Inferential Statistics Project - Business Report

By: Aaryani Kadiyala

PGP-Data Science and Business Analytics

(PGPDSBA.O.JAN24.A)

Table of Contents

List of Tables :

List of Figures:

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: Data for Problem 1

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

Ans: P(Randomly Chosen Player would suffer an injury) = Total Player Injured/Total Player
P(Randomly Chosen Player would suffer an injury) = 145/235 = 0.6170212765957447
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?

Ans: P(Player is a forward or a winger) = P(Forward)+P(Winger)
P(Player is a forward or a winger) = (94/235 + 29/235) = 123/235 = 0.5234042553191489
Probability that a player is a forward or a winger is 52.3%

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

Ans: P (A randomly chosen player plays in a striker position and has a foot injury) =
P (Striker ∩ Foot Injury) = 45/235 = 0.19148936170212766
Probability that a randomly chosen player plays in a striker position and has a foot injury
Is 19.1%.

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

Ans: P (A randomly chosen injured player is a striker) = P (Striker | Injured)
= P (Striker ∩ Foot Injury)/P(Injured)
= (45/235) / (145/235) = 0.3103448275862069
The probability that a randomly chosen injured player is a striker is 31.0%

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)

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

Ans: z-score = (3.17-5)/1.5 = -1.21

Using the standard distribution table, the proportion to the left of z-score -1.21 is 0.1131.

Therefore, the proportion of the gunny bags have a breaking strength less than 3.17 kg per sq cm is 0.1131 or approximately 11.31%.

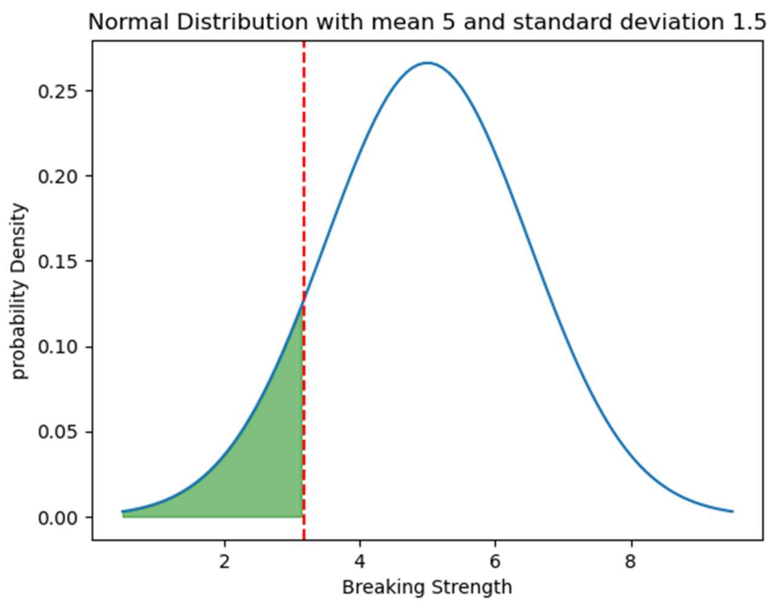

Figure 1 : Visual Representation of 2.1

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

Ans: z-score = (3.6-5)/1.5 = -0.93

Using the standard distribution table, the proportion to the left of z-score -0.93 is 0.1762

Therefore, the proportion of the gunny bags have a breaking strength at least 3.6 kg per sq cm Is 1-0.1762 = 0.8238 or approximately 82.38%.

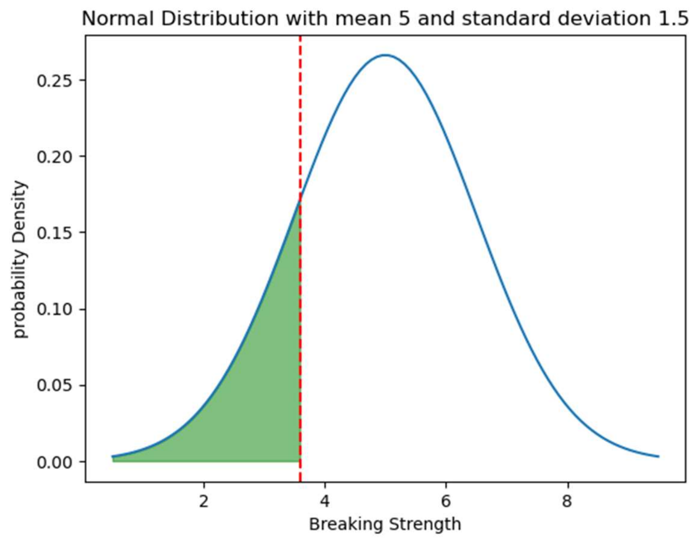

Figure 2 : Visual Representation of 2.2

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

Ans: z1 = (5-5)/1.5 = 0

z2 = (5.5-5)/1.5 = 0.33

Using the standard normal distribution table, the area/proportion to the left of z-score 0.33 is 0.6293 and the area/proportion to the left of z-score 0 is 0.5.

Therefore, the proportion of the gunny bags have a breaking strength between 5 and 5.5 kg per sq cm is 0.6293 – 0.5 = 0.1293 or approximately 12.93%.

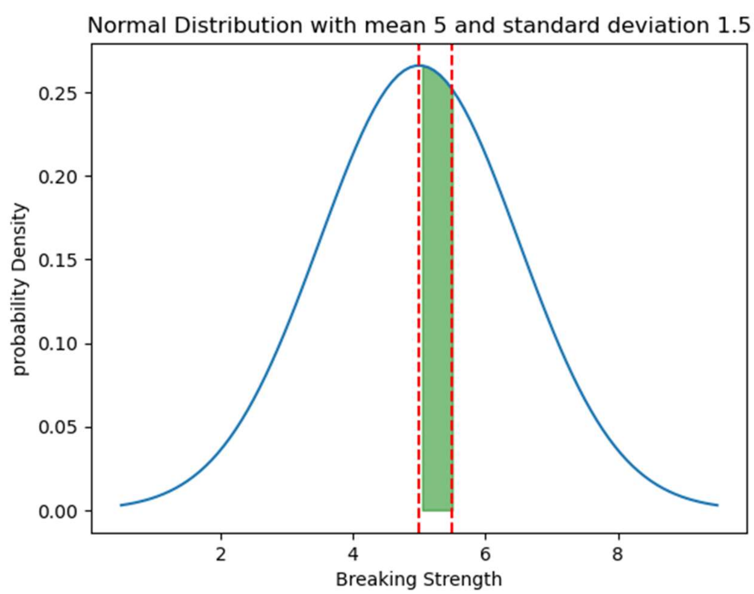

Figure 3 : Visual Representation of 2.3

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

Ans: z1 = (3-5)/1.5 = -1.33

z2 = (7.5-5)/1.5 = 1.67

Using the standard normal distribution table, the area/proportion to the left of z-score -1.33 is 0.0918 and the area/proportion to the left of z-score 1.67 is 0.9525.

Therefore, the proportion of the gunny bags have a breaking strength is between 3 and 7.5 kg per sq cm is 0.9525 – 0.0918 = 0.8607 or approximately 86.07%.

So, the proportion of the gunny bags have a breaking strength is not between 3 and 7.5 kg per sq cm is 1 – 0.8607 = 0.1393 or approximately 13.93%.

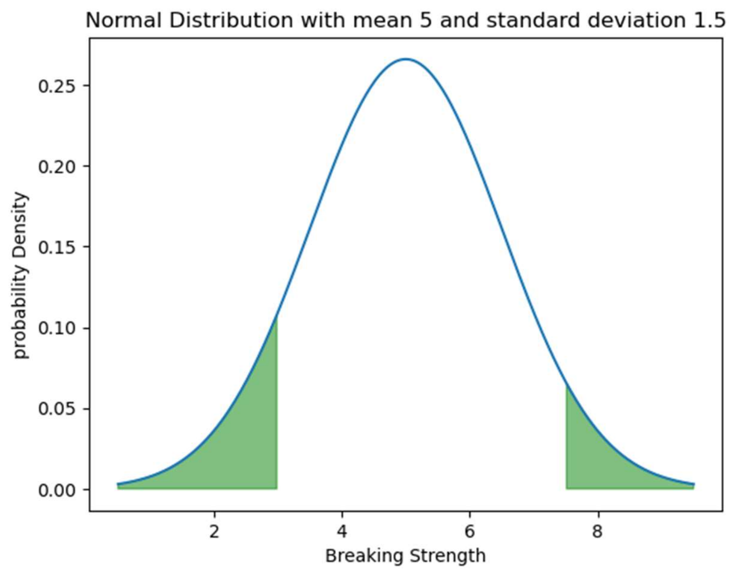

Figure 4 : Visual Representation of 2.4

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);

3.1 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?

Ans: Zingaro thinks that unpolished stone is not suitable for printing that means average: mu Brinell's hardness index less than 150 (For printing it should be at least 150).

Null hypothesis: Unpolished stone is suitable for printing: H0: Mu >= 150

Alternative Hypothesis: Unpolished stone is not suitable for printing: H1: Mu < 150

Sample average: X= 134.1105

Level of significance: 0.05

Standard deviation of sample:33.041804

Degree of Freedom:74

Identify the test statistic:

Assumptions:

- We have two samples, considering one for the treatment
- Sample sizes for both samples are same.
- Both are independent and Continuous in nature
- The large sample, n > 30. So, we use the T test
- Both the variables are normally distributed

T-stat value: -4.854269, One -sample t-test p value: 3.036694197077475e-06

Conclusion: Since p value is less than Level of significance (0.05). Hence, we have enough evidence to reject the null hypothesis in favour of alternative hypothesis

We conclude that there is significance evidence that the mean hardness of the unpolished stones is not suitable for printing

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

Ans:

Assumptions: Mp: mean hardness of polished stones

Mu: mean hardness of unpolished stones

Null hypothesis: The mean hardness of both polished and unpolished stones is same H0: Mp=Mu

Alternative Hypothesis: The mean hardness of polished and unpolished stones is not same H1: Mp > Mu or Mp! =Mu

Sample average:X1= 134.1105, X2= 147.7881

Level of significance: 0.05

Degree of Freedom: 74

Identify the test statistic:

Assumptions:

- We have two samples and their variance are not equal
- Sample sizes for both samples are same.
- Both are independent and Continuous in nature
- The large sample, n > 30. So, we will use the 2 sample T test
- Both the variables are normally distributed

t-test p-value=1.0703549280233378e-07, T-stat value: -5.718103

Conclusion: Since p value is less than level significance (0.05), Hence we reject null hypothesis. Accepts Alternate Hypothesis where there is significance evidence that Unpolished stone is not suitable for printing and also leads us to know mean hardness of the polished stones are greater than unpolished stones and are not same.

Problem 4:

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.

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

Ans: If we perform Hypothesis Test for the response Variables.

	df	sum_sq	mean_sq	F	PR(>F)
Dentist	4.0	1.577946e+05	39448.638889	1.934537	0.112066
Residual	85.0	1.733301e+06	20391.776471	NaN	NaN

Figure 5 : Output of Hypothesis Test for 4.1

We see P Value is greater than alpha(0.05). hence, we fail to reject Null Hypothesis.

Thus the Mean Hardness is same across all dentists.

Now Lets perform for Alloy1 and Alloy2

Alloy_1:

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

Figure 6: Output of Hypothesis Test for Alloy_1 for 4.1

Alloy_2:

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

Figure 7: Output of Hypothesis Test for Alloy_2 for 4,1

Now, We see that the corresponding P-value is greater than alpha (0.05). Thus, We fail to reject NULL Hypothesis.
Thus, for Both Alloy1 and Alloy2, the Mean Hardness of Alloy1 and Alloy2 is same across all Dentists.

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

Ans: Hypothesis Test for Alloy 1 data-

Null hypothesis (for Alloy 1): There is no difference in means among the Methods interms of implant hardness for Alloy 1.

Alternative hypothesis (for Alloy 1): There is a difference in means among the Methods in terms of implant hardness for Alloy 1.

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

Figure 8: Output of Hypothesis Test for Alloy_1 for 4.2

P-value is 0.004163 which is lesser than alpha i.e., 0.05. Hence, we can reject null hypothesis and consider there is a difference in means among the methods in terms of implant hardness for Alloy 1.

ANOVA Test for Alloy 2 data-

Null hypothesis (for Alloy 2): There is no difference among the methods in terms of implant hardness for Alloy 2.

Alternative hypothesis (for Alloy 2): There is a difference among the methods interms of implant hardness for Alloy 2.

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

Figure 9: Output of Hypothesis Test for Alloy_2 for 4.2

P-value is 0.000005 which is very less than alpha i.e., 0.05. Hence, we can reject null hypothesis and consider there is a difference in means among the methods in terms of implant hardness for Alloy 2.

Method 3 has variation when compared with the other two.

```
  Multiple Comparison of Means - Tukey HSD, FWER=0.05
=======================================================
group1 group2  meandiff p-adj    lower     upper   reject
-------------------------------------------------------
     1      2   10.4333 0.9415  -64.7584   85.6251  False
     1      3    -166.8    0.0 -241.9917  -91.6083   True
     2      3 -177.2333    0.0 -252.4251 -102.0416   True
-------------------------------------------------------
```

Figure 10: Output of Multiple Comparison on all methods

Based on the results above, method 3 has a significantly different impact on the hardness level for Alloy 1.

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

Ans : Null hypothesis (for Alloy 1): There is no difference among the Interaction effect between dentist and method levels in terms of implant hardness for Alloy 1.

Alternative hypothesis (for Alloy 1): There is a difference among the Interaction effect between dentist and method levels in terms of implant hardness for Alloy 1.

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

Figure 11: Output of Hypothesis Test for Alloy_1 for 4.3

P-value is 0.000221 which is lesser than alpha i.e., 0.05. Hence, we can reject null hypothesis and consider there is a difference in means among the Interaction effect between dentist and method levels in terms of implant hardness for Alloy 1.

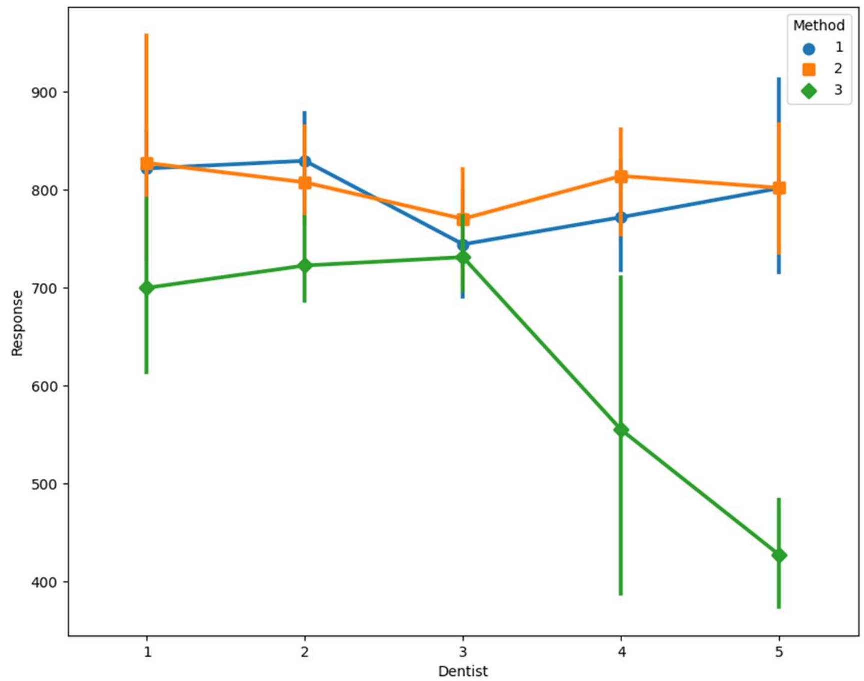

Figure 12: Interaction plot between dentist and method - alloy 1

Hypothesis Test for Alloy 2 data-

Null hypothesis (for Alloy 2): There is no difference in means among the Interaction effect between dentist and method levels in terms of implant hardness for Alloy 2.

Alternative hypothesis (for Alloy 2): There is a difference in means among the Interaction effect between dentist and method levels in terms of implant hardness for Alloy 2.

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

Figure 13: Output of Hypothesis Test for Alloy_2 for 4.3

P-value is 0.000482 which is lesser than alpha i.e., 0.05. Hence, we can reject null hypothesis and consider there is significant difference in means among the Interaction effect between dentist and method levels in terms of implant hardness for Alloy 2.

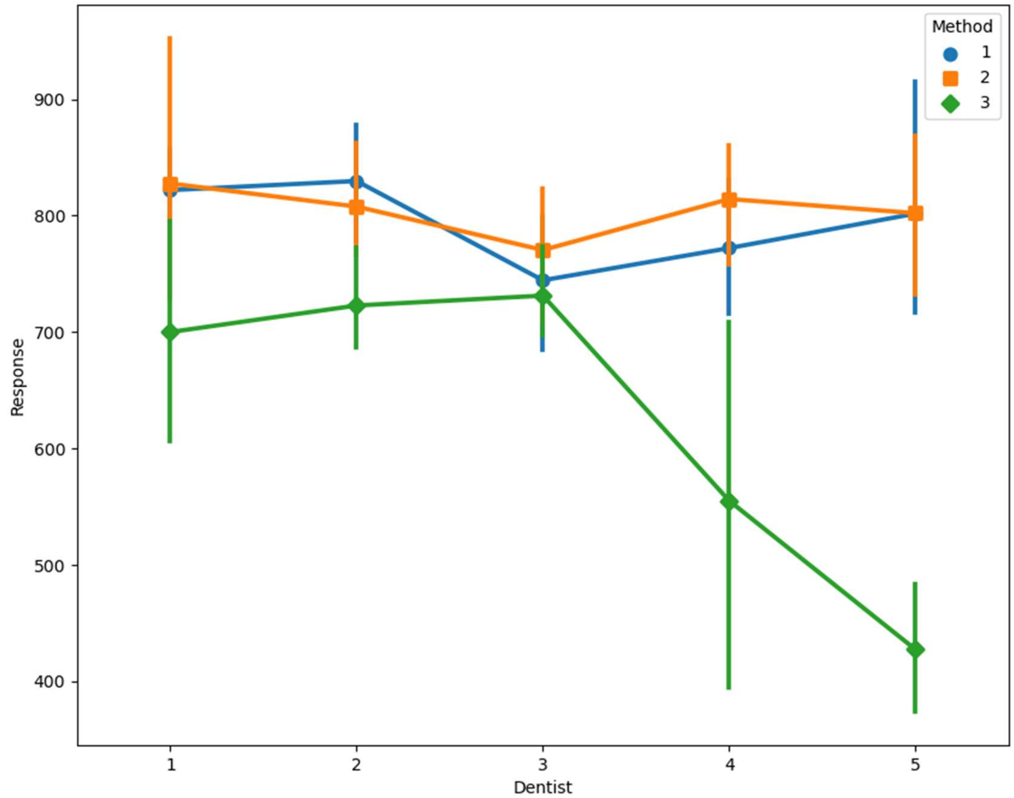

Figure 14: Interaction plot between dentist and method - alloy 2

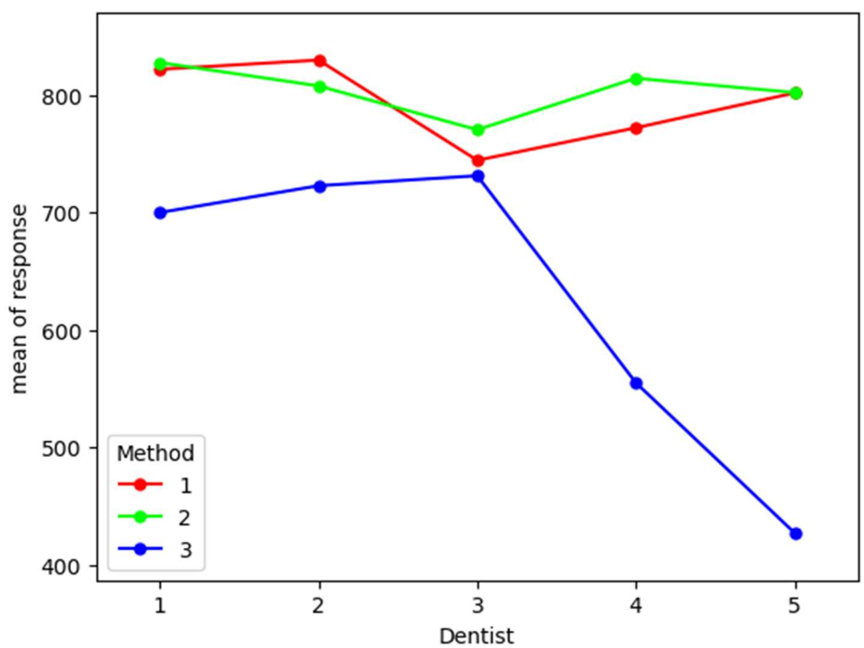

Figure 15: Interaction plot between dentist and method

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

Ans: Hypothesis Test for Alloy 1 data-

Null hypothesis (for Alloy 1): There is no difference among the factors Dentist and Method levels in terms of implant hardness for Alloy 1.

Alternative hypothesis (for Alloy 1): There is a difference among the factors Dentist and Method levels in terms of implant hardness for Alloy 1.

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

Figure 16: Output of Hypothesis Test for Alloy_1 for 4.4

P value for Method is 0.002211 which is less than 0.05. Hence, we have enough evidence to reject null hypothesis and consider that at least one pair of method means is different for Alloy 1.

This means the methods have an impact on the hardness level

From the results above, Method 3 is causing the impact on the hardness level

Hypothesis Test for Alloy 2 data-

Null hypothesis (for Alloy 2): There is no difference among the factors Dentist and Method levels in terms of implant hardness for Alloy 2.

Alternative hypothesis (for Alloy 2): There is a difference among the factors Dentist and Method levels in terms of implant hardness for Alloy 2.

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

Figure 17: Output of Hypothesis Test for Alloy_2 for 4.4

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
1	2	10.4333	0.9415	-64.7584	85.6251	False
1	3	-166.8	0.0	-241.9917	-91.6083	True
2	3	-177.2333	0.0	-252.4251	-102.0416	True

Figure 18: Output of Multiple Comparison on all methods

Method 3 is causing the impact on the hardness level.