BUSINESS REPORT

IS Project Coded

Anand B R | Batch PGPDSBA.O.JAN24

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

Table 1.1 Margin Data of the Football team with position and Injury

	Striker	Forward	Attacking	Winger	Total	
Player injured	45	56	24	20	145	
Players not injured	32	38	11	9	90	
Total	77	94	35	29	235	
1.1	61,70213					
1.2	52,34043					
1.3	19,14894					
1.4	31,03448					

Table 1.2 is an Excel sheet to calculate Marginal probability.

WHAT IS THE PROBABILITY THAT A RANDOMLY CHOSEN PLAYER WOULD SUFFER AN INJURY?

The probability of a Player suffering an Injury is the ratio of Players Injured by Total Players.

= 61.7%

1.1 WHAT IS THE PROBABILITY THAT A PLAYER IS A FORWARD OR A WINGER?

Probability that is player is forward or a winger is the sum of Probability of Player is Forward with Probability of Player is Winger:

probability (player is forward) = 94/235 =
probability (player is winger) = 29/235

= 52.34%

1.2 WHAT IS THE PROBABILITY THAT A RANDOMLY CHOSEN PLAYER PLAYS IN A STRIKER POSITION AND HAS A FOOT INJURY?

Probability that player plays in Striker and has foot injury is the ratio of Striker Injured Player by Total Player

= 19.15%

1.3 WHAT IS THE PROBABILITY THAT A RANDOMLY CHOSEN INJURED PLAYER IS A STRIKER?

Probability of Injured player is a Striker.

Bayes Theorem:

Probability(striker|injured) = Probability (striker) * Probability (injured|striker) / Probability (Injured)

Probability (striker) = 77/235, probability (injured|striker) = 45/77, probability (injured) = 145/235

= 31.03%

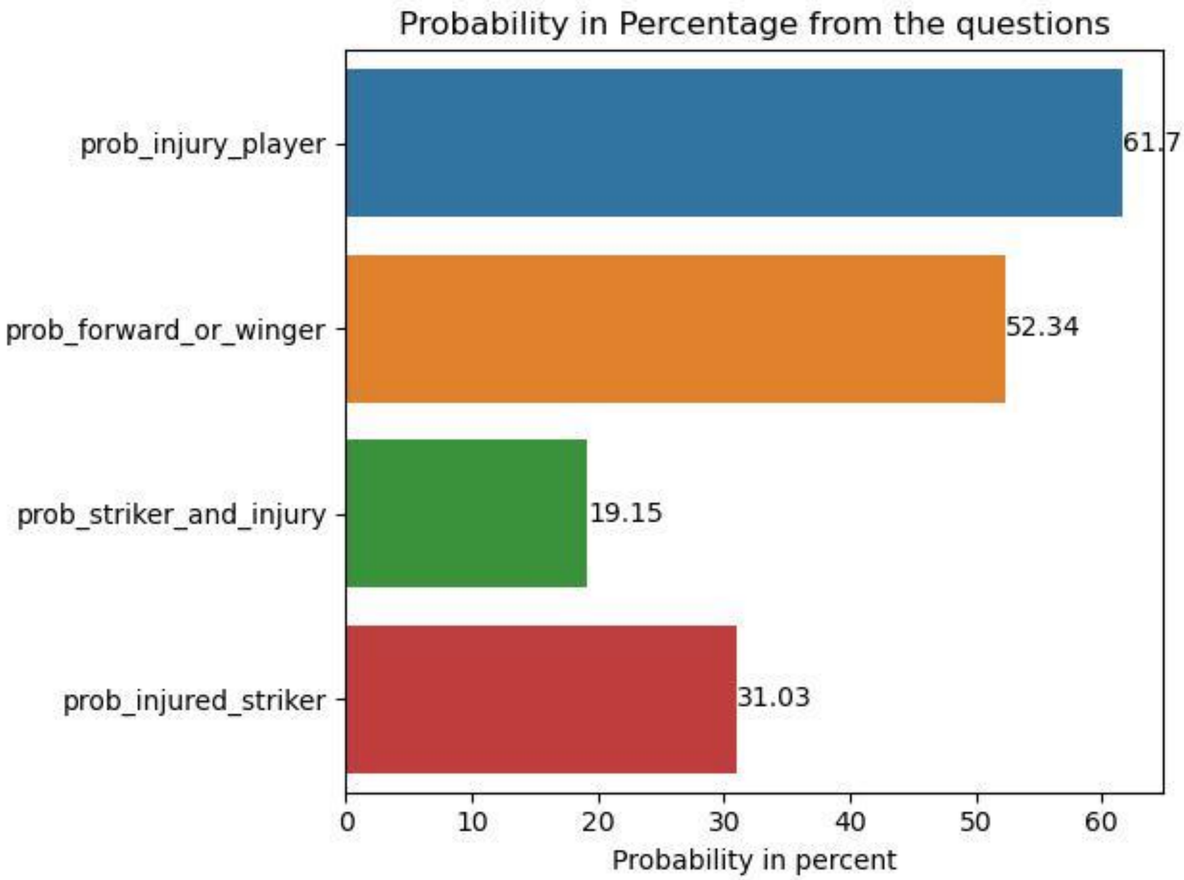


fig1. visual representation of Probability for Problem 1

INFERENCE

- Analyzing Injury probabilities by player positions can provide information on which position is prone to Injury.

PROBLEM 2

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)**

Population Mean = 5Kg/sq cm
Population standard deviation = 1.5 kg/sq cm
Population is Normally distributed.

2.1 WHAT PROPORTION OF THE GUNNY BAGS HAVE A BREAKING STRENGTH OF LESS THAN 3.17 KG PER SQ CM?

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

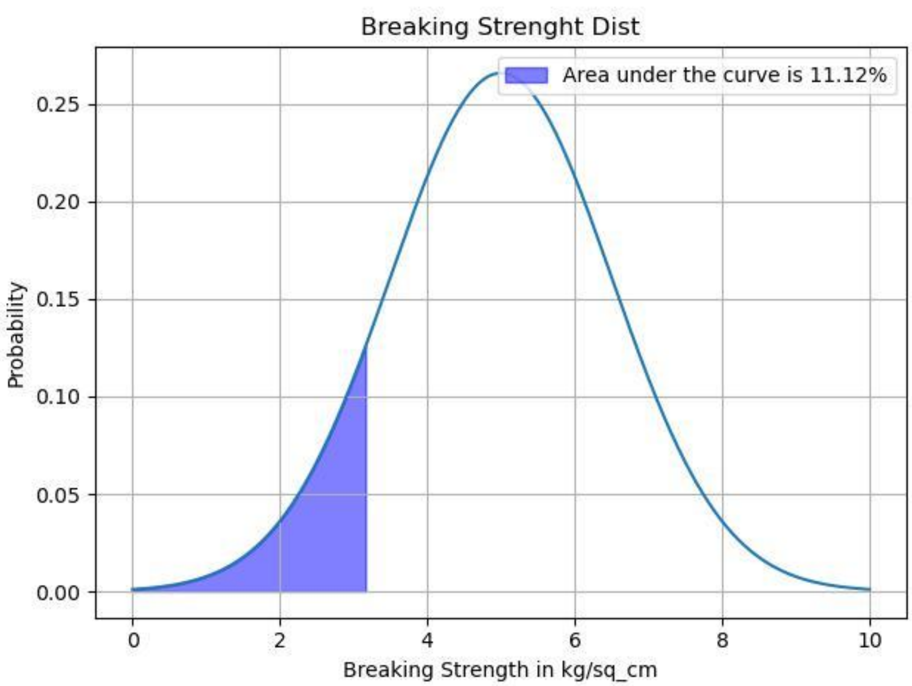


fig 2.1 Percentage of Gunny bags with breaking strength less than 3.17 Kg/sq cm

2.2 WHAT PROPORTION OF THE GUNNY BAGS HAVE A BREAKING STRENGTH OF AT LEAST 3.6 KG PER SQ CM.?

The proportion of gunny bags have a breaking strength of at least 3.6 kg per sq cm 82.47%.

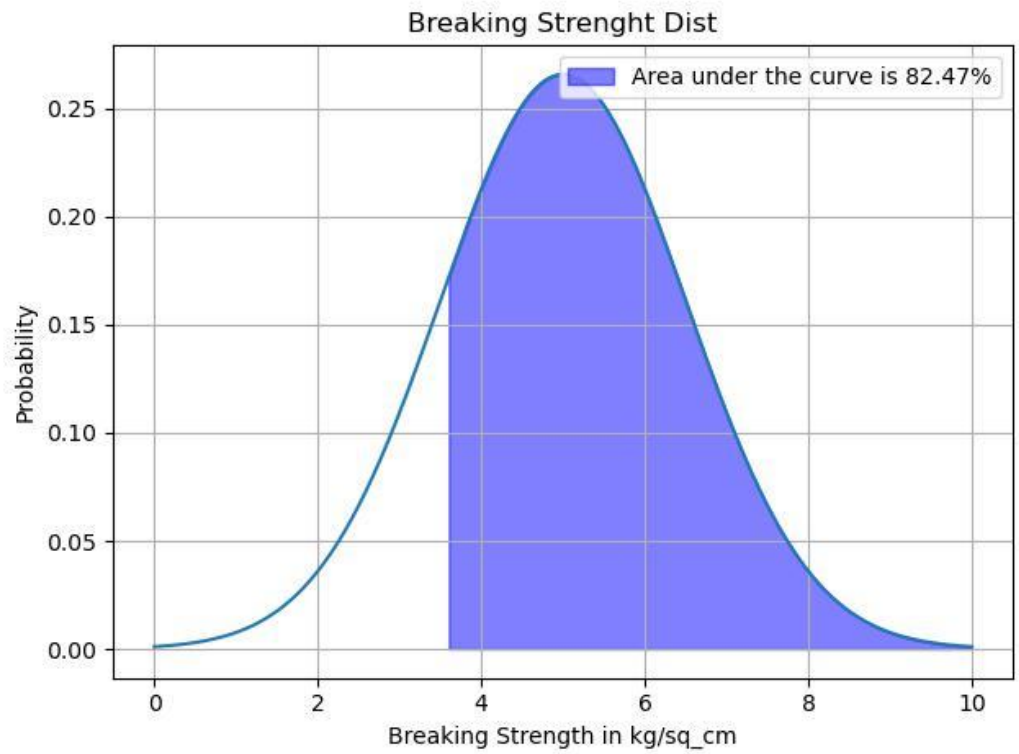


fig 2.2 Percentage of Gunny bags with breaking strength at least 3.6 Kg/sq cm

2.3 WHAT PROPORTION OF THE GUNNY BAGS HAVE A BREAKING STRENGTH BETWEEN 5 AND 5.5 KG PER SQ CM.?

The proportion of the gunny bags have a breaking strength between 5 and 5.5 kg per sq cm 13.06%.

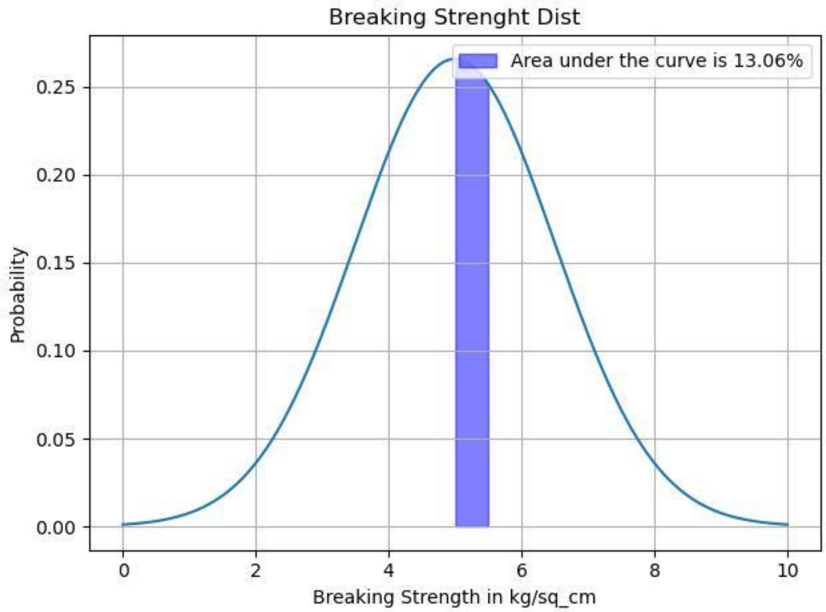


fig 2.3 Percentage of Gunny bags with breaking strength between 5 and 5.5 kg per sq cm

2.4 WHAT PROPORTION OF THE GUNNY BAGS HAVE A BREAKING STRENGTH NOT BETWEEN 3 AND 7.5 KG PER SQ CM.?

A proportion of the gunny bags have a breaking strength NOT between 3 and 7.5 kg per sq cm 13.9%.

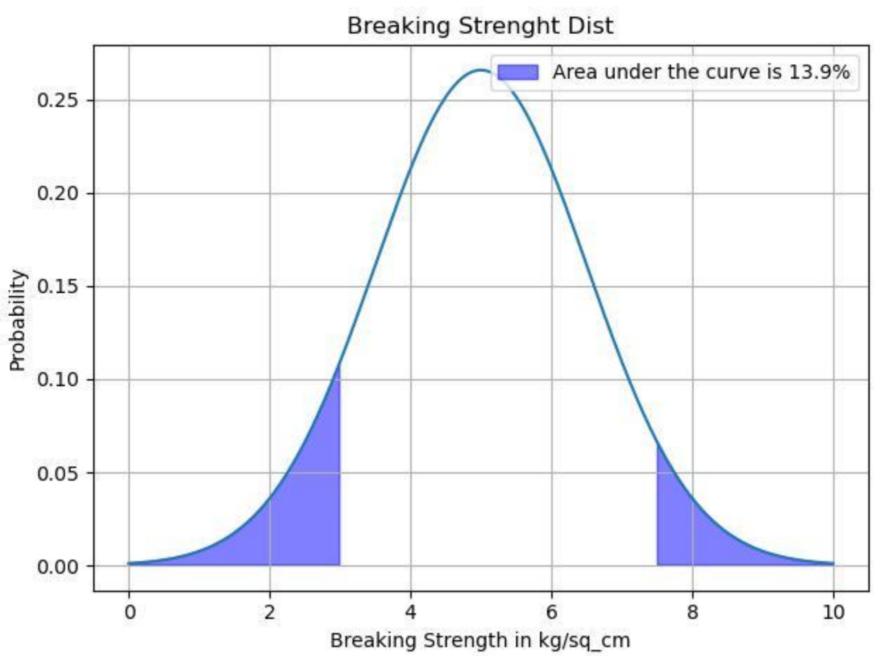


fig 2.4 Percentage of Gunny bags with breaking strength NOT between 3 and 7.5 kg per sq cm

PROBLEM 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)

Hardness index at least 150.
Samples from: Unpolished, Polished Stones
Significance level (alpha): 0.05

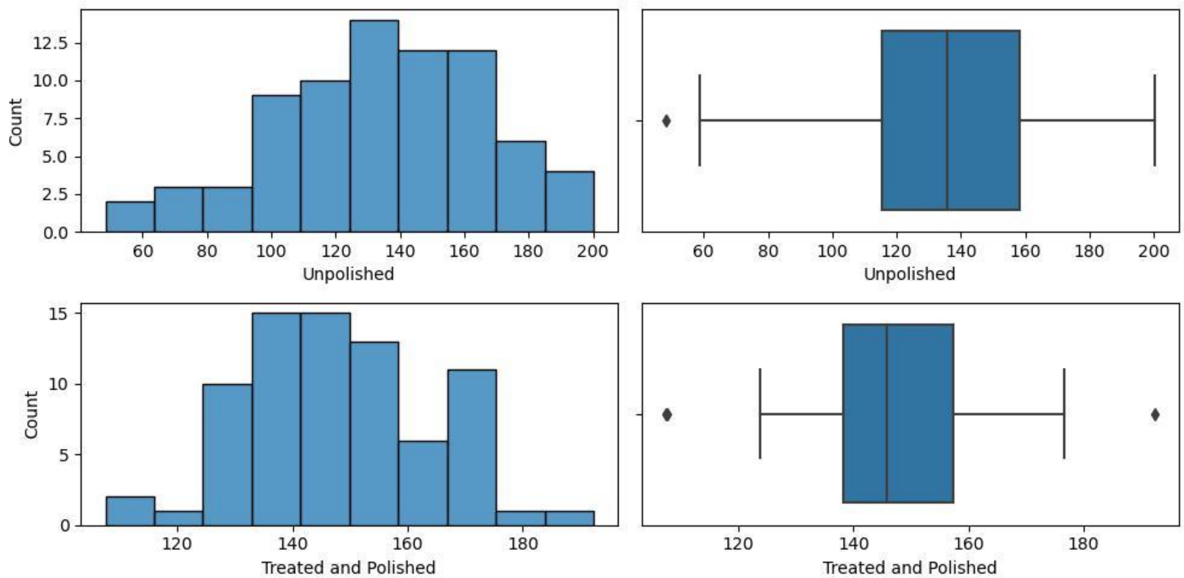


fig 3. Histogram and Box plot of Polished and unpolished stone samples.

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?

Analyze are Unpolished stones Not suitable for printing.

Step1: Define the Hypothesis

Null Hypothesis: Hardness of Unpolished stones is at least Brinell's hardness index 150.
Ho: ≥ 150

Alternate Hypothesis: Hardness of Unpolished stones is less than Brinell's Hardness index 150.
H1: < 150

- One sample LEFT Tailed test.

Step2: Significance level

Alpha:

α = 0.05

Step3: Calculate Test Statistics & Critical Value/P Value

- 1 sample
  - Since the sample size is 75 (>30), We can conduct.
    - One sample t-test

      Or
    - 1 sample Z test

Result of One Sample t-test:

	1 sample T-test	1 sample Z test
T statistics	-4.164629601426757	-4.164629601426757
P value	4.171286997419652e-05	1.5592921935294406e-05

Table 3.1 t-statistics and p values for unpolished stones samples

Step 4: Conclusion

Since P value (4.171286997419652e-05) is less than alpha (α = 0.05)

- Reject the Null Hypothesis

There is evidence that the Hardness Index of Unpolished Stones is significantly less than 150. Zingaro's thinking that unpolished stones are not suitable for printing is justified.

3.2 IS THE MEAN HARDNESS OF THE POLISHED AND UNPOLISHED STONES THE SAME?

Analyze if the mean hardness index of Unpolished stones and Polished are similar.

Step1: Define the Hypothesis

Null Hypothesis: Mean hardness of Polished Stones is equal to Mean hardness of Unpolished stones.

Ho: μ(polished) = μ (unpolished)

Alternate Hypothesis: Mean hardness of Polished Stones and Unpolished stones are different.

H1: μ(polished) ≠ μ (unpolished)

- Two samples two tailed test.

Step2: Significance level

Alpha:
$\alpha = 0.05$

Step3: Calculate Test Statistics & Critical Value/P Value

- Two sample 2 tailed test:
  - As sample size is > 30, we can conduct:
    - 2 sample t-test independent with equal variance
      Or
    - 2 sample Z test

Result of Two Sample t-test:

	2 sample T-test independent	2 sample Z test
T statistics	3.2422320501414053	3.242232050141406
P value	0.0014655150194628353	0.0011859741319915178

Table 3.2 t-statistics and p values for unpolished & Polished Stones stones samples

Step 4: Conclusion

Since P value (0.0014655150194628353) is less than alpha ($\alpha = 0.05$)

- Reject the Null Hypothesis

There is evidence that the mean Hardness Index of Unpolished Stones and Polished stones is Different.

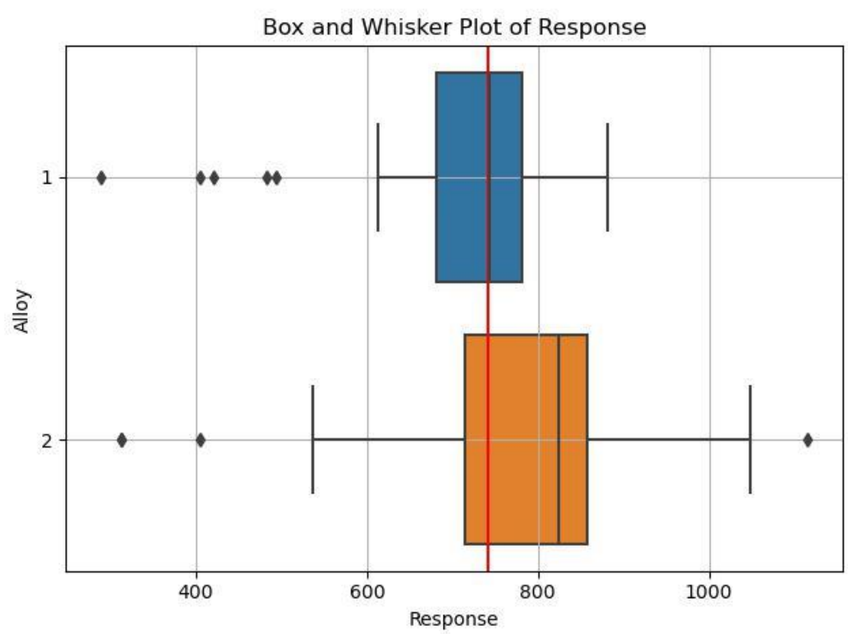
PROBLEM 4

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 favor one method above another and may work better in his/her favorite method. The response is the variable of interest.

Data Insights:

- Categorical variables (independent):
  - Dentist, Temp, Method
  - Alloy
- Dependent Variable (Numerical/Continuous):
  - Response
- All questions Note to consider Alloy type separately.
  - Create child datasets unique to Alloy type.
- No null values
- ANOVA test

<u>Plotting Response Mean v/s Median for Each alloy type</u>.



<u>fig 4.1 Box Plot of Response for each Alloy Type</u>

Point Estimate for Alloy 1:

	Alloy	***Response***
count	*45.0*	*45.000000*
mean	*1.0*	*707.488889*
std	*0.0*	*121.194551*
min	*1.0*	*289.000000*
25%	*1.0*	*681.000000*
50%	*1.0*	*743.000000*
75%	*1.0*	*782.000000*
max	*1.0*	*882.000000*

Point Estimate for Alloy 1:

	Alloy	***Response***
count	*45.0*	*45.000000*
mean	*2.0*	*776.066667*
std	*0.0*	*160.892595*
min	*2.0*	*312.000000*
25%	*2.0*	*715.000000*
50%	*2.0*	*824.000000*
75%	*2.0*	*858.000000*
max	*2.0*	*1115.000000*

CONDUCT NORMALITY TEST

a. Shapiro-Wilk Test: Assess Whether given sample data comes from a Normally Distributed Population

Results:

Alloy1:

```
Shapiro-Wilk Test Statistics for Alloy_1: 0.8304629921913147
Shapiro-Wilk p value for Alloy_1: 1.1945070582441986e-05
```

Alloy 2:

```
Shapiro-Wilk Test Statistics for Alloy_2: 0.887769341468811
Shapiro-Wilk p value for Alloy_2: 0.00040293222991749644
```

b. Anderson-Darling Test: Assess Whether given sample data comes from a Normally Distributed Population

Results:

Alloy 1:

```
Anderson-Darling Test Statistics for Alloy_1:  2.561066309273201
critical value for Alloy_1: [0.535 0.609 0.731 0.853 1.014]
significance Value for Alloy_1: [15.  10.   5.   2.5  1. ]
```

Alloy 2:

```
Anderson-Darling Test Statistics for Alloy_2:  1.8931311726356412
critical value for Alloy_2: [0.535 0.609 0.731 0.853 1.014]
```

```
significance Value for Alloy_2: [15.  10.   5.   2.5 1. ]
```

INFERENCE FROM SHAPIRO AND ANDERSON TEST:

- Shapiro Test p value is extremely low than alpha (0.05), which Rejects HO or Data might not be Normally distributed.
- Anderson-Darling Test statistic greater than Critical Value w.r.t to Significance level(α) which suggest significance evid ence to reject HO or Data sample of Population which might not be Normally distributed.
- Distributions for Both Alloy 1 and Alloy 2 are not normally Distributed.
- Outliers (visualized in the Plot) in the Data and ANOVA are sensitive to Outliers.

c. Levene Test: Test for Variance equality

Results:

```
Levene Test Statistics for Alloy groups: 1.4194717470917784
Levene p value for Alloy groups: 0.23669380462584474
```

INFERENCE FROM LEVENE TEST:

- Levene's Test p value is greater than alpha (0.05), which suggests Variances across the groups are equal approximately or No significance evidence to reject Ho.

4.1 HOW DOES THE HARDNESS OF IMPLANTS VARY DEPENDING ON DENTISTS?

Step1: Define the Hypothesis

Null Hypothesis: Mean implant hardness is similar across Dentists.

Ho: $\mu1 = \mu2 = \mu3 = \mu4 = \mu5$

Alternate Hypothesis: Mean Implant hardness is Different across Dentists (at least one mean is different from others)

H1: Not all means are equal.

Step2: Significance level

Alpha:

$\alpha = 0.05$

Step3: Calculate Test Statistics & Critical Value/P Value

One way ANOVA test:

Alloy 1:

Result:

```
One-way Anova Test stats Alloy_1: 1.9771119908770842
One-way Anova P value Alloy_1: 0.11656712140267628
```

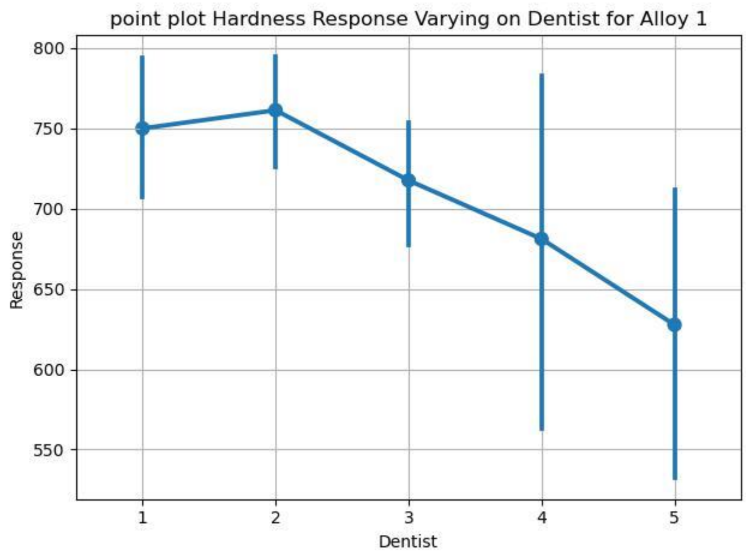


Fig 4.2 Point plot Hardness Response varying on Dentist for Alloy 1.

Alloy 2:

Result:

```
One-way Anova Test stats Alloy_2: 0.5248351000282961
One-way Anova P value Alloy_2: 0.7180309510793431
```

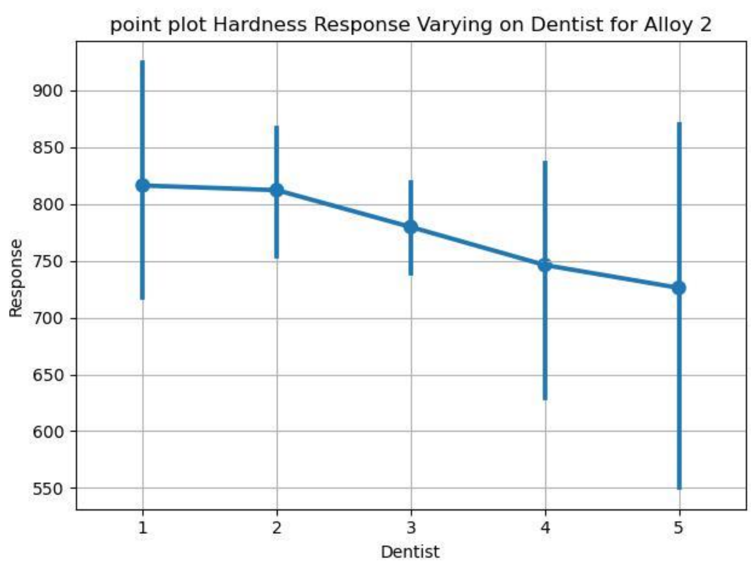


Fig 4.3 Point plot Hardness Response varying on Dentist for Alloy 2.

Step 4: Conclusion

Alloy 1:

- P Value (0.1165) > alpha (0.05)
- Failed to Reject Null Hypothesis (Ho)
- No sufficient evidence that there is significant difference between the mean implant hardness varying depending on Dentists.

Alloy 2:

- P Value (0.7180) > alpha (0.05)
- Failed to Reject Null Hypothesis (Ho)
- No sufficient evidence that there is significant difference between the mean implant hardness varying depending on Dentists.

4.2 HOW DOES THE HARDNESS OF IMPLANTS VARY DEPENDING ON METHODS?

Step1: Define the Hypothesis

Null Hypothesis: Mean implant hardness is similar across Methods.

Ho: $\mu1 = \mu2 = \mu3$

Alternate Hypothesis: Mean Implant hardness is Different across Methods (at least one mean is different from others)

H1: Not all means are equal.

Step2: Significance level

Alpha:

$\alpha = 0.05$

Step3: Calculate Test Statistics & Critical Value/P Value

Results:

Alloy 1:

```
One-way Anova Test stats Alloy_1: 6.263326635486233
One-way Anova P value Alloy_1: 0.004163412167505543
```

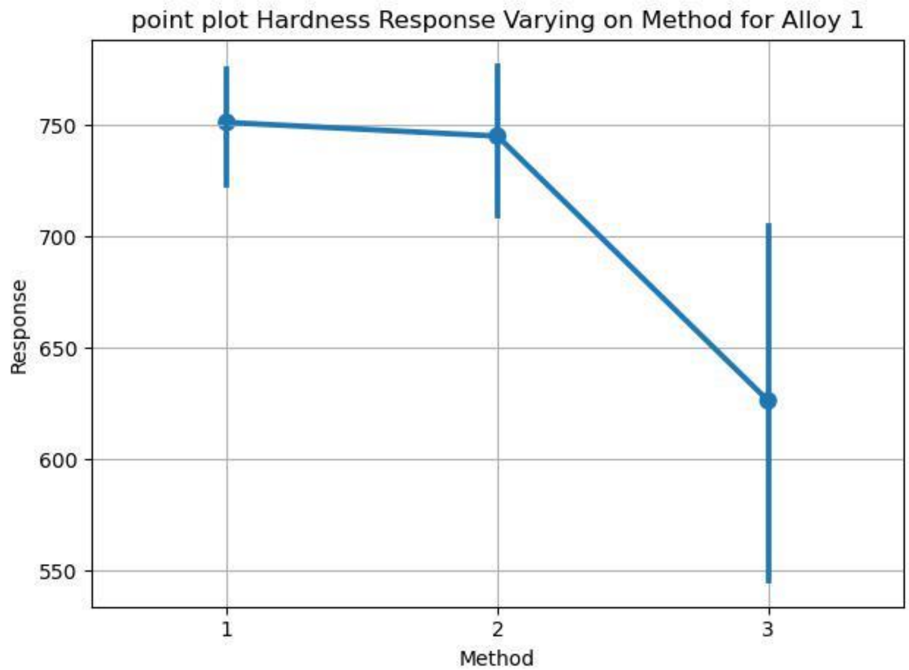


Fig 4.4 Point plot Hardness Response varying on Methods for Alloy 1.

Alloy 2:

```
One-way Anova Test stats Alloy_2: 16.41079988438482
One-way Anova P value Alloy_2: 5.415871051443187e-06
```

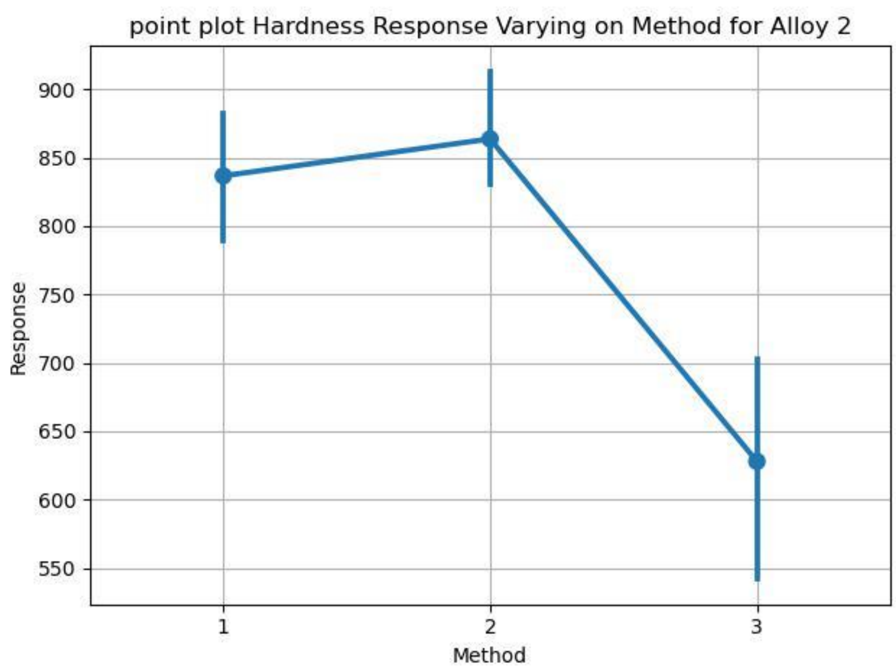


Fig 4.5 Point plot Hardness Response varying on Methods for Alloy 2.

Step 4: Conclusion

Alloy 1:

- P Value (0.004) < alpha (0.05)
- Reject Null Hypothesis (Ho)
- Sufficient evidence that there is significant difference between the mean implant hardness varying with Methods.

Alloy 2:

- P Value (5.4×10^{-6}) < alpha (0.05)
- Reject Null Hypothesis (Ho)
- Sufficient evidence that there is significant difference between the mean implant hardness varying with Methods.

Further Analyzing which Method, the mean is significantly different using Tukey's Comparison:

Alloy 1:

```
 Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====================================================
group1 group2  meandiff p-adj    lower    upper  reject
-----------------------------------------------------
     1      2   -6.1333  0.987  -102.714  90.4473  False
     1      3    -124.8 0.0085 -221.3807 -28.2193   True
     2      3 -118.6667 0.0128 -215.2473  -22.086   True
-----------------------------------------------------
```

Mean difference for Method 3 is high for Alloy 1 - resulting Low P-Value

Fig 4.6 Tukey's Comparison for Alloy 1

Tukey's hsd test for comparison states that mean implant hardness significantly different between

- Method 1 and 3
- Method 2 and 3
- Mean difference is high with Method 3 when compared with Method 1 and Method 2

Alloy 2:

```
 Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====================================================
group1 group2 meandiff p-adj    lower     upper  reject
-----------------------------------------------------
     1      2     27.0 0.8212  -82.4546  136.4546  False
     1      3   -208.8 0.0001 -318.2546  -99.3454   True
     2      3   -235.8    0.0 -345.2546 -126.3454   True
-----------------------------------------------------
```

Mean difference for Method 3 is high for Alloy 2 - resulting Low P-Value

Fig 4.7 Tukey's Comparison for Alloy 2

Tukey's hsd test for comparison states that mean implant hardness significantly different between

- Method 1 and 3
- Method 2 and 3
- Mean difference is high with Method 3 when compared with Method 1 and Method 2

Visualization under section 3 also demonstrates the same

4.3 WHAT IS THE INTERACTION EFFECT BETWEEN THE DENTIST AND METHOD ON THE HARDNESS OF DENTAL IMPLANTS FOR EACH TYPE OF ALLOY?

Step1: Define the Hypothesis

Null Hypothesis: Interaction effect between Dentist and Method does not exist.
Alternate Hypothesis: Interaction effect between Dentist and Method exist.

Step2: Significance level

Alpha:

$\alpha = 0.05$

Step3: Calculate Test Statistics & Critical Value/P Value

Results:

Alloy 1:

```
****************************************************************************************************
Interaction between Dentist and Method on the Hardness for Alloy_1
****************************************************************************************************
                        degree of freedom    sum square    mean square    F stats       P value
-------------------  -------------------  ------------  -------------  ---------  -------------
C(Dentist)                             4        106684        26670.9    3.89964    0.0114843
C(Method)                              2        148472        74236.1   10.8543     0.000284041
C(Dentist):C(Method)                   8        185941        23242.7    3.39838    0.00679275
Residual                              30        205180         6839.33  nan        nan
```

p-value = 0.00679

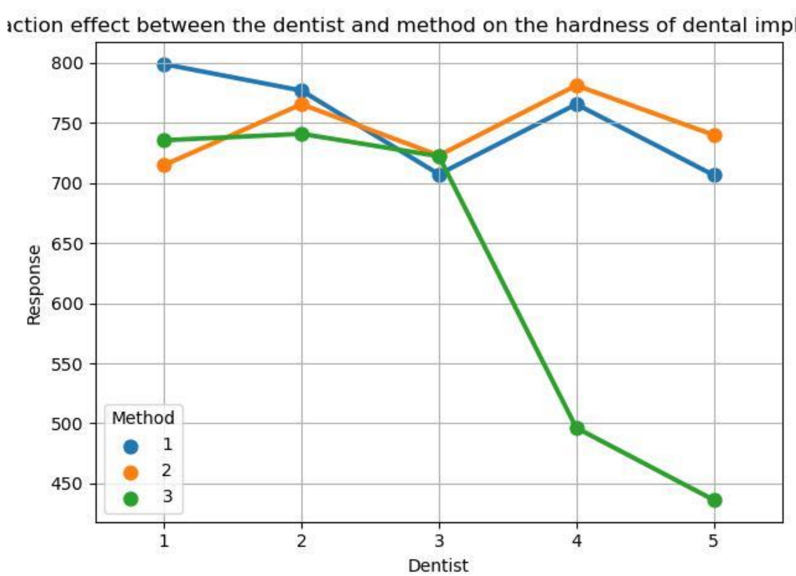


Fig 4.8 Interaction between dentist and method on hardness for Alloy 1

Alloy 2:

****

Interaction between Dentist and Method on the Hardness for Alloy_2

****

	degree of freedom	sum square	mean square	F stats	P value
C(Dentist)	4	56797.9	14199.5	1.10615	0.371833
C(Method)	2	499640	249820	19.4612	3.81425e-06
C(Dentist):C(Method)	8	197460	24682.5	1.92279	0.093234
Residual	30	385105	12836.8	nan	nan

p-value = 0.0932

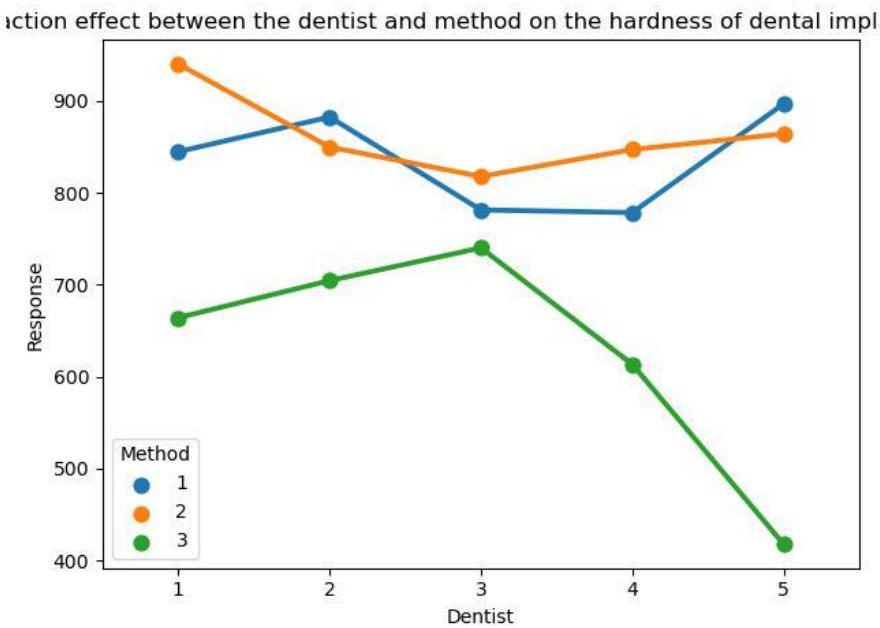


Fig 4.9 Interaction between dentist and method on hardness for Alloy 2

Step 4: Conclusion

Alloy 1:

- P value 0.00679 < alpha 0.05
- Reject Null Hypothesis
- Interaction between Dentist and Method Exist.

Alloy2:

- P Value 0.0932 > alpha 0.05
- Failed to Reject Null Hypothesis
- No significant evidence to state that Interaction between Dentist and Method DO NOT EXIST.

Plots

1.

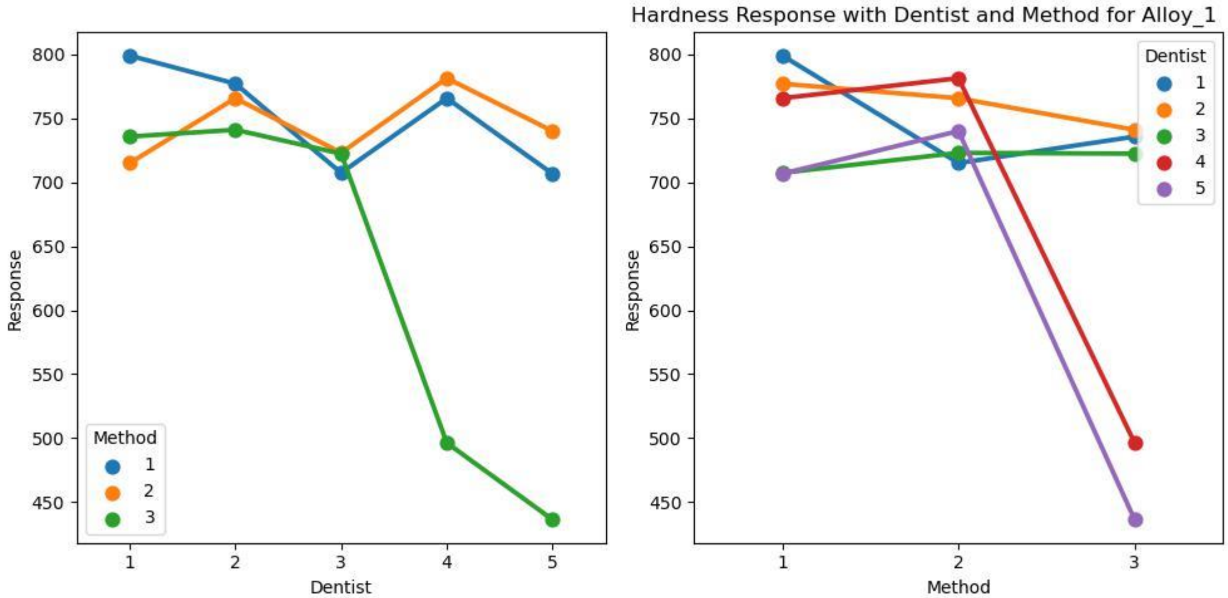


Fig 4.10 Hardness Response with Dentist and Method for Alloy 1

2.

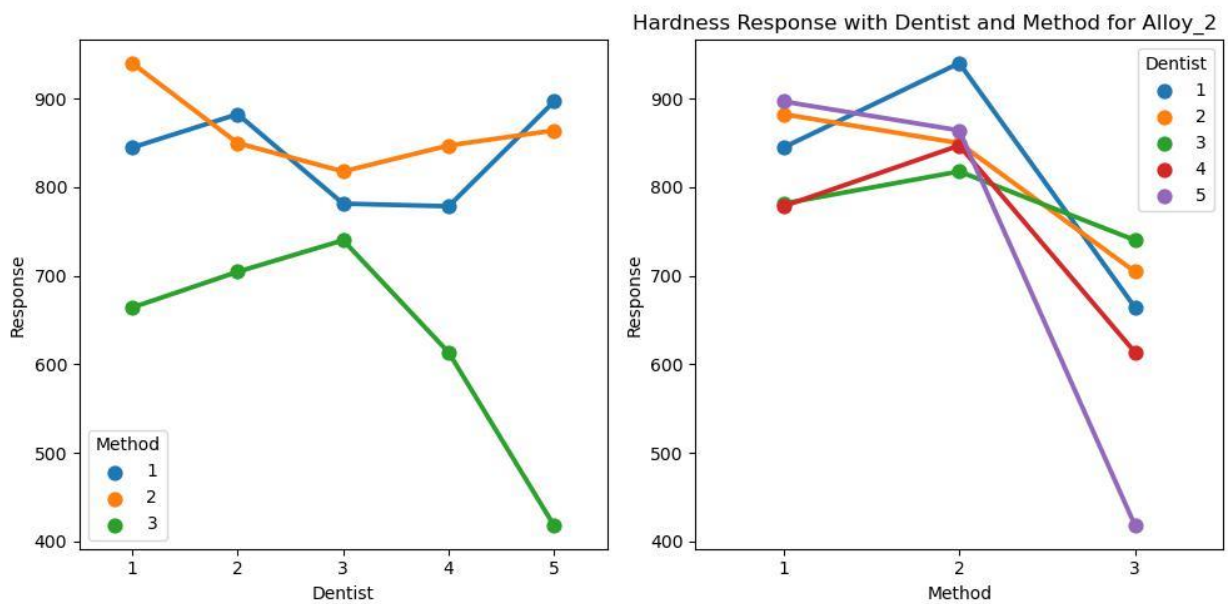


Fig 4.11 Hardness Response with Dentist and Method for Alloy 2

Inference of Interaction Plot

- Plot for Alloy 1 state interaction between all Methods and Dentists.
- Plot for Alloy 2 state interaction only between Method 1 and 2 with Dentists.
- Method 3 significantly varies as relative to Method 1 and 2 for Both Alloys.
- Combination of Dentist 4 and 5 with Method 3 significantly vary for Alloy 1.
- Combination of Dentist 5 with Method 3 significantly varies for Alloy 2.
- Combination Method 1 and 2 with any Dentist

4.4 HOW DOES THE HARDNESS OF IMPLANTS VARY DEPENDING ON DENTISTS AND METHODS TOGETHER?

Step1: Define the Hypothesis

Hypothesis for Method:

- Null Hypothesis: Mean hardness of Implants for different Method is similar.
  Ho: $\mu1 = \mu2 = \mu3$
- Alternate Hypothesis: At least one Method's Mean is different.

Hypothesis for Dentist:

- Null Hypothesis: Mean hardness of Implants for different Dentist is similar.
  Ho: $\mu1 = \mu2 = \mu3 = \mu4 = \mu5$
- Alternate Hypothesis: At least one Dentist's Mean is different.

Step2: Significance level

Alpha:

$\alpha = 0.05$

Step3: Calculate Test Statistics & Critical Value/P Value

Alloy 1:

Results:

Interaction between Dentist and Method on the Hardness for Alloy_1

	df	Sum_sq	Mean_sq	F	PR(>F)
C(Dentist)	4.0	106683.688889	26670.92	2.59	0.051875
C(Method)	2.0	148472.17777	74236.08	7.21	0.002211
Residual	38.0	391121.3777	10292.66	NaN	NaN

Alloy 2:

Results:

`Interaction between Dentist and Method on the Hardness for Alloy_2`

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

Step 4: Conclusion

Alloy1:

- Dentist:
  - P value 0.051875 > alpha 0.05
  - Failed to Reject Ho
  - No sufficient evidence to state that there is a significant difference between Mean hardness of Implant with different Dentists.
- Method:
  - P value 0.002211 < alpha 0.05
  - Reject Ho
  - Sufficient evidence to state that there is significant difference between Mean hardness of implant with different Methods.

Alloy 2:

- Dentist:
  - P value 0.458933 > alpha 0.05
  - Failed to Reject Ho
  - No sufficient evidence to state that there is a significant difference between Mean hardness of Implant with different Dentists.
- Method:
  - P value 0.000008 < alpha 0.05
  - Reject Ho
  - Sufficient evidence to state that there is significant difference between Mean hardness of implant with different Methods.

which dentists and methods combinations are different:

- For both Alloy1 and Alloy2, there is NO significant difference in Mean Implant Hardness among different Dentists.
- For both Alloy1 and Alloy2, there is sufficient evidence to state that there is significant difference in Mean implant Hardness among different Methods.
    - Conducting Tukey's HSD Comparison

Analyzing for which Method, the mean is significantly different using Tukey's Comparison:

Alloy 1:

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
1	2	-6.1333	0.987	-102.714	90.4473	False
1	3	-124.8	0.0085	-221.3807	-28.2193	True
2	3	-118.6667	0.0128	-215.2473	-22.086	True

Mean difference for Method 3 is high for Alloy 1 - resulting Low P-Value

Fig 4.12 Tukey's Comparison for Alloy 1

Tukey's hsd test for comparison states that mean implant hardness significantly different between

- Method 1 and 3
- Method 2 and 3
- Mean difference is high with Method 3 when compared with Method 1 and Method 2

Alloy 2:

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
1	2	27.0	0.8212	-82.4546	136.4546	False
1	3	-208.8	0.0001	-318.2546	-99.3454	True
2	3	-235.8	0.0	-345.2546	-126.3454	True

Mean difference for Method 3 is high for Alloy 2 - resulting Low P-Value

Fig 4.13 Tukey's Comparison for Alloy 2

Tukey's hsd test for comparison states that mean implant hardness significantly different between

- Method 1 and 3
- Method 2 and 3
- Mean difference is high with Method 3 when compared with Method 1 and Method 2

Which interaction levels are different?

- Alloy 1:
    - Interaction between Dentist and Method:
        - P value 0.006793 < alpha 0.05
        - Interaction between Dentist and Method Exist.
    - Interaction between Dentist and Temp:
        - P value 0.862862 > alpha 0.05
        - Interaction between Dentist and Temp DOES NOT Exist.
    - Interaction between Temp and Method:
        - P value P value 0.898357 > alpha 0.05
        - Interaction between Method and Temp DOES NOT Exist.

- Alloy 2:
    - Interaction between Dentist and Method:
        - P value 0.093234 > alpha 0.05
        - Interaction between Dentist and Method DOES NOT Exist.
    - Interaction between Dentist and Temp:
        - P value 0.825318 > alpha 0.05
        - Interaction between Dentist and Temp DOES NOT Exist.
    - Interaction between Temp and Method:
        - P value P value 0.675983 > alpha 0.05
        - Interaction between Method and Temp DOES NOT Exist.