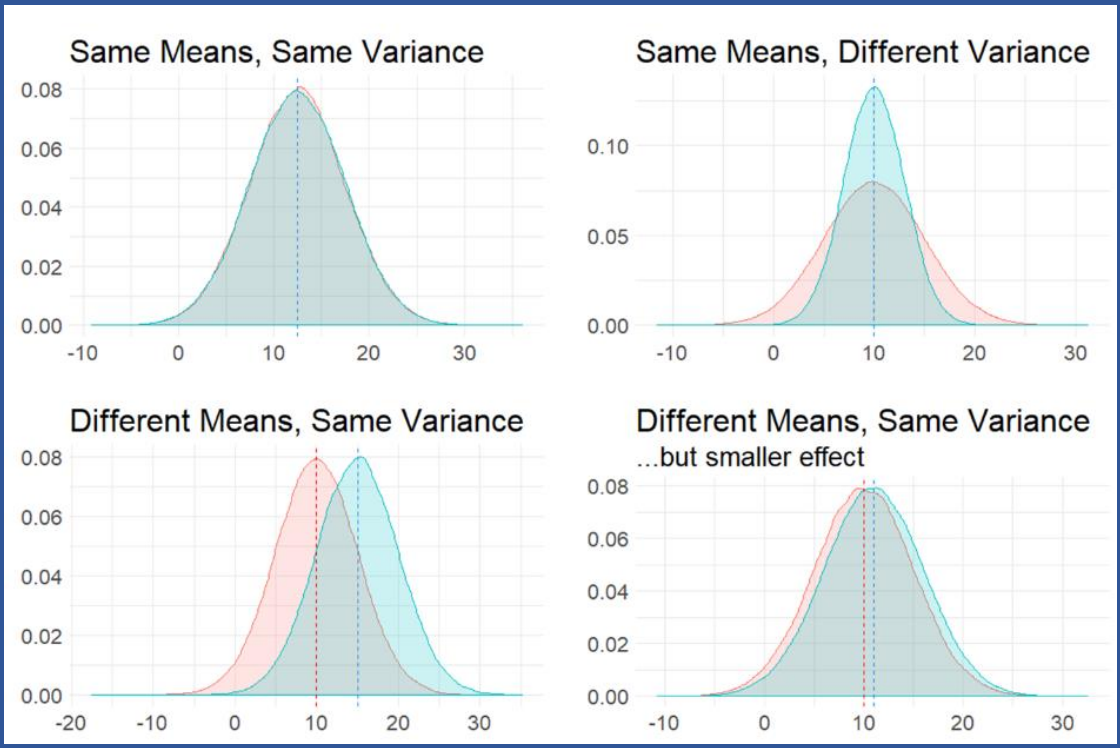
INFERENTIAL STATISTICS

(I S)PROJECT



GREAT LEARNING

DESIGNED BY :DURGESH KUMAR JHA

Session :2024-25

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**

Based on the above data, answer the following questions.

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

Total number of players played = 235

Total number of Injured Player = 145

Probability being Injured Player = 145/235 = 0.617

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

Total number of players played = 235

Total number of Forward Players = 94

Total number of Wigner = 29

Probability being Forward or Wigner = 123/235 = 0.523

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

Total number striker players played = 77

Total player having foot injury in striker position = 45

Probability being Injured while in striker position = 45/77 = 0.584

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

Total number of players played = 235

Total player having foot injury in striker position = 45

Probability being Injured while in striker position = 45/235 = 0.191

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

For solving the below questions, we need to use the Standard Normal Distribution function (SND) which is; -

$$Z = (X - \mu) / \sigma$$

Where; - X = Observed value

μ = mean = 5

σ = Standard Deviation = 1.5

Z = z-score

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

Z = (3.17 – 5) / 1.5

Z = -1.21

By using the SND table, the proportion to the left of Z score is -1.21 is 0.1131%.

So, the proportion of the gunny bags with having breaking strength of less than 3.17 kg per sq cm is approx. 11.31%.

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

Z = (3.60 – 5) / 1.5

Z = -0.93

By using the SND table, the proportion to the left of Z score is -0.93 is 0.1762.

So, the proportion of the gunny bags with having breaking strength of at least 3.60 kg per sq cm is approx. (1- 0.1762) % approx. 82.38%.

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

Z1 = (5 – 5) / 1.5

Z1 = 0

Z1 = (5.5 – 5) / 1.5

Z1 = 0.33

By using the SND table, the proportion to the left of Z score 0.33 is 0.629 and area proportion to the left of z-score is 0.5.

So, the proportion of the gunny bags with having breaking strength b\w 5 and 5.5 kg per sq cm is 0.629 – 0.5 = 0.129 approx. 12.9%.

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

Z1 = (3 – 5) / 1.5

Z1 = -1.33

Z2 = (7.5 - 5) / 1.5

Z2 = 1.67

By using the SND table, the proportion to the left of Z score -1.33 is 0.0918 and area proportion to the left of z-score 1.67 is 0.9525.

As, the proportion of the gunny bags with having breaking strength having b\w 3 and 7.5 kg per sq cm is 0.9525 – 0.0918 = 0.8607 approx. 86.07%.

So, the proportion of the gunny bags with having breaking strength not having b\w 3 and 7.5 kg per sq cm is 1 – 0.8607 = 0.1393 approx. 13.93%.

Q 3. 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)

	Unpolished	Treated and Polished
0	164.481713	133.209393
1	154.307045	138.482771
2	129.861048	159.665201
3	159.096184	145.663528
4	135.256748	136.789227

Image 1

Above figure shows the first five rows and columns of data set

	Unpolished	Treated and Polished
70	123.067611	142.293544
71	171.822218	140.124092
72	88.135994	141.393091
73	145.150397	131.370530
74	170.854823	144.502647

Image 2

Above figure shows the bottom five rows and columns of data set

	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

Image 3

Above figure describes the data set.

We have 2 Columns and 75 rows.

```
Data columns (total 2 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Unpolished           75 non-null     float64
 1   Treated and Polished 75 non-null     float64
```

Image 4

Above figure shows we have not any null values in our data set.

3.1 Zingaro has reason to believe that the unpolished stones may not be suitable for printing. Do you think Zingaro is justified in thinking so?

We have the confidence interval of 5% $\alpha = 0.05$, and number of samples = 75

Significance of the test	**Assumptions**	**Test Statistic Distributions**
Test for the Population mean $H_0: \mu = \mu_0$	Continuous data Normally Distributed and sample size < 30 Unknown Standard Deviation Random sampling from the population	T distribution One-sample t-test

Let's test our null Hypothesis and alternate hypothesis

We will test the Null hypothesis, against the Alternate Hypothesis

H_0: stone found to be had adequate hardness >= 150

H_a: unpolished stone not suitable for printing < 150

The p-value is 4.171286997419652e-05

Insight- Zingaro is justified in thinking that the unpolished stones may not be suitable for printing.

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

Significance of the test	Assumptions	Test Statistic Distributions
Test for the Population mean H_0: $\mu = \mu_0$	Continuous data Normally Distributed and sample size < 30 Unknown Standard Deviation Random sampling from the population	T distribution 2-sample t-test

We will test the Null hypothesis, against the Alternate Hypothesis

H_0: The mean hardness for both unpolished and polished stones are same

H_a: The mean hardness for both unpolished and polished stones are not same

The p-value is 0.0014655150194628353

Insight- The mean hardness of the polished and unpolished stones is the different.
We also found the same result with the Describe function: -

```
The mean hardness of the Unpolished stone is 134.11052653373332
The mean hardness of the Unpolished stone is 147.78811718133335
```

Image 5

Q 4. Dental implant data: The hardness of metal implants 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 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.

	Dentist	Method	Alloy	Temp	Response
0	1	1	1	1500	813
1	1	1	1	1600	792
2	1	1	1	1700	792
3	1	1	2	1500	907
4	1	1	2	1600	792

Image 6

Above figure shows the top 5 rows and columns of data set

	Dentist	Method	Alloy	Temp	Response
85	5	3	1	1600	483
86	5	3	1	1700	405
87	5	3	2	1500	536
88	5	3	2	1600	405
89	5	3	2	1700	312

Image 7

Above figure shows the bottom 5 rows and columns of data set

To solve the given problems, we changed the data set based on alloy type as, Alloy1_Data and Alloy2_Data.

	Dentist	Method	Alloy	Temp	Response
0	1	1	1	1500	813
1	1	1	1	1600	792
2	1	1	1	1700	792
6	1	2	1	1500	782
7	1	2	1	1600	698

Image 8

Above figure shows the top 5 rows and columns of data set Alloy1_Data

	Dentist	Method	Alloy	Temp	Response
3	1	1	2	1500	907
4	1	1	2	1600	792
5	1	1	2	1700	835
9	1	2	2	1500	1115
10	1	2	2	1600	835

Image 9

Above figure shows the bottom 5 rows and columns of data set Alloy2_Data.

	Dentist	Method	Alloy	Temp	Response
count	45.000000	45.000000	45.0	45.000000	45.000000
mean	3.000000	2.000000	1.0	1600.000000	707.488889
std	1.430194	0.825723	0.0	82.572282	121.194551
min	1.000000	1.000000	1.0	1500.000000	289.000000
25%	2.000000	1.000000	1.0	1500.000000	681.000000
50%	3.000000	2.000000	1.0	1600.000000	743.000000
75%	4.000000	3.000000	1.0	1700.000000	782.000000
max	5.000000	3.000000	1.0	1700.000000	882.000000

Image 10

Above figure shows the Statistical figure of data set Alloy1_Data

	Dentist	Method	Alloy	Temp	Response
count	45.000000	45.000000	45.0	45.000000	45.000000
mean	3.000000	2.000000	2.0	1600.000000	776.066667
std	1.430194	0.825723	0.0	82.572282	160.892595
min	1.000000	1.000000	2.0	1500.000000	312.000000
25%	2.000000	1.000000	2.0	1500.000000	715.000000
50%	3.000000	2.000000	2.0	1600.000000	824.000000
75%	4.000000	3.000000	2.0	1700.000000	858.000000
max	5.000000	3.000000	2.0	1700.000000	1115.000000

Image 11

Above figure shows the Statistical summary of data set Alloy2_Data

	Dentist	Method	Temp	Response
count	90.000000	90.000000	90.000000	90.000000
mean	3.000000	2.000000	1600.000000	741.777778
std	1.422136	0.821071	82.107083	145.767845
min	1.000000	1.000000	1500.000000	289.000000
25%	2.000000	1.000000	1500.000000	698.000000
50%	3.000000	2.000000	1600.000000	767.000000
75%	4.000000	3.000000	1700.000000	824.000000
max	5.000000	3.000000	1700.000000	1115.000000

Image 12

Above figure shows the Statistical summary of data set df_excel, after the correction in the original data set.

After the correction we don't have the Alloy column in the original data set.

4.1 How does the hardness of implants vary depending on dentists?

Significance of the test	Assumptions	Test Statistic Distributions
Test for more than two populations H_0: All population means are equal	The populations are normally distributed. Samples are independent simple random samples. Population variance is equal	F distribution 1-way AVOVA F-Test

Let's write the Null and Alternate Hypothesis

H_0: There is no significance difference in implants hardness amongst different dentist

Ha: There is significance difference in implants hardness amongst different dentist

The p-value for Alloy1_Data is 0.0004324350

The p-value for Alloy2_Data is 0.0004324350

As per the requirement of the question we will run test separately base on the data set Alloy1_Data and Alloy2_Data.

We perform Levene test with Alloy1_Data

The p-value is 7.025697275572713e-07

We perform F-one-way ANOVA test with Alloy1_Data

The p-value is 2.613905810174214e-57

We perform Levene test with Alloy2_Data

The p-value is 2.5314647179229246e-07

We perform F-one-way ANOVA test with Alloy1_Data

The p-value is 1.6648108558440017e-50

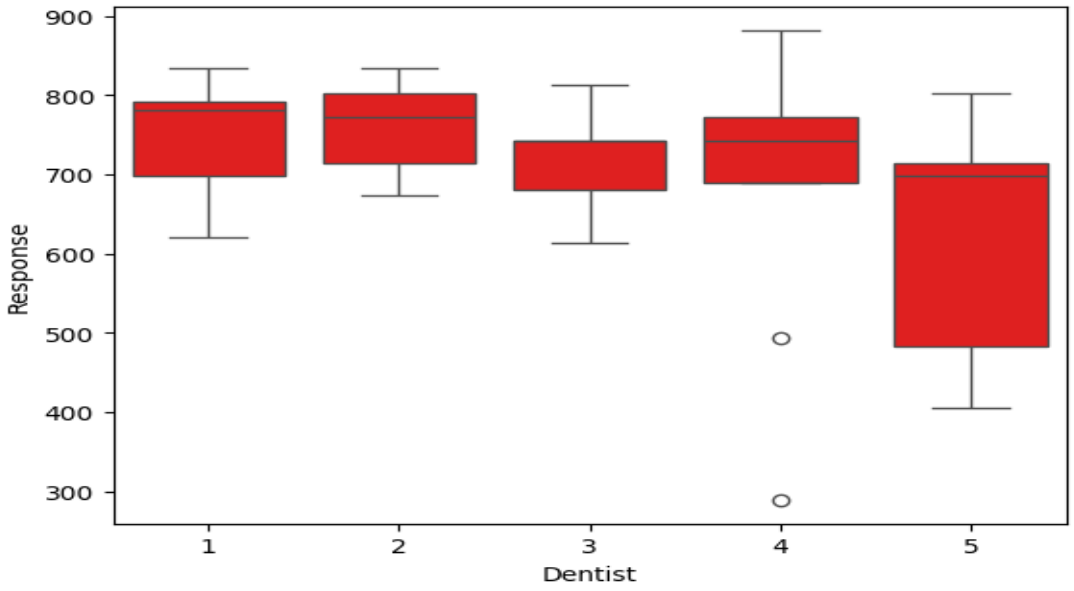


Image 13

Above figure shows the relationship b\w the Dentist and the response or Hardness for Alloy1_Data

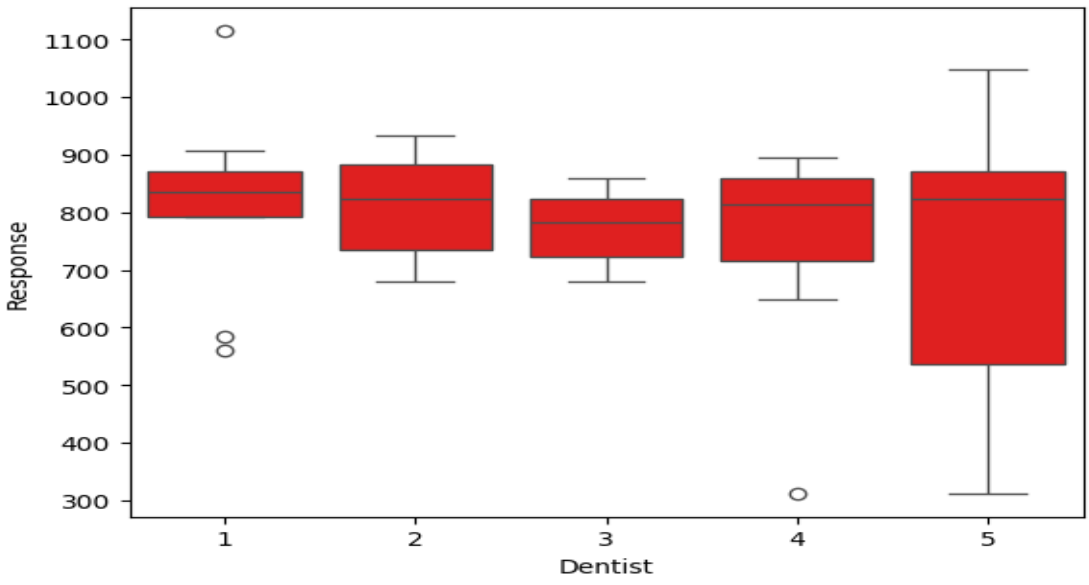


Image 14

Above figure shows the relationship b\w the Dentist and the response or Hardness for Alloy2_Data

As we found out the value of p-value which is much less than the Level of significance so we reject the Null Hypothesis

4.2 How does the hardness of implants vary depending on methods?

Significance of the test	**Assumptions**	**Test Statistic** **Distributions**
Test for more than two populations H_0: All population means are equal	The populations are normally distributed. Samples are independent simple random samples. Population variance is equal	F distribution 1-way AVOVA F-Test

Let's write the Null and Alternate Hypothesis

H_0: There is no significance difference in implants hardness amongst different Methods

Ha: There is significance difference in implants hardness amongst different Methods

The p-value for Alloy1_Data is 1.74074114056e-06

The p-value for Alloy2_Data is 1.74074114056e-06

As per the requirement of the question we will run test separately base on the data set Alloy1_Data and Alloy2_Data.

We perform Levene test with Alloy1_Data

The p-value is 6.029123452339682e-07

We perform F-one-way ANOVA test with Alloy1_Data

The p-value is 2.313824455454923e-57

We perform Levene test with Alloy2_Data

The p-value is 2.2470939093460585e-07

We perform F-one-way ANOVA test with Alloy1_Data

The p-value is 1.4956915786544367e-50

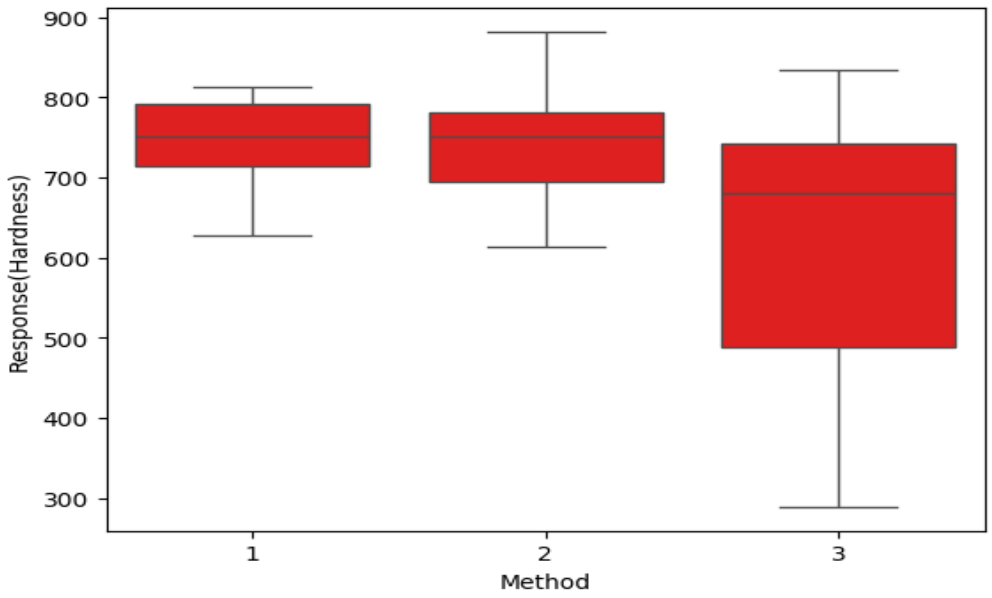


Image 15

Above figure shows the relationship b\w the Method and the response or Hardness for Alloy1_Data

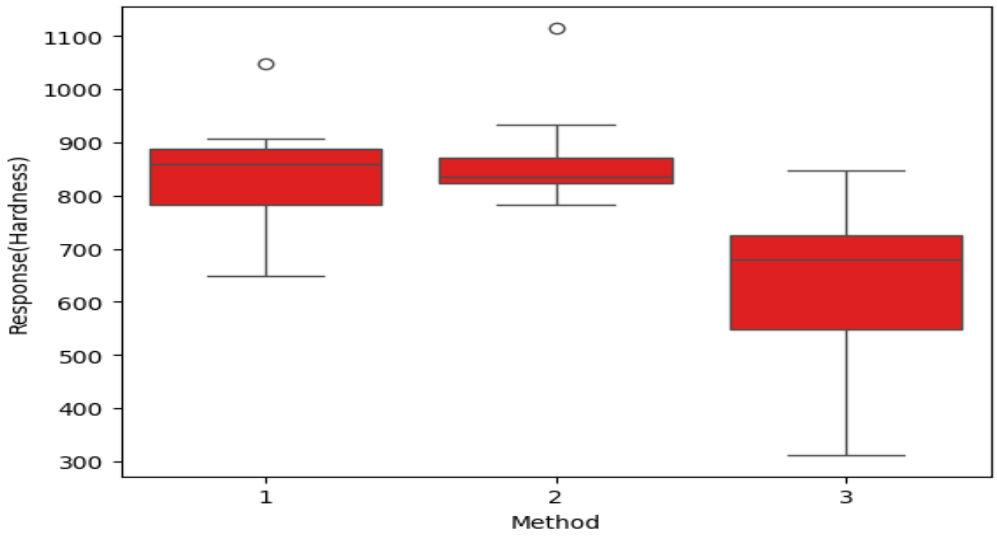


Image 16

Above figure shows the relationship b\w the Method and the response or Hardness for Alloy2_Data

As we found out the value of p-value which is much less than the Level of significance so we reject the Null Hypothesis.

So, there is a great significance in implants hardness amongst different Methods.

4.3 What is the interaction effect between the dentist and method on the hardness of dental implants for each type of alloy?

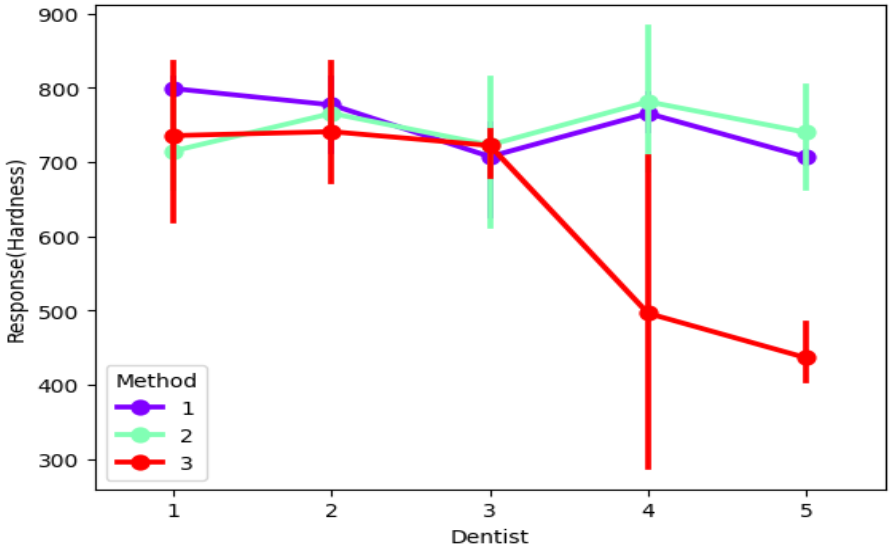


Image 17

The above figure shows the interaction effect b\w the dentist and method on the hardness of dental implants for Alloy type 1.

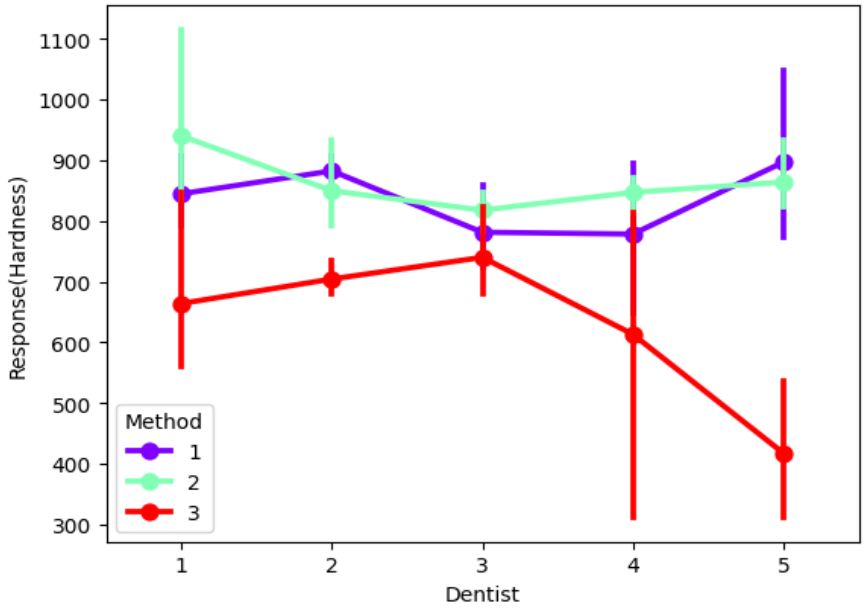


Image 18

The above figure shows the interaction effect b\w the dentist and method on the hardness of dental implants for Alloy type 2.

4.4 How does the hardness of implants vary depending on dentists and methods together?

Significance of the test	Assumptions	Test Statistic Distributions

Test for more than two populations H_0: All population means are equal	The populations are normally distributed. Samples are independent simple random samples. Population variance is equal	F distribution 2-way AVOVA F-Test

Let's write the Null and Alternate Hypothesis

H_0: There is no significance difference in implants hardness based on different dentist and methods

Ha: There is significance difference in implants hardness based on different dentist and methods

	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

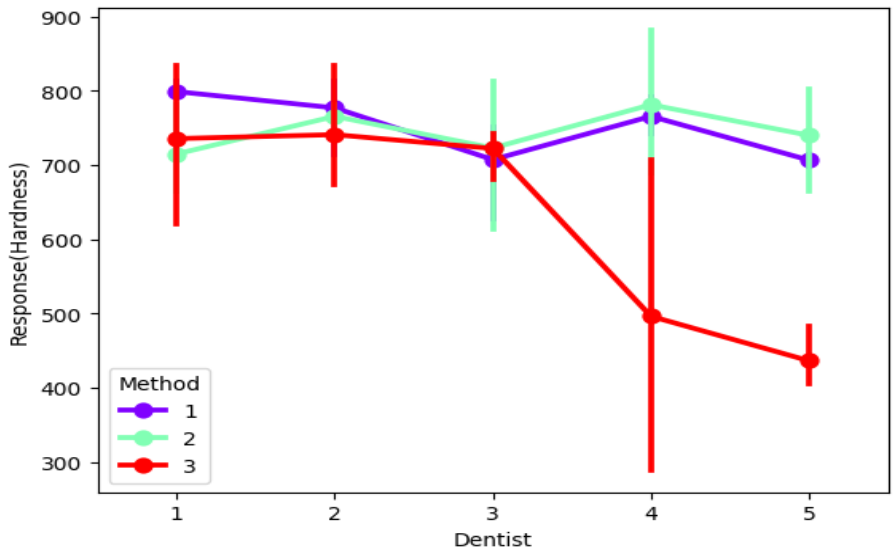


Image 19

Above Figure and graphical representation shows the Relationship between the Hardness with both dentist and method used for Alloy type 1

The p-value we have by comparing both dentist and method used with the Hardness (Response) is 0.0063.

As we found out the value of p-value which is much less than the Level of significance so we reject the Null Hypothesis.

	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

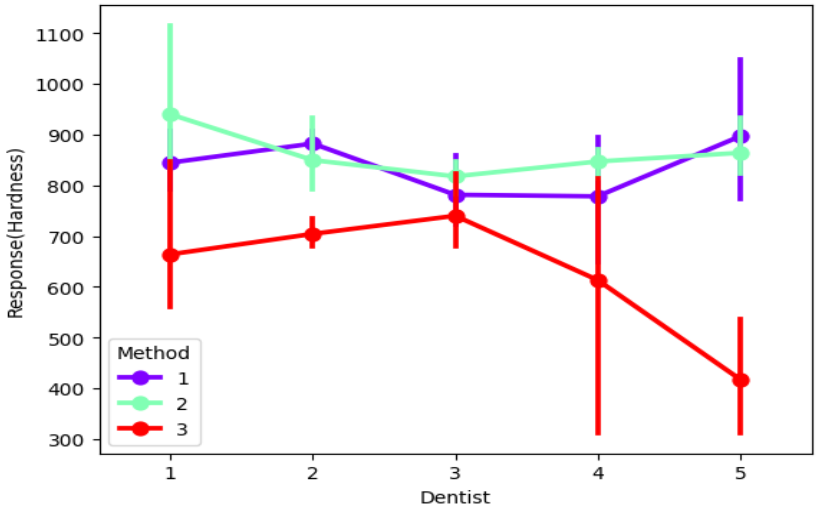


Image 20

Above figure and graphical representation shows the difference in implants hardness based on different dentist and method for Alloy type 2

The p-value we have by comparing both dentist and method used with the Hardness (Response) is 0.093234.

As we found out the value of p-value which is more than the Level of significance so can't reject the Null Hypothesis.