Inferential Statistics

Coded Project

Anirudh Sardiwal

28 January 2024

Table of Contents

1.0 Problem 1 - Football

	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?

$$P = \frac{\text{Total number of players Injured}}{\text{Total number of Players}} = \frac{145}{235} = 0.6170 = 61.7\,\%$$

Therefore the probability that a randomly chosen player would suffer an injury is ***61.7%***

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

P(F) = Player is a Forward

P(W) = Player is a Winger

P(F or W) = P(F U W) = P(F) + P(W) — for independent groups

$$P(F) = \frac{\text{Total number of Forwards}}{\text{Total number of Players}} = \frac{94}{235} = 0.40 = 40.00\,\%$$

$$P(W) = \frac{\text{Total number of Wingers}}{\text{Total number of Players}} = \frac{29}{235} = 0.1234 = 12.34\,\%$$

$$P(F\,or\,W) = 40.00\ +\ 12.34 = 52.34\,\%$$

Therefore the probability that a player is a Forward or a Winger is ***52.34%***

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

$$P = \frac{\text{Number of Injured Strikers}}{\text{Total number of Players}} = \frac{45}{235} = 0.1914 = 19.14\,\%$$

Probability that a player is a Striker and has an Injury is ***19.14%***

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

In this case the sample space will be total number of injured players.

$$P = \frac{\text{Number of Injured Strikers}}{\text{Total number of injured Players}} = \frac{45}{145} = 0.3103 = 31.03\,\%$$

Therefore the probability that a randomly chosen injured player is a Striker is ***31.03%***

2.0 Problem 2 - Gunny Bags

Given:

Population Mean μ = 5 kg/sq.cm

Population Standard Deviation σ = 1.5 kg/sq.cm.

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

The proportion of gunny bags having a breaking strength of less than 3.17 kg per sq cm is **11.12%.**

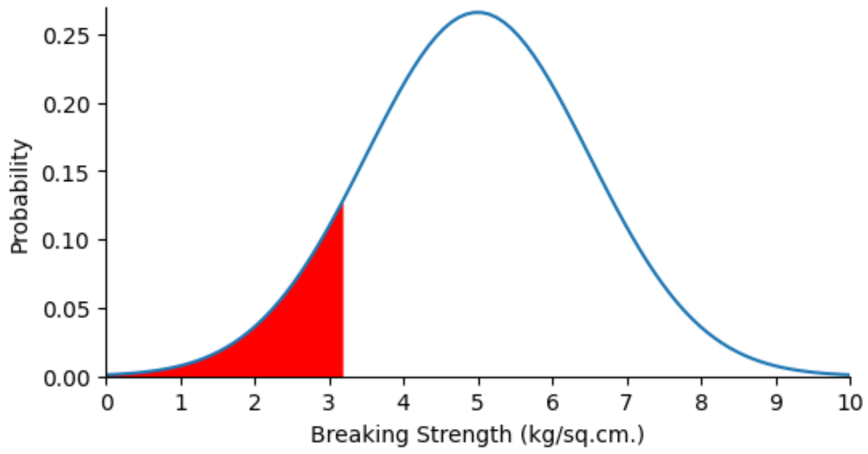

Fig. 1 - Proportion of Gunny Bags below 3.17 kg/sq.cm. Breaking Strength

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 having a breaking strength of at least 3.6 kg per sq cm is **82.47%.**

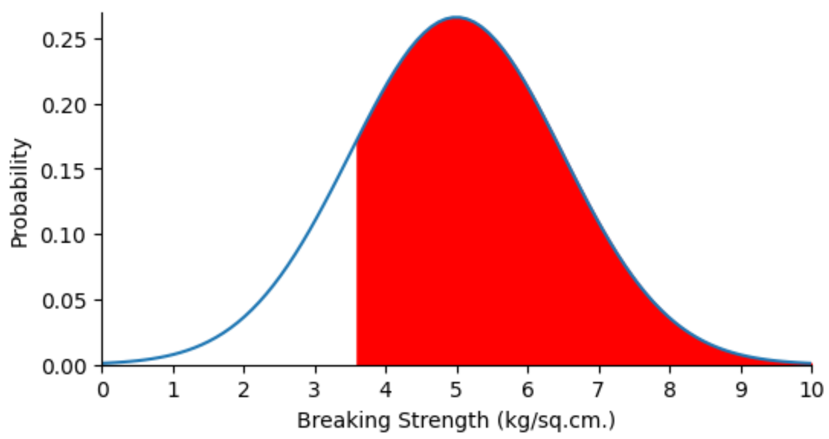

Fig. 2 - Proportion of Gunny Bags above 3.6 kg/sq.cm. Breaking Strength

2.3 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 a breaking strength between 5 and 5.5 kg per sq cm is **13.06%.**

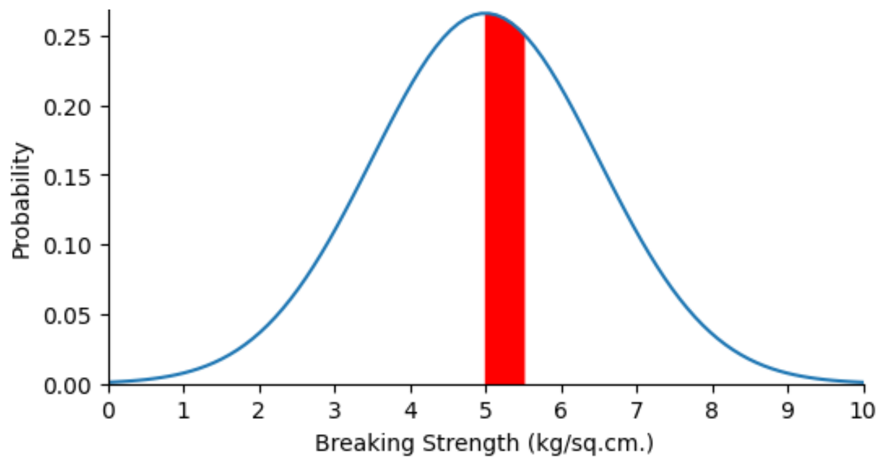

Fig. 3 - Proportion of Gunny Bags having Breaking Strength between 5 and 5.5 kg/sq.cm.

2.4 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 a breaking strength NOT between 3 and 7.5 kg per sq cm is **13.9%.**

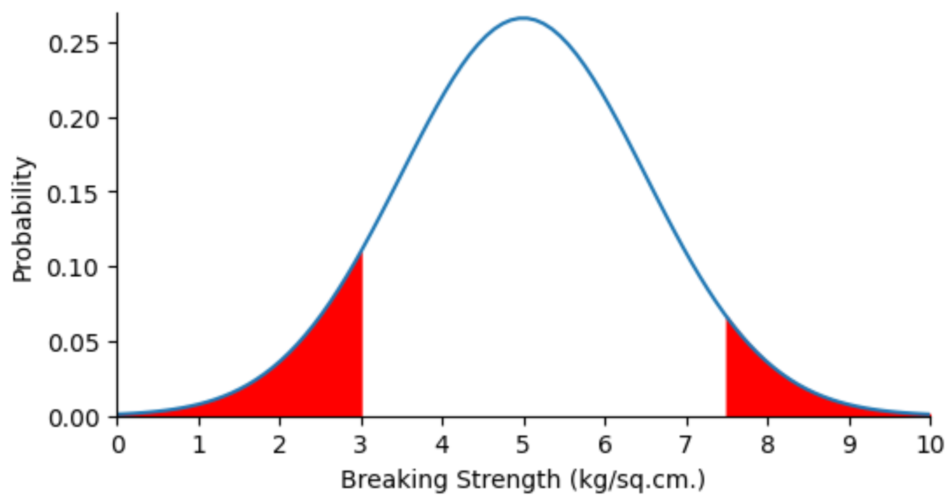

Fig. 4 - Proportion of Gunny Bags NOT between 3 and 7.5 kg/sq.cm.

3.0 Problem 3 - Zingaro

Given : For polishing the minimum Hardness Index required is 150

$\mu = 150 \qquad \alpha = 0.05$

The data consists of 75 rows and 2 columns and gives the stone hardness for Polished or Treated and Polished stones.

	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

Fig. 5 - Head of Zingaro data

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?

$\alpha = 0.05$

H_0 : Stone Hardness ≥ 150 (Stone Hardness is greater than or equal to 150)

H_1 : Stone Hardness < 150 (Stone Hardness is less than 150)

For this problem we will use a One Sample, One Tailed t-test.

The t-statistic comes out to be -4.16 and the p-value is 0.0001.

Since p-value $< \alpha$, we reject the Null Hypothesis. The stone hardness is 150 and therefore Unpolished stones are not suitable for printing.

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

$\alpha = 0.05$

$H_0 : \mu_{polished} = \mu_{unpolished}$

(The mean hardness of Polished and Unpolished stones is the same)

$H_1 : \mu_{polished} \neq \mu_{unpolished}$

(The mean hardness of Polished and Unpolished stones is different)

For this problem we will use a Two Sample, Two Tailed t-test.

The t-statistic comes out to be -3.24 and the p-value is 0.0015.

Since this is a two-tailed test, the p-value will be divided by 2.

The adjusted p-value is 0.0007.

Since p-value $< \alpha$, we reject the Null Hypothesis. The mean hardness of Polished and Unpolished stones is significantly different.

4.0 Problem 4 - Dental Implants

4.1 Introduction

	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

Fig. 6 - Top 5 rows of Dental Implants data

Dental Implants data has 90 rows and 5 columns. The dependent variable is Response, which is the hardness of dental implants. Indepependent variables are Dentist and Method. There are 5 types of Dentists and 3 types of Methods. There are 2 types of Alloys for which the calculations will be different. The variable Temp is redundant for this problem.

4.2 Visualizing the Effects

4.2.1 Alloy 1

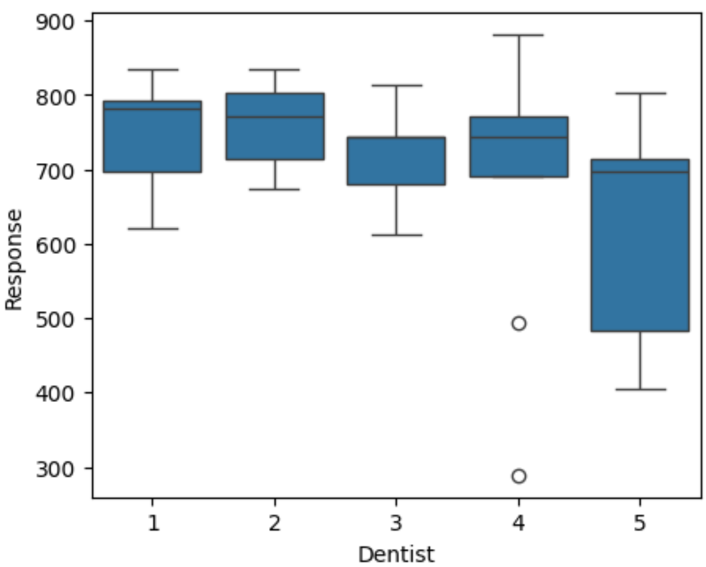

Fig. 7 - Effect of Dentist on Response for Alloy 1

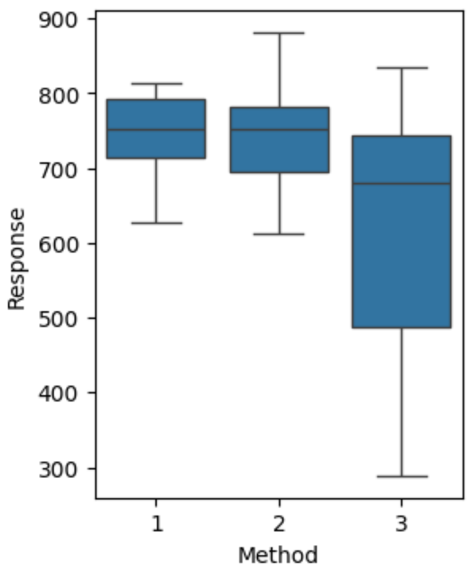

Fig. 9 - Effect of Method on Response for Alloy 1

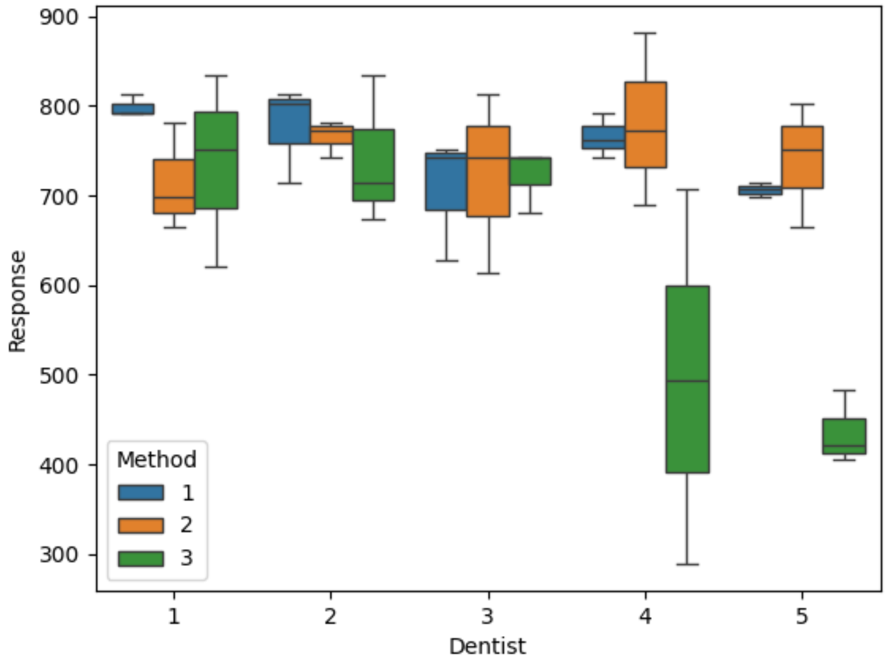

Fig. 10 - Effect of Method by Dentist on Response for Alloy 1

4.2.2 Alloy 2

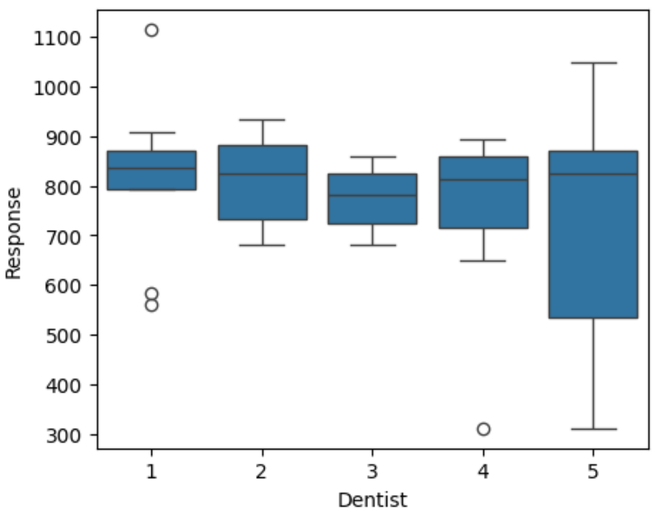

Fig. 7 - Effect of Dentist on Response for Alloy 2

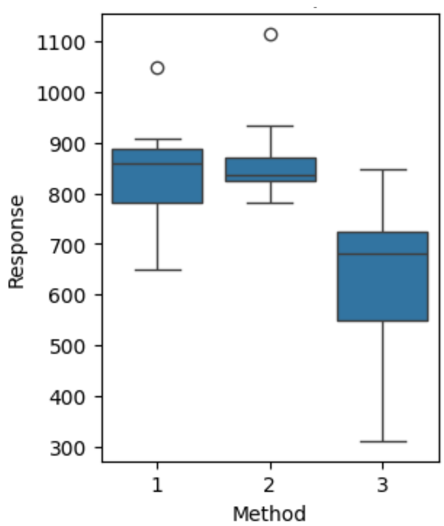

Fig. 7 - Effect of Method on Response for Alloy 2

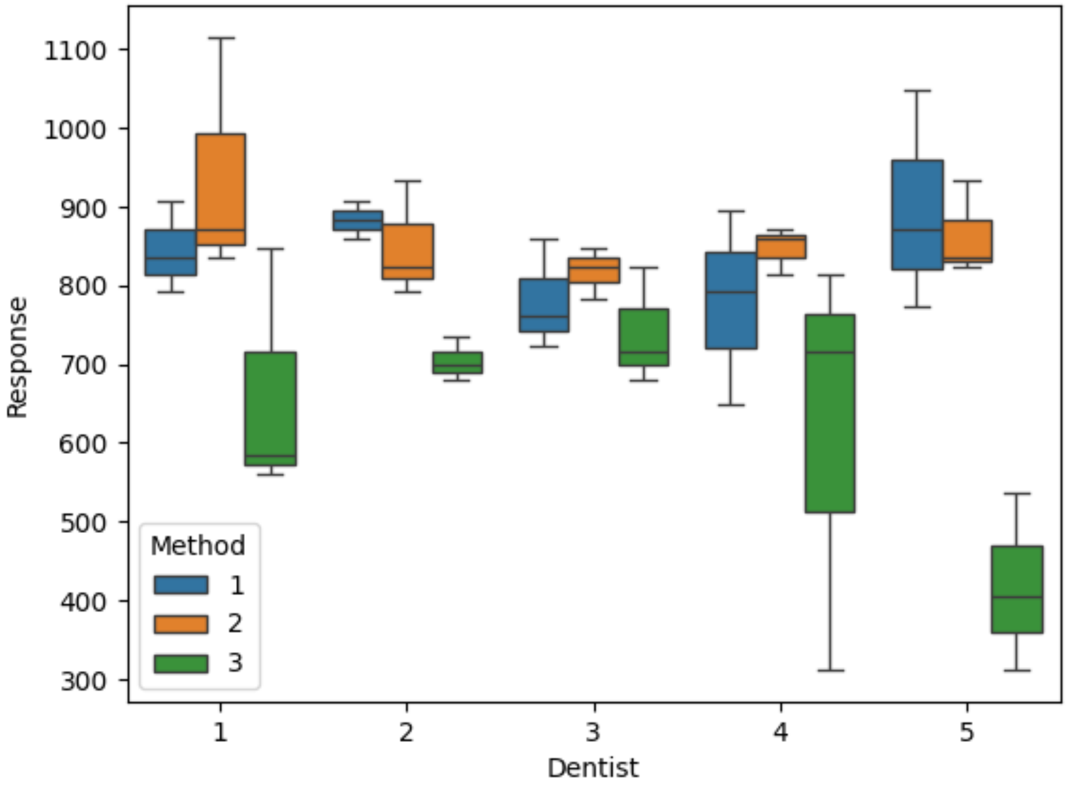

Fig. 13 - Effect of Method by Dentist on Response for Alloy 2

4.3 Assumption Testing

4.3.1 Shapiro-Wilk's Test for Normality

H_0 : Values of hardness in the Response variable follow a Normal distribution.

H_1 : Values of hardness in the Response variable do not follow a Normal distribution.

Level of Significance $\alpha = 0.05$

The p-value comes out to be 1.945×10^{-5}

Since p-value $< \alpha$, we reject the Null hypothesis. The dependent variable Response does not follow a Normal distribution.

4.3.2 Levene's test for Homogeneity of Variance

For Dentist factor

H_0 : Variances of Responses from all Dentists are equal.

H_1 : At least one variance is different from the rest.

Level of Significance $\alpha = 0.05$

The p-value comes out to be 0.0079

Since p-value $< \alpha$, we reject the Null hypothesis. Not all Dentist Response variances are equal. At least one variance is different from the rest.

For Method factor

H_0 : Variances of Responses from all Methods are equal.

H_1 : At least one variance is different from the rest.

Level of Significance $\alpha = 0.05$

The p-value comes out to be 0.0041

Since p-value $< \alpha$, we reject the Null hypothesis. Not all Method Response variances are equal. At least one variance is different from the rest.

4.4 Questions

4.4.1 How does the hardness of implants vary depending on Dentists?

Alloy 1

$H_0 : \mu_{D1} = \mu_{D2} = \mu_{D3} = \mu_{D4} = \mu_{D5}$ The means of hardness of implants by all Dentists are equal for Alloy 1.

H_1 : At least one mean is different from the rest.

Level of Significance $\alpha = 0.05$

	Source	SS	DF	MS	F	p-unc	np2
0	Dentist	106683.688889	4	26670.922222	1.977112	0.116567	0.165074
1	Within	539593.555556	40	13489.838889	NaN	NaN	NaN

The p-value comes out to be 0.12

Since p-value $> \alpha$, we fail to reject the Null hypothesis. All means of hardness of implants by all Dentists are equal for Alloy 1.

Alloy 2

$H_0 : \mu_{D1} = \mu_{D2} = \mu_{D3} = \mu_{D4} = \mu_{D5}$ The means of hardness of implants by all Dentists are equal for Alloy 2.

H_1 : At least one mean is different from the rest.

Level of Significance $\alpha = 0.05$

	Source	SS	DF	MS	F	p-unc	np2
0	Dentist	5.679791e+04	4	14199.477778	0.524835	0.718031	0.049866
1	Within	1.082205e+06	40	27055.122222	NaN	NaN	NaN

The p-value comes out to be 0.72

Since p-value $> \alpha$, we fail to reject the Null hypothesis. All means of hardness of implants by all Dentists are equal for Alloy 2.

4.4.2 How does the hardness of implants vary depending on Methods?

Alloy 1

$H_0 : \mu_{M1} = \mu_{M2} = \mu_{M3}$ The means of hardness of implants by all Methods are equal for Alloy 1.

H_1 : At least one mean is different from the rest.

Level of Significance $\alpha = 0.05$

	Source	SS	DF	MS	F	p-unc	np2
0	Method	148472.177778	2	74236.088889	6.263327	0.004163	0.229734
1	Within	497805.066667	42	11852.501587	NaN	NaN	NaN

Fig. 14 - Anova test results for Method factor

The p-value comes out to be 0.004

Since p-value $< \alpha$, we reject the Null hypothesis. At least one mean is different from the rest for Alloy 1.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
1	2	-6.1333	0.987	-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

Fig. 14a - Tukey HSD test for means of Response by Methods for Alloy 1

Pairs 1-3 and 2-3 differ in their means for Alloy 1.

Alloy 2

$H_0 : \mu_{M1} = \mu_{M2} = \mu_{M3}$ The means of hardness of implants by all Methods are equal for Alloy 2.

H_1 : At least one mean is different from the rest.

Level of Significance $\alpha = 0.05$

	Source	SS	DF	MS	F	p-unc	np2
0	Method	499640.4	2	249820.200000	16.4108	0.000005	0.438665
1	Within	639362.4	42	15222.914286	NaN	NaN	NaN

Fig. 15 - Anova test results for Dentist factor

The p-value comes out to be 0.000005

Since p-value $< \alpha$, we reject the Null hypothesis. At least one mean is different from the rest for Alloy 2.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
1	2	27.0	0.8212	-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

Fig. 15a - Tukey HSD test for means of Response by Methods for Alloy 2

Pairs 1-3 and 2-3 differ in their means for Alloy 1.

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

Alloy 1

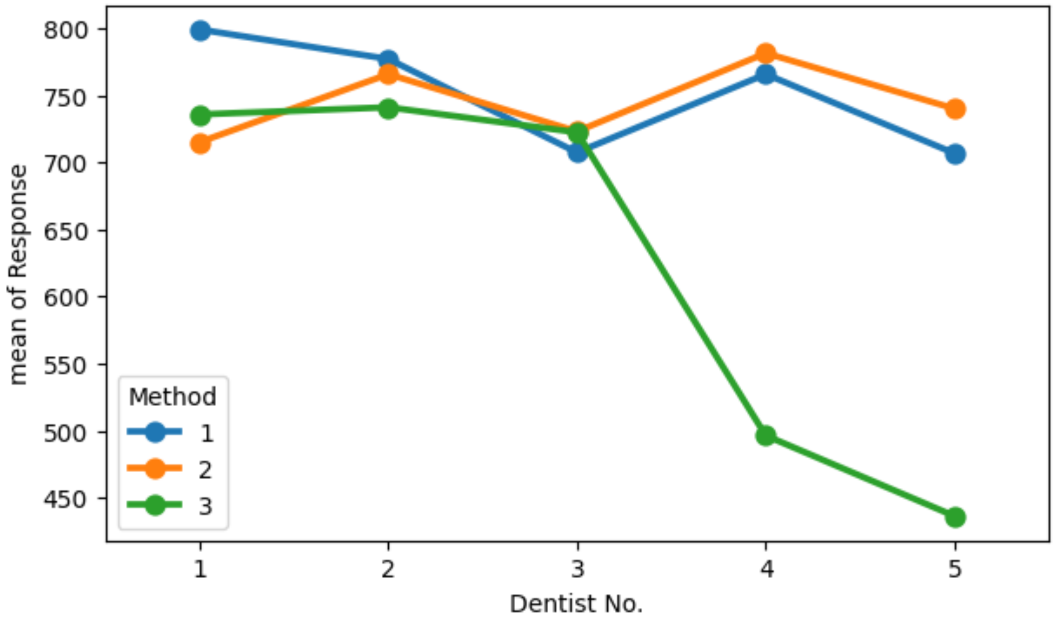

Fig. 16 - Interaction plot for Alloy 1

Inferences

- The mean levels of hardness fall mostly between 700 and 800.
- The only exceptions are with method 3 for dentists 4 and 5. That means dentists 4 and 5 are using a particular method which is significantly reducing the mean hardness of the dental implants.
- There is greater variation in the Response from dentists 4 and 5. Dentist 3 is able to maintain the hardness in a close range with any method.

Alloy 2

Inferences

- Overall there's greater variation between means of Response in Alloy 2.
- The min. and max. means are respectively lower and higher than Alloy 1.

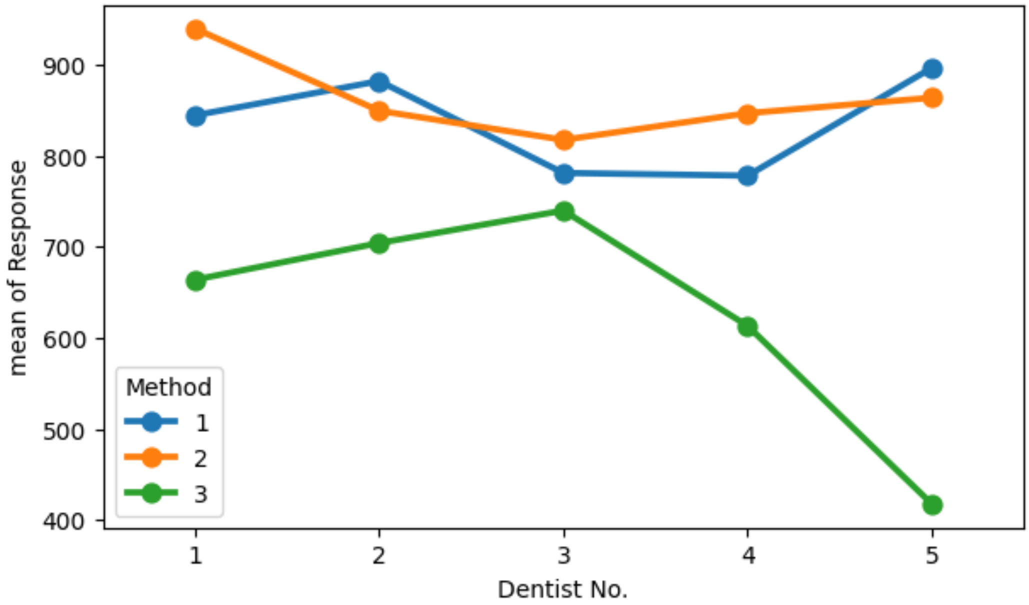

Fig. 17 - Interaction plot for Alloy 2

- The is greatest variation in the Response is from dentists 5. On the contrary, Dentist 3 is able to maintain the hardness in a close range with any method.
- Methods 1 and 2 seem all right, but there seems to be some issues with Method 3, because it's means is varying considerably depending on the Dentist.

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

Alloy 1

H_0 : Interaction between Dentists and Methods is having no effect on Response for Alloy 1.

H_1 : Interaction between Dentists and Methods is having at least some effect on Response for Alloy 1.

Level of Significance $\alpha = 0.05$

	Source	SS	DF	MS	F	p-unc	np2
0	Dentist	106683.688889	4	26670.922222	3.899638	0.011484	0.342084
1	Method	148472.177778	2	74236.088889	10.854287	0.000284	0.419825
2	Dentist * Method	185941.377778	8	23242.672222	3.398383	0.006793	0.475406
3	Residual	205180.000000	30	6839.333333	NaN	NaN	NaN

Fig. 18 - Anova result for the interaction between Methods and Dentists for Alloy 1

The p-value comes out to be 0.007

Since p-value $< \alpha$, we reject the Null hypothesis. Interaction between Dentists and Methods has an effect on the Response for Alloy 1.

```
  Multiple Comparison of Means - Tukey HSD, FWER=0.05
==========================================================
group1 group2  meandiff p-adj    lower     upper   reject
----------------------------------------------------------
  M1D1   M1D2     -22.0    1.0 -270.8283  226.8283  False
  M1D1   M1D3  -91.6667 0.9853   -340.495  157.1617  False
  M1D1   M1D4  -33.3333    1.0 -282.1617   215.495  False
  M1D1   M1D5  -92.3333 0.9844 -341.1617   156.495  False
  M1D1   M2D1     -84.0 0.9933 -332.8283  164.8283  False
  M1D1   M2D2  -33.3333    1.0 -282.1617   215.495  False
  M1D1   M2D3     -76.0 0.9975 -324.8283  172.8283  False
  M1D1   M2D4  -17.6667    1.0   -266.495  231.1617  False
  M1D1   M2D5     -59.0 0.9998 -307.8283  189.8283  False
  M1D1   M3D1  -63.3333 0.9996 -312.1617   185.495  False
  M1D1   M3D2     -58.0 0.9999 -306.8283  190.8283  False
  M1D1   M3D3  -76.6667 0.9972   -325.495  172.1617  False
  M1D1   M3D4 -302.6667  0.007   -551.495  -53.8383   True
  M1D1   M3D5 -362.6667 0.0007   -611.495 -113.8383   True
  M1D2   M1D3  -69.6667  0.999   -318.495  179.1617  False
  M1D2   M1D4  -11.3333    1.0 -260.1617   237.495  False
```

Fig. 19 - Pairwise Tukey HSD results for Alloy 1 (in part)

Pairs

- **M3D5 with** M1D1, M1D2, M1D3, M1D4, M1D5, M2D1, M2D2, M2D3, M2D4, M2D5, M3D1, M3D2, M3D3 and
- **M3D4 with** M1D1, M1D2, M1D4, M2D2 , M2D4

are having an effect on Response variable.

Alloy 2

H_0 : Interaction between Dentists and Methods is having no effect on Response for Alloy 2.

H_1 : Interaction between Dentists and Methods is having at least some effect on Response for Alloy 2.

Level of Significance $\alpha = 0.05$

	Source	SS	DF	MS	F	p-unc	np2
0	Dentist	56797.911111	4	14199.477778	1.106152	0.371833	0.128530
1	Method	499640.400000	2	249820.200000	19.461218	0.000004	0.564728
2	Dentist * Method	197459.822222	8	24682.477778	1.922787	0.093234	0.338949
3	Residual	385104.666667	30	12836.822222	NaN	NaN	NaN

Fig. 20 - Anova result for interaction between Methods and Dentists for Alloy 2

The p-value comes out to be 0.093

Since p-value $>$ α, we fail to reject the Null hypothesis. Interaction between Dentists and Methods does not have a significant effect on Response for Alloy 2.