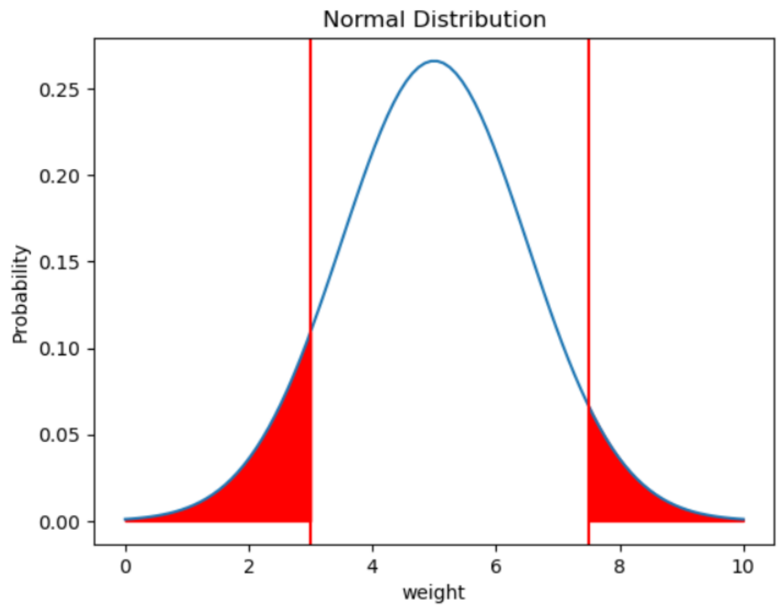
3.3 What proportion of the gunny bags have breaking strength between 5 and 5.5 kg per sq cm.?

A) p = 0.1306



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

A) p = 0.139

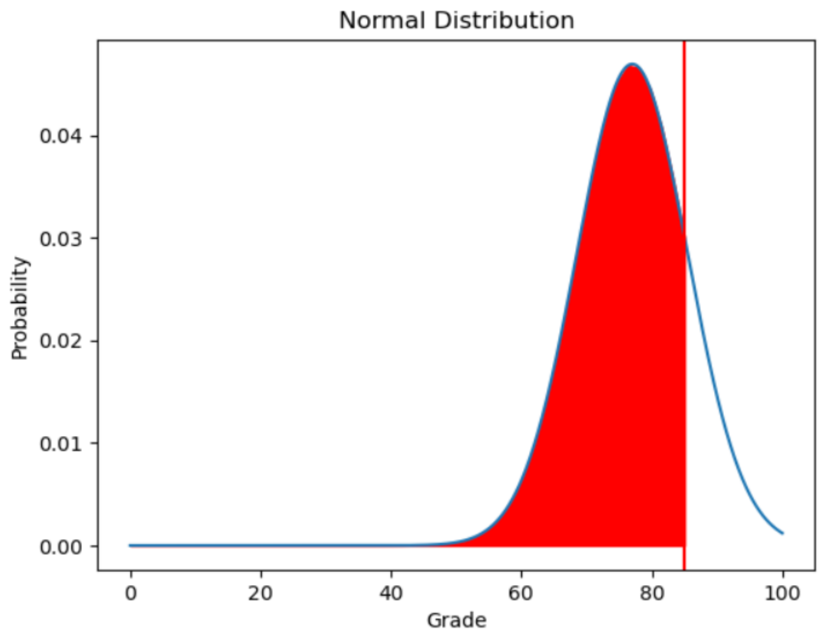


Problem 4

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. Based on the given information answer the questions below.

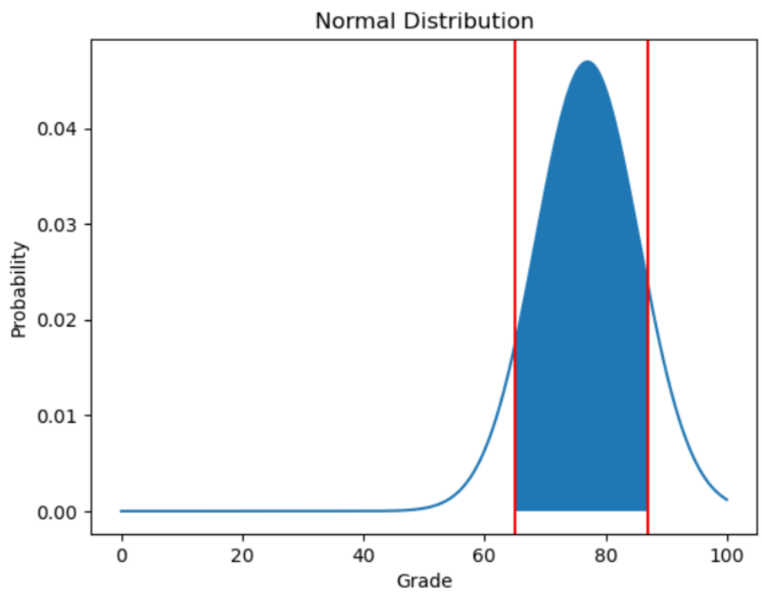
4.1 What is the probability that a randomly chosen student gets a grade below 85 on this exam?

A) p = 0.8267



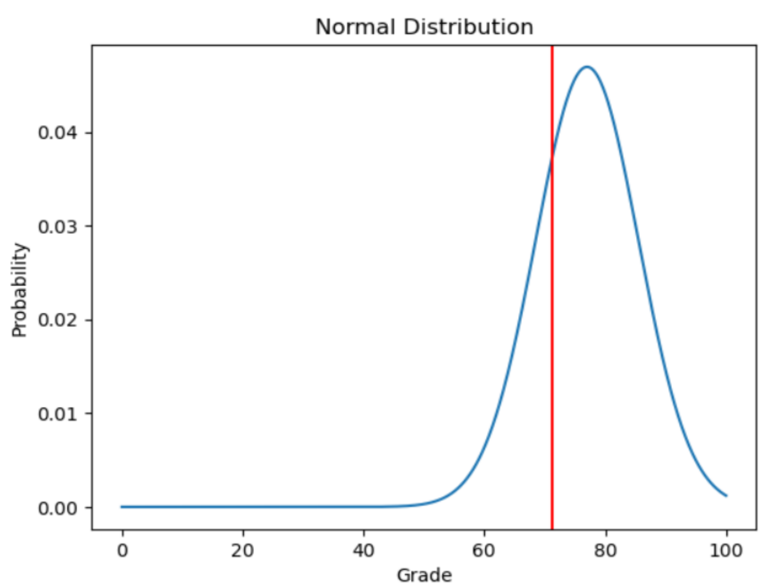
4.2 What is the probability that a randomly selected student scores between 65 and 87?

A) p = 0.8013



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

A) Cut off Grade = 71.2668



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);

mu = 150

alpha = 0.05

2 independent samples of polished and unpolished stones are given , and no population standard deviation is known

hence a 2 sample , independent t test is a viable test

5.1 2 Sample Independent T Test.

Level of Significance : α = 0.05

Type of Test : 2 Sample Independent T Test - 2 Tailed

Hypothesis :

Null Hypothesis: Mean hardness of unpolished stones is equal to Mean hardness of Treated and Polished stone

H0 : μ Unpolished = μ Treated and Polished

Alternate Hypothesis: Mean hardness of unpolished stones is not equal to Mean hardness of Treated and Polished stone

Ha : μ Unpolished ≠ μ Treated and Polished

```
t_statistic :  -3.2422
p_value : 0.0015
```

0.0015 < 0.05 i.e p_value is less then level of significance (α), so we reject the null hypothesis , hence Mean hardness of Unpolished stones is not equal to Mean hardness of Treated and Polished stones

Therefore Zingaro has reason to believe now that the unpolished stones may not be suitable for printing.

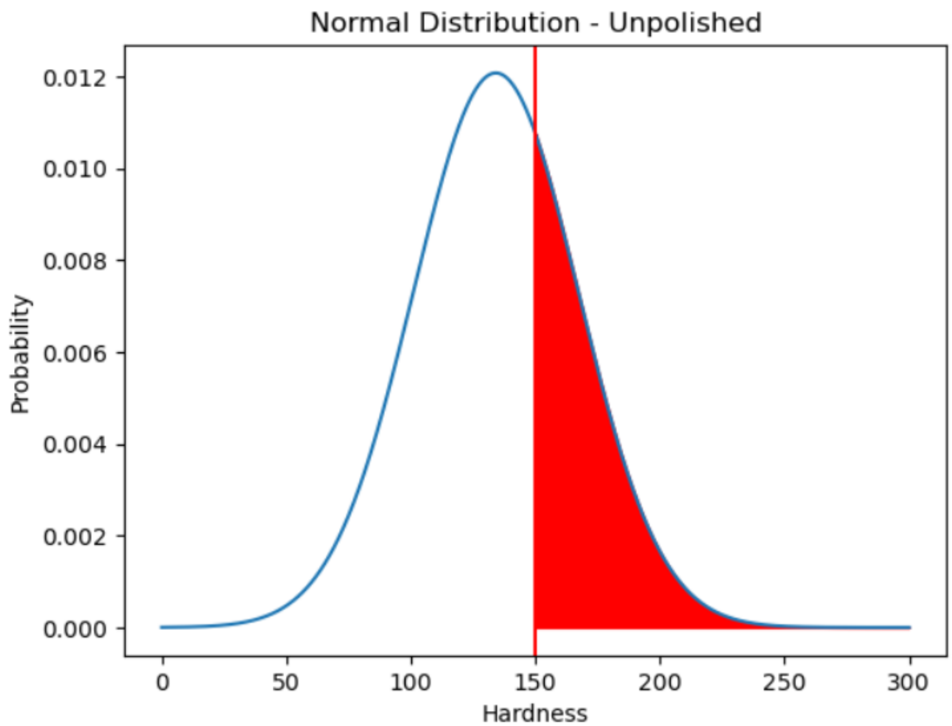
5.2 Is the mean hardness of the polished and unpolished stones the same?

	count	mean	std	min	25%	50%	75%	max
Unpolished	75.0	134.110527	33.041804	48.406838	115.329753	135.597121	158.215098	200.161313
Treated and Polished	75.0	147.788117	15.587355	107.524167	138.268300	145.721322	157.373318	192.272856

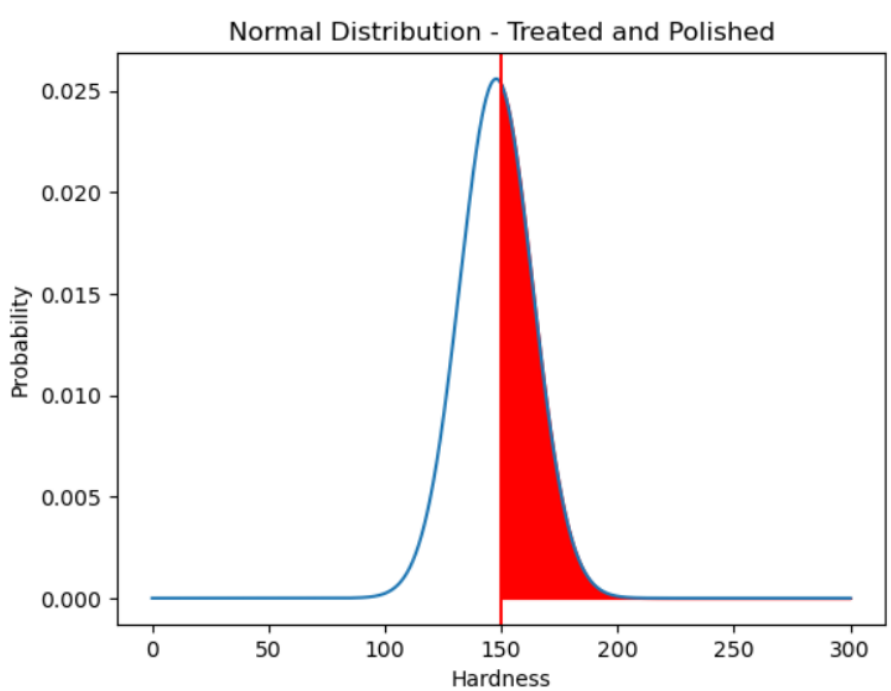
The Mean hardness of Treated and Polished Stones are way higher than Mean Hardness of the Unpolished Stones , and also as the Standard Deviation of Treated and Polished Stones is lower than that of Unplolished Stones , distribution of Treated and Polished stones hardness tend to lie more closely to their mean.

This can be illustrated further by , visualizing the probability that the hardness in Brinell's hardness index of at least 150, for Treated and Polished , Unpolished Stones.

$P_{\text{hardness} > 150} = 0.3153$ for Unpolished Stones



$P_{\text{hardness} > 150} = 0.4436$ - for Treated and Polished 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?

To test weather the program had effect on Before and After pushup performance

Level of Significance : α = 0.05

Type of Test : Paired T Test - One tailed

Hypothesis :

Null Hypothesis: Mean number of pushups before the program is equal to Mean number of pushups after the program

$H0 : \mu \text{ Before} = \mu \text{ After}$

Alternate Hypothesis: Mean number of pushups after the program is greater than Mean number of pushups before the program

$Ha : \mu \text{ Before} < \mu \text{ After}$

```
t_statistic : -19.322619811082458
p_value : 1.1460209626255983e-35
```

P- value is less than alpha , so we reject the null hypothesis therefore Mean number of pushups after the program is greater than Mean number of pushups before the program

But 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.

Now we need to perform a one tailed one sample t test on the difference against 6 as, if number of pushups greater than 5, the program is considered successful .

Level of Significance : α = 0.05

Type of Test : One Sample T Test - One tailed

Hypothesis :

Null Hypothesis: Mean number of pushups After program is at most 5 more than Mean number of pushups Before program

H0 : μ After - μ Before ≤ 5

Alternate Hypothesis: Mean number of pushups After program is at least 6 more than Mean number of pushups Before program

Ha : μ After - μ Before > 5

New Data Frame :

	Sr no.	Before	After	Difference
0	1	39	44	5
1	2	25	25	0
2	3	39	39	0
3	4	6	13	7
4	5	40	44	4
...	...	...	...	...
95	96	16	18	2
96	97	19	28	9
97	98	24	28	4
98	99	14	24	10
99	100	30	39	9

```
t_statistic -1.5666989036012808
p_value 0.060188015134156087
p_value<alpha:  False
```

As P values greater than alpha , we fail to reject the null hypothesis i.e Mean number of pushups After program is at most 5 more than Mean number of pushups Before program.

Hence we can conclude from the Paired T test and One Sample T test that , though the program was able to improve the number of pushups , according to its success criteria it was unsuccessful.

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.

Level of Significance : α = 0.05

Type of Test : one way ANOVA

Hypothesis :

Alloy 1

H0 : The mean implant hardness among all dentists is same for Alloy 1 i.e H0 : μ Dentist 1 = μ Dentist 2 = μ Dentist 3 = μ Dentist 4 = μ Dentist 5

Ha : The mean implant hardness is different in atleast one of the dentists for Alloy 1

Alloy 2

H0 : The mean implant hardness among all dentists is same for Alloy 2

Ha : The mean implant hardness is different in atleast one of the dentists for Alloy 2

7.2 Assumptions

Assumptions of one way Anova

1. The Samples are drawn from different populations are independent and random
2. Populations are continuous and ideally normally distributed
3. Variances are equal (at least approximately)

Assumptions of two way Anova

1. The dependent variable should be continuous
2. Two independent variables should each consist of two or more categorical , independent groups
3. No Signficant outliers
4. Dependent variable should be approximately normally distributed for each combination of the groups of the two independent variables

7.3 Conclusion regarding whether implant hardness depends on dentists?

Alloy 1

	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

P_value > alpha hence , we fail to reject the null hypothesis ie The mean implant hardness among all dentists is same for Alloy 1

If the null hypothesis were to be rejected, it is not possible to identify which pairs of Dentists differ, for alloy 1

Alloy 2

	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

P_value > alpha hence , we fail to reject the null hypothesis ie The mean implant hardness among all dentists is same for Alloy 2

If the null hypothesis were to be rejected, it is not possible to identify which pairs of Dentists differ for alloy 2

7.4 Test whether there is any difference among the methods on the hardness of dental implant, separately for the two types of alloys.

Alloy 1

H0 : The mean implant hardness among all Methods is same for Alloy 1.

Ha : The mean implant hardness is different in atleast one of the Methods for Alloy 1

	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

P_value < alpha , hence we reject the null hypothesis i.e The mean implant hardness is different in atleast one of the Methods for Alloy 1 .If the null hypothesis is rejected, it is not possible to identify which pairs of methods differ.

Alloy 2

H0 : The mean implant hardness among all Methods is same for Alloy 2.

Ha : The mean implant hardness is different in atleast one of the Methods for Alloy 2.

	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

P_value < alpha , hence we reject the null hypothesis i.e The mean implant hardness is different in atleast one of the Methods for Alloy 2 .If the null hypothesis is rejected, it is not possible to identify which pairs of methods differ.

7.5 Test whether there is any difference among the temperature levels on the hardness of dental implant, separately for the two types of alloys.

Alloy 1

H0 : The mean implant hardness among all Temperature Levels is same for Alloy 1.

Ha : The mean implant hardness among all Temperature Levels is not same for Alloy 1

	df	sum_sq	mean_sq	F	PR(>F)
Temp	1.0	10083.333333	10083.333333	0.681527	0.413618
Residual	43.0	636193.911111	14795.207235	NaN	NaN

P_value > alpha , hence we fail to reject the null hypothesis i.e The mean implant hardness among all Temperature Levels is same for Alloy 1. .If the null hypothesis is rejected, it is not possible to identify which levels of temperatures differ for Alloy 1

Alloy 2

H0 : The mean implant hardness among all Temperature Levels is same for Alloy 2.

Ha : The mean implant hardness among all Temperature Levels is not same for Alloy 2

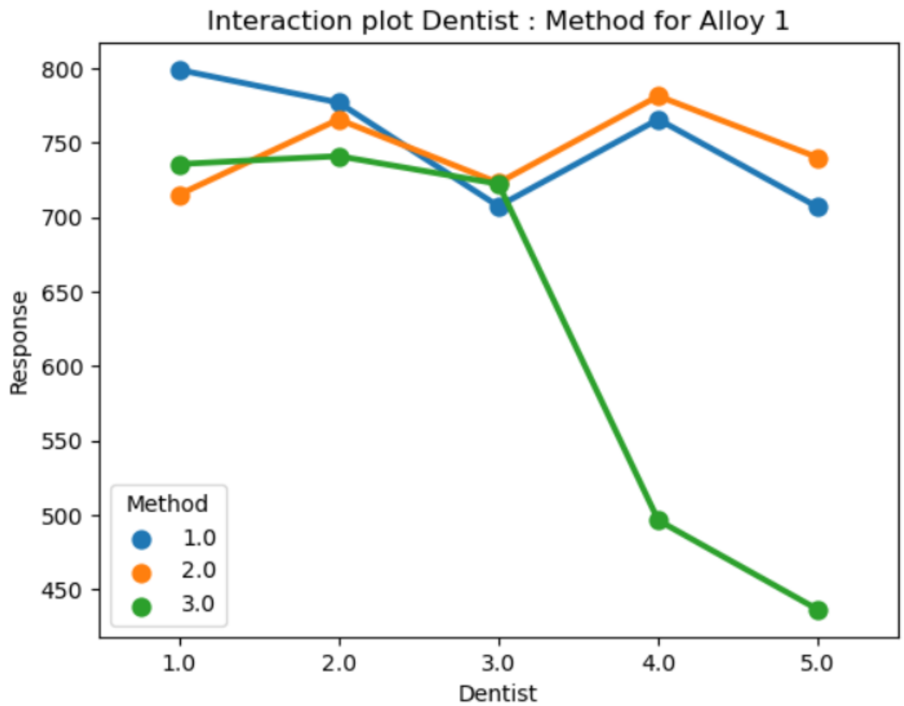
	df	sum_sq	mean_sq	F	PR(>F)
Temp	1.0	8.629603e+04	86296.033333	3.524941	0.067246
Residual	43.0	1.052707e+06	24481.552713	NaN	NaN

P_value > alpha , hence we fail to reject the null hypothesis i.e The mean implant hardness among all Temperature Levels is same for Alloy 2. .If the null hypothesis is rejected, it is not possible to identify which levels of temperatures differ for Alloy 2.

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

Alloy 1

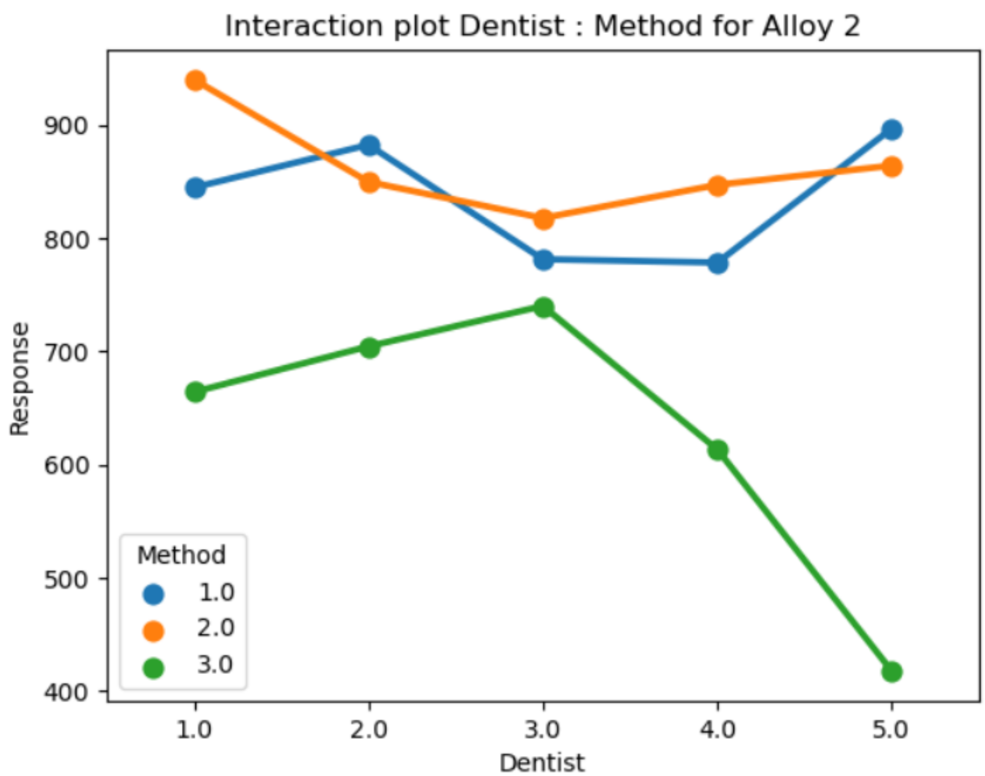
	df	sum_sq	mean_sq	F	PR(>F)
C(Dentist):C(Method)	14.0	441097.244444	31506.946032	4.606728	0.000221
Residual	30.0	205180.000000	6839.333333	NaN	NaN



We can see that our lines are intersecting, which means there's an interaction between Dentist, Method, and Implant hardness. The results are expected, as the P-value from the ANOVA test told us there's a significant interaction effect between them. i.e low p-value (p-value < 0.05)

Alloy 2

	df	sum_sq	mean_sq	F	PR(>F)
C(Dentist):C(Method)	14.0	753898.133333	53849.866667	4.194953	0.000482
Residual	30.0	385104.666667	12836.822222	NaN	NaN



We can see that our lines are intersecting, but not as much as alloy 1, which means there's some interaction between Dentist, Method, and Implant hardness. The results are expected, as the P-value from the ANOVA test told us there's a significant interaction effect between them. i.e low p-value (p-value < 0.05).

When we compare the F values for the Dentist Method interaction , its Slightly higher for Alloy 1 which is reflected by a greater number of intersection in the interaction plot.

7.7 Effect of both factors, dentist, and method, separately on each alloy.

Alloy 1

H0 : The mean implant hardness among all Dentists, all Methods and all interaction levels of Dentists and Methods is equal for Alloy 1.

Ha : The mean implant hardness among all Dentists, all Methods and all interaction levels of Dentists and Methods is not equal for Alloy 1.

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

As the p-value is less than alpha for all factors and interaction of factors , we reject the null hypothesis and hence The mean implant hardness among all Dentists, all Methods and all interaction of Dentists and Methods is not equal for Alloy 1.

But unfortunately ,it is not possible to identify which dentists are different, which methods are different, and which interaction levels are different.

Alloy 2

H0 : The mean implant hardness among all Dentists, all Methods and all interaction levels of Dentists and Methods is equal for Alloy 2.

Ha : The mean implant hardness among all Dentists, all Methods and all interaction levels of Dentists and Methods is not equal for Alloy 2.

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

As the p-value is more than alpha for dentists , we fail to reject the null hypothesis and hence The mean implant hardness among all Dentists, all Methods and all interaction levels of Dentists and Methods is equal for Alloy 2

But unfortunately ,it is not possible to identify which dentists are different, which methods are different, and which interaction levels are different.