Hypothesis Testing

Zingaro

Aquarius Health Club

Metal Implants in Dental Cavities

APOORV P

JITENDRA S

KASHYAP

PREYAL C

RATHI S

SHARATH A

PGP-AIML June’23

22/10/2023

Contents

Executive Summary

Zingaro

Zingaro stone printing is a company that specializes in printing images or patterns on polished or unpolished stones. 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. They have made some claims about the hardness quality which we will be proving in this document.

Aquarius Health Club

Aquarius Health Club, one of the largest and most popular crossfit 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 5 more push-ups as compared to when he/she enrolled in the program. We will be helping Aquarius Health Club to find out whether their program was successful or not using the data provided by them.

Metal Implants in Dental Cavities

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. We have been given such data with responses of metal implants and have been asked to test whether there is a difference on metal implants for different dentists and different methods.

Introduction

Through this exercise we will be performing different tests of Hypothesis for each problem and try to prove against the Null Hypothesis. For each dataset we will look at its distribution and basic characteristics before starting to test each one of them. This will help a person to understand Hypothesis Testing, different T-tests, Test of ANOVA, and basic statistics for the data.

Zingaro

Data Description

	Unpolished	Treated and Polished
47	165.948268	149.561327
30	98.865709	142.156784
63	121.407525	107.524167
43	58.981371	141.887153
10	95.349030	145.828526

Fig-1 Zingaro Sample Data

```
 0   Unpolished            75 non-null     float64
 1   Treated and Polished  75 non-null     float64
```

The shared data has 75 records and just 2 columns namely ‘Unpolished’ and ‘Treated and Polished’ which indicates the hardness of unpolished and polished stones respectively.

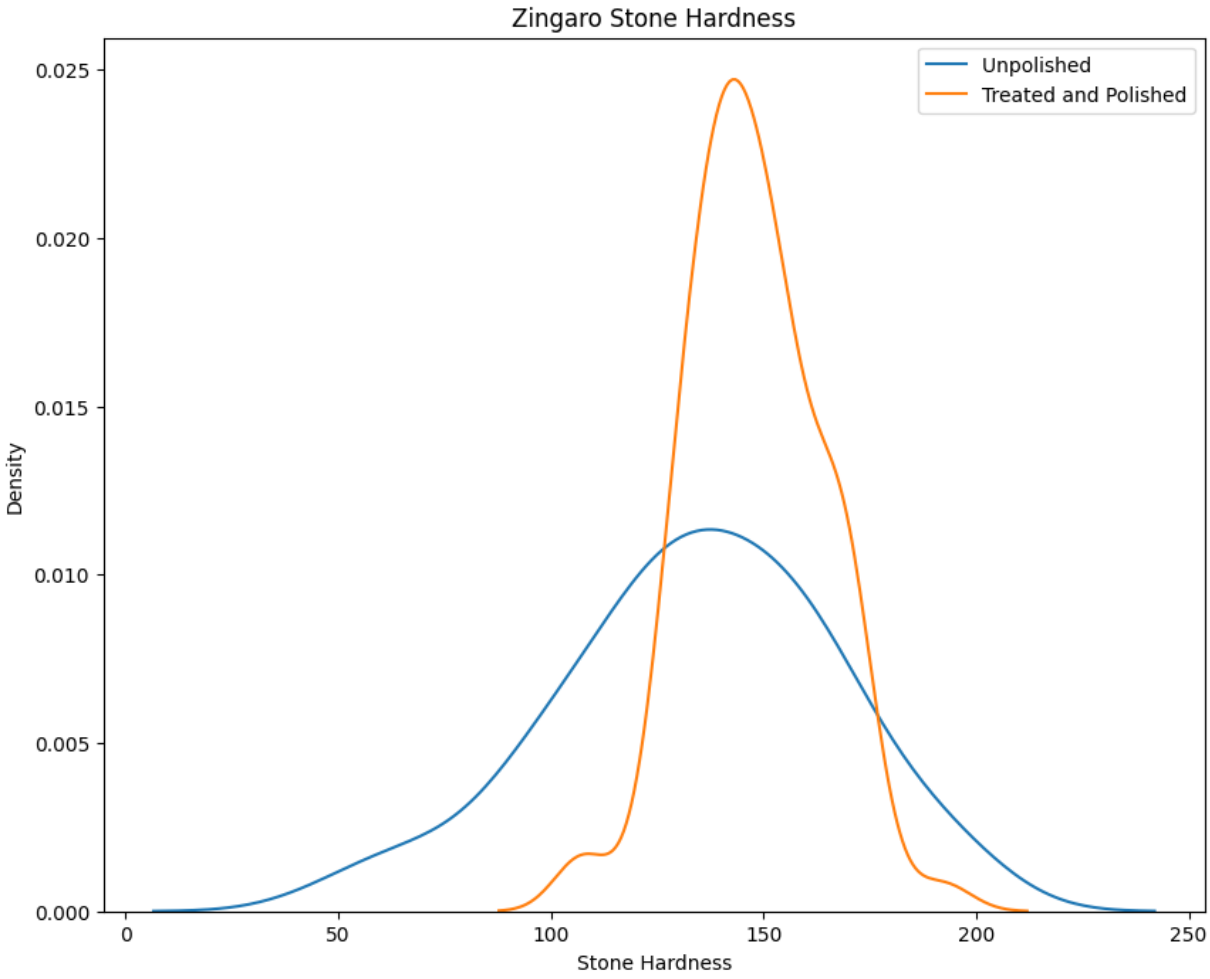

Fig-2 Zingaro Data Distribution

Both of the data follow normal distribution, with unpolished data slightly towards perfect normal.

Zingaro believes that the unpolished stones may not be suitable for printing. Do you think Zingaro is justified in thinking so?

Earlier experience of Zingaro with this particular client is favorable as the stone surface was found to be of adequate hardness. However, Zingaro has reason to believe now that the unpolished stones may not be suitable for printing.

Hypothesis Test

As the claim is that the hardness of unpolished stones is not adequate for printing, our Null hypothesis would be that the unpolished stones are alright for printing, i.e. mean to be at least 150 for Unpolished data. The Alternate hypothesis would be that the unpolished stones are not

good for printing i.e. mean to be less than 150. We need to conduct a One-Tailed One Sample t Test.

$H_0 \rightarrow \mu_{unpolished} >= 150$
$H_a \rightarrow \mu_{unpolished} < 150$

For level of significance 5%, we get the p-value of 4.1712869974196533e-05. Hence, we **Reject the Null Hypothesis**. Zingaro's claim about the unpolished stones not being of adequate hardness is correct.

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

We need to verify for Zingaro that the mean of hardness for Unpolished and Polished Stones is the same or not.

Hypothesis Test

To find this, we need to conduct a Two-Tailed Two Sample t Test. For the same, our Null hypothesis would be that both the means are equal and the Alternate hypothesis would be exact opposite of it i.e. both the means are not equal.

$H_0 \rightarrow \mu_{unpolished} = \mu_{polished}$
$H_a \rightarrow \mu_{unpolished} \neq \mu_{polished}$

For level of significance 5%, we get the p-value of 0.0014655150194628351. Hence, we **Reject the Null Hypothesis**. The mean hardness of unpolished and polished stones are not the same.

Aquarius Health Club

Data Description

	Sr no.	Before	After
79	80	23	25
20	21	35	42
71	72	25	32
67	68	17	26
52	53	31	36

Fig-3 Aquarius Health Club Sample data

```
0   Sr no.   100 non-null    int64
1   Before   100 non-null    int64
2   After    100 non-null    int64
```

The shared data has 100 records and 3 columns. Each record corresponding to one particular individual's push up counts before and after the program.

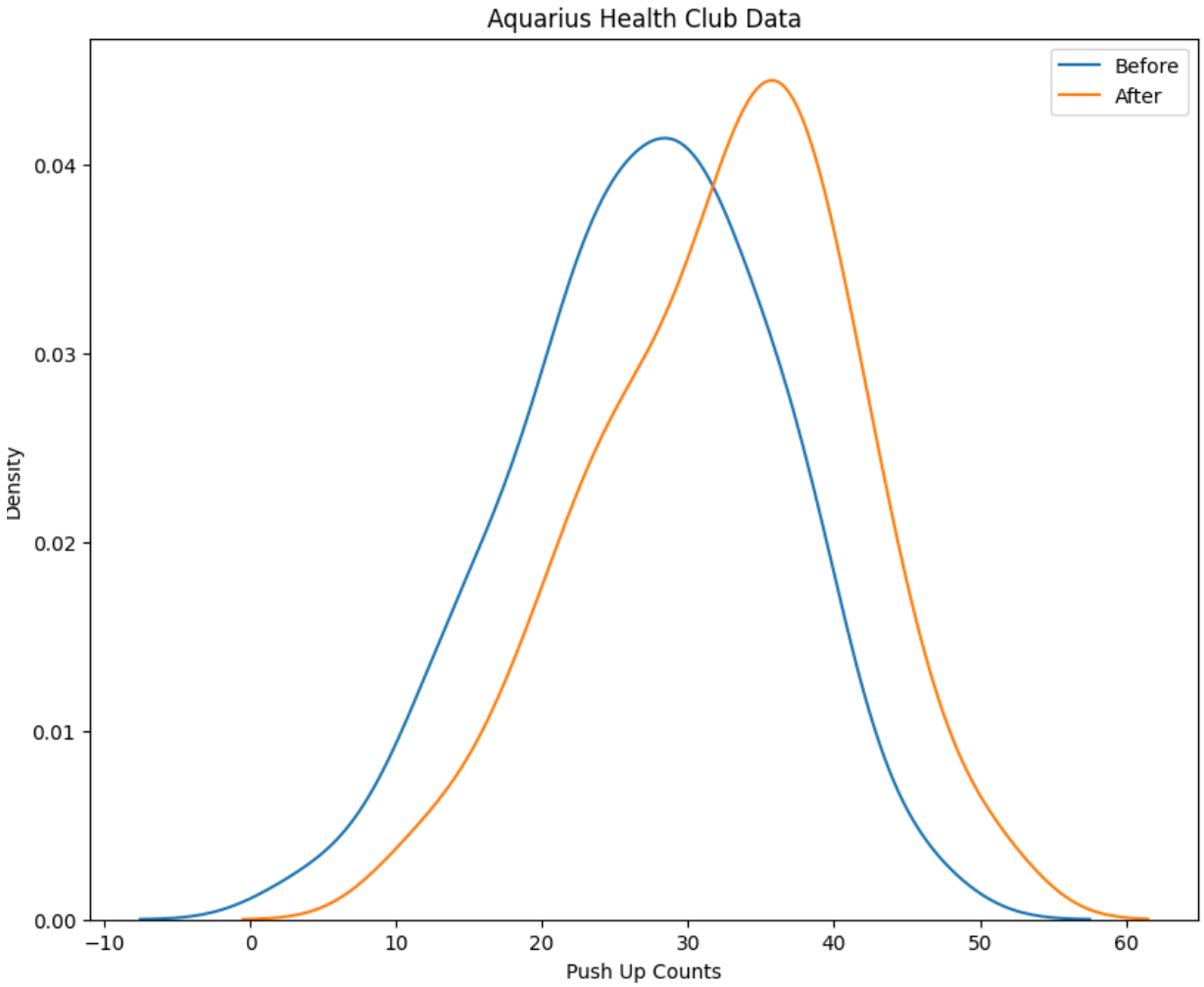

Fig-4 Aquarius Health Club Push Up Count Distribution

Before and After push up counts data follow very close to normal distribution.

Can we conclude whether the program is successful?

The program is considered successful if the candidate is able to do 5 more push-ups as compared to when he/she enrolled in the program.

This is a problem of a Paired T-test.

Hypothesis Testing

For a paired t-test, the following formula is considered:-

$\mu_{\text{before}} = \mu_{\text{after}}$

That means, there is no difference between an individual's push up counts before and after the program. Here the problem given considers a count of 5 more push ups than before is considered successful, so we can consider the above formula for paired t-test as below and then form our hypothesis.

$$\mu_{difference} = 0$$

This means the difference between after and before is 0.

Now, for the program to be successful the person's push up count should be at least 5 more than before i.e. the program is considered unsuccessful if a person is still doing 4 or less push ups. This is our Null Hypothesis.

$$H_0 \rightarrow \mu_{difference} <= 4$$
$$H_a \rightarrow \mu_{difference} > 4$$

And the Alternate Hypothesis would be the success claim of the program that the person is now doing 5 or more push ups than before.

For conducting this, we create another column in the dataset namely 'Difference' and perform a One Sample One-tailed t-test on that column.

For level of significance 5%, we get the p-value of 2.3271721890386642e-07. Hence, we **Reject the Null Hypothesis**. The program is successful as an individual can do at least 5 push ups more than what they were doing before the program.

Metal Implants in Dental Cavities

Data Description

	Dentist	Method	Alloy	Temp	Response
83	5	2	2	1700	835
41	3	1	2	1700	724
4	1	1	2	1600	792
89	5	3	2	1700	312
45	3	2	2	1500	824

Fig-4 Metal Implants Sample Data

```
 0   Dentist    90 non-null     int64
 1   Method     90 non-null     int64
 2   Alloy      90 non-null     int64
 3   Temp       90 non-null     int64
 4   Response   90 non-null     int64
```

There are 90 records and 5 columns. The Response column is the output column for metal implants and the rest 4 columns are features affecting it like Dentists, Methods, Alloys and Temperature.

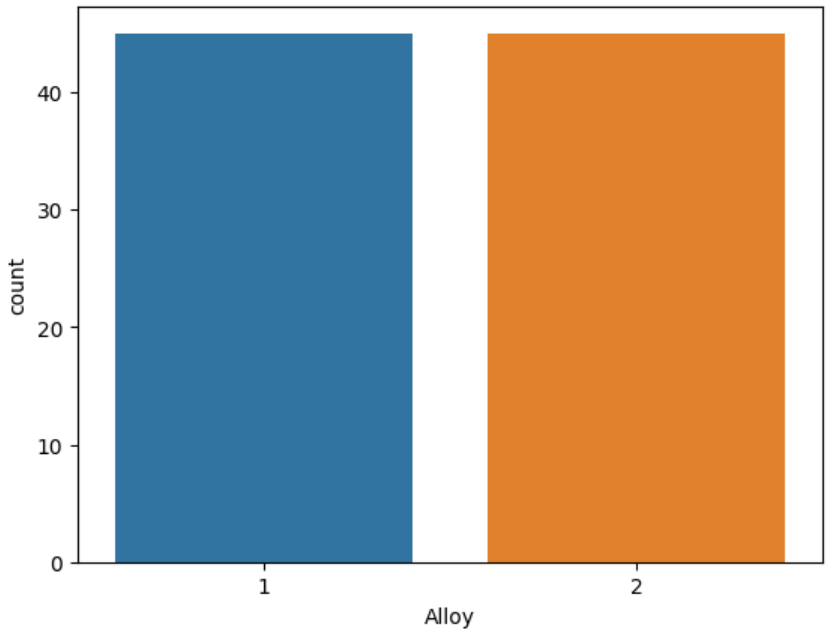

Fig-5 Metal Implants Different Alloys

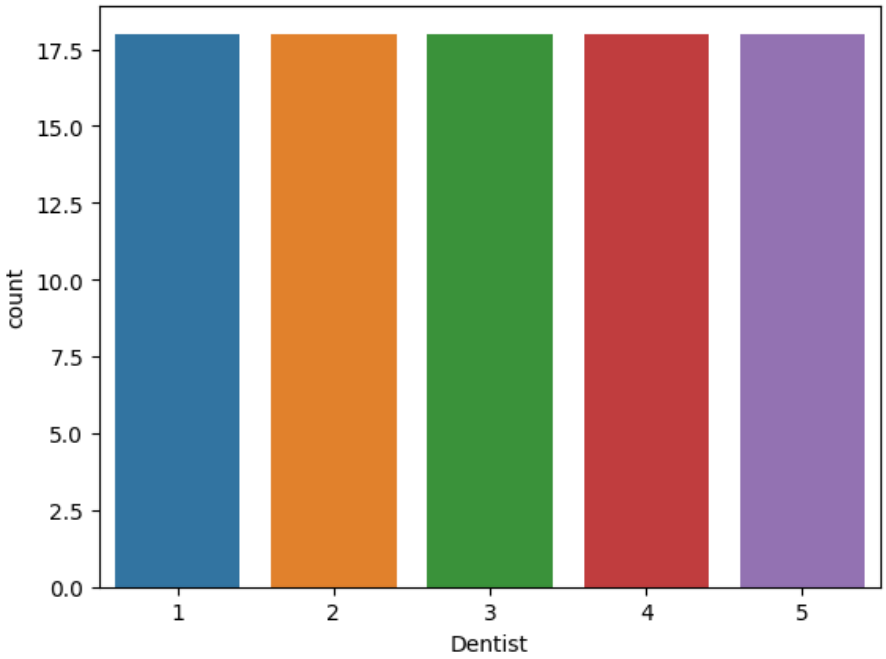

Fig-6 Metal Implants Different Dentists

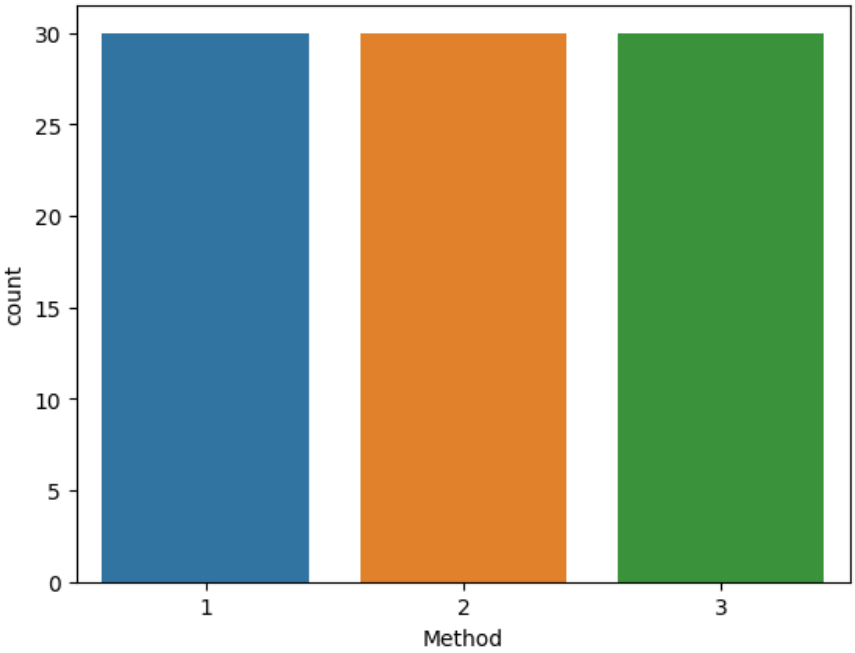

Fig-7 Metal Implants DIfferent Methods

There are 2 different alloys, 5 different dentists and 3 different methods that might or might not affect the Response variable.

Assumptions for ANOVA

1. The responses for each factor level have a normal population distribution.
2. These distributions have the same variance.
3. The data are independent.

Checking the assumptions are correct or not for Both the alloys differently, and based on different Dentists and different Methods.

Alloy 1 - Dentists

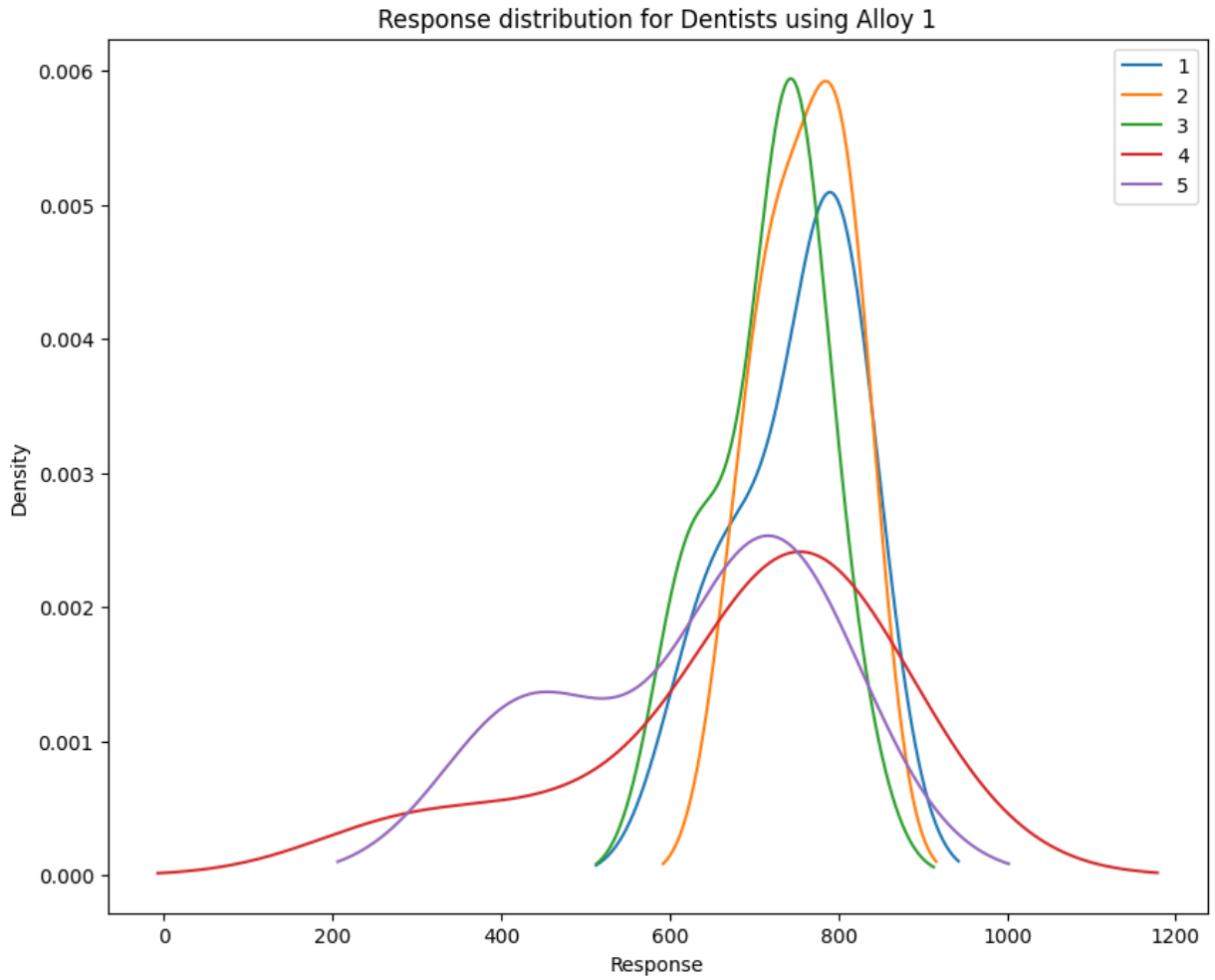

Fig-8 Response Data for Dentists when Alloy 1 is used

From the above graph it is quite clear that not all dentists data follow normal distribution for the response variable. On finding out the variances, there is a significant difference between Dentist 4 and 5 against the rest of the dentists. The sample size is also less than 30. Hence, the assumptions are not correct for Alloy 1 - Dentists data.

Alloy 2 - Dentists

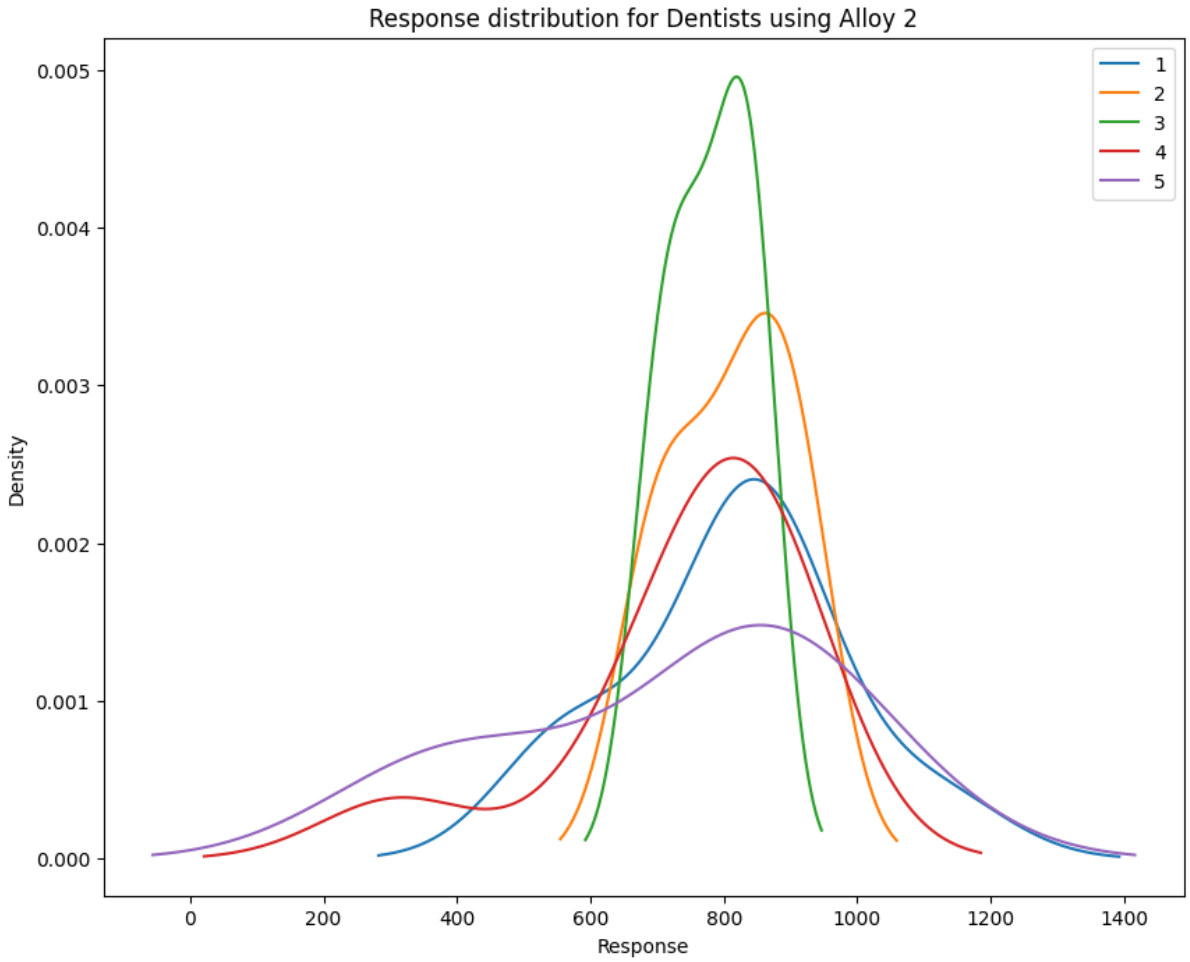

Fig-9 Response Data for Dentists when Alloy 2 is used

From the above graph it is quite clear that not all dentists data follow normal distribution for the response variable. On finding out the variances, there is a significant difference between Dentist 1, 4 and 5 against the rest of the dentists. The sample size is also less than 30. Hence, the assumptions are not correct for Alloy 2 - Dentists data.

Alloy 1 - Methods

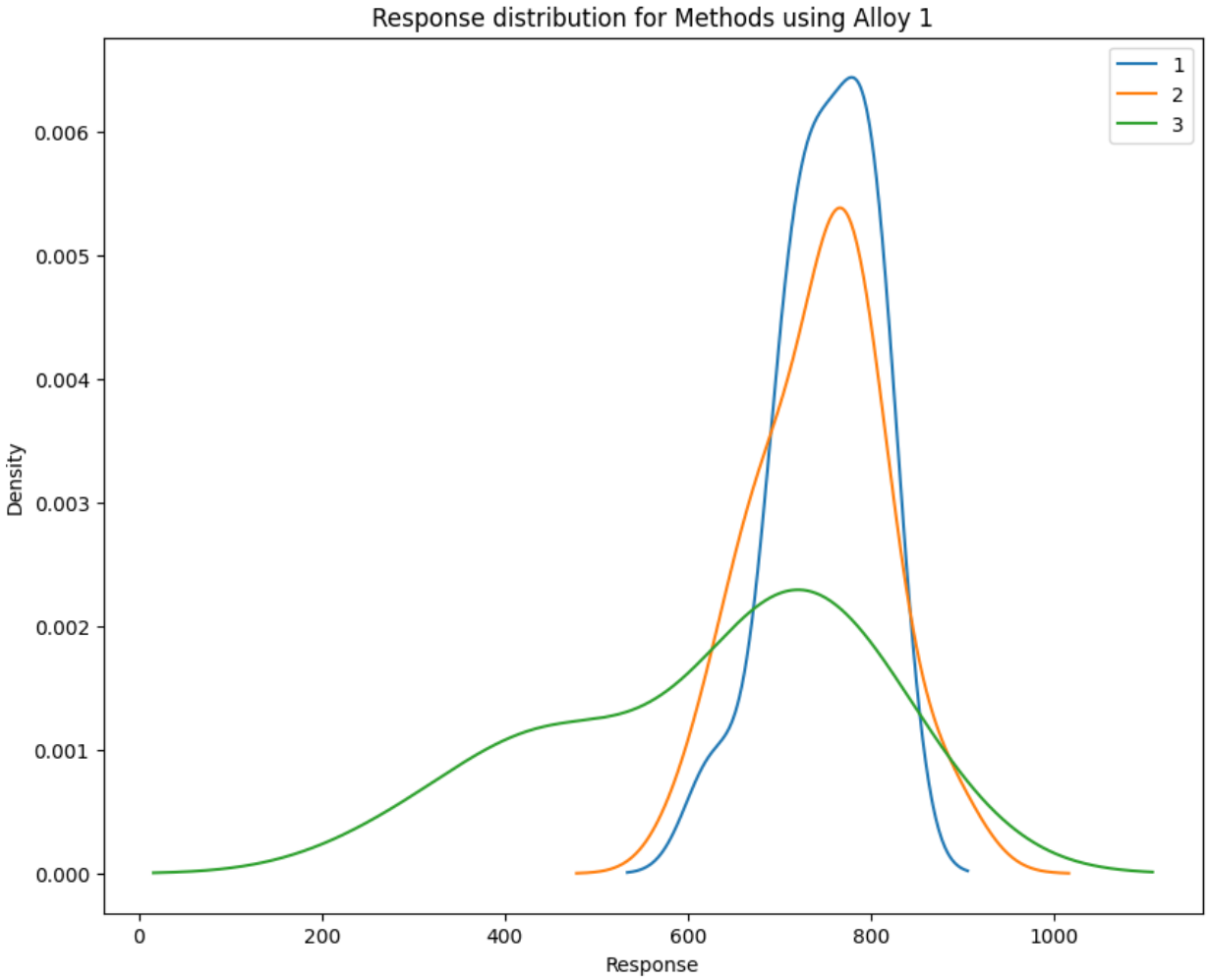

Fig-10 Response Data for Methods when Alloy 1 is used

From the above graph it is quite clear that not all methods data follow normal distribution for the response variable. On finding out the variances, there is a significant difference between Method 3 against the rest of the methods. The sample size is also less than 30. Hence, the assumptions are not correct for Alloy 1 - Methods data.

Alloy 2 - Methods

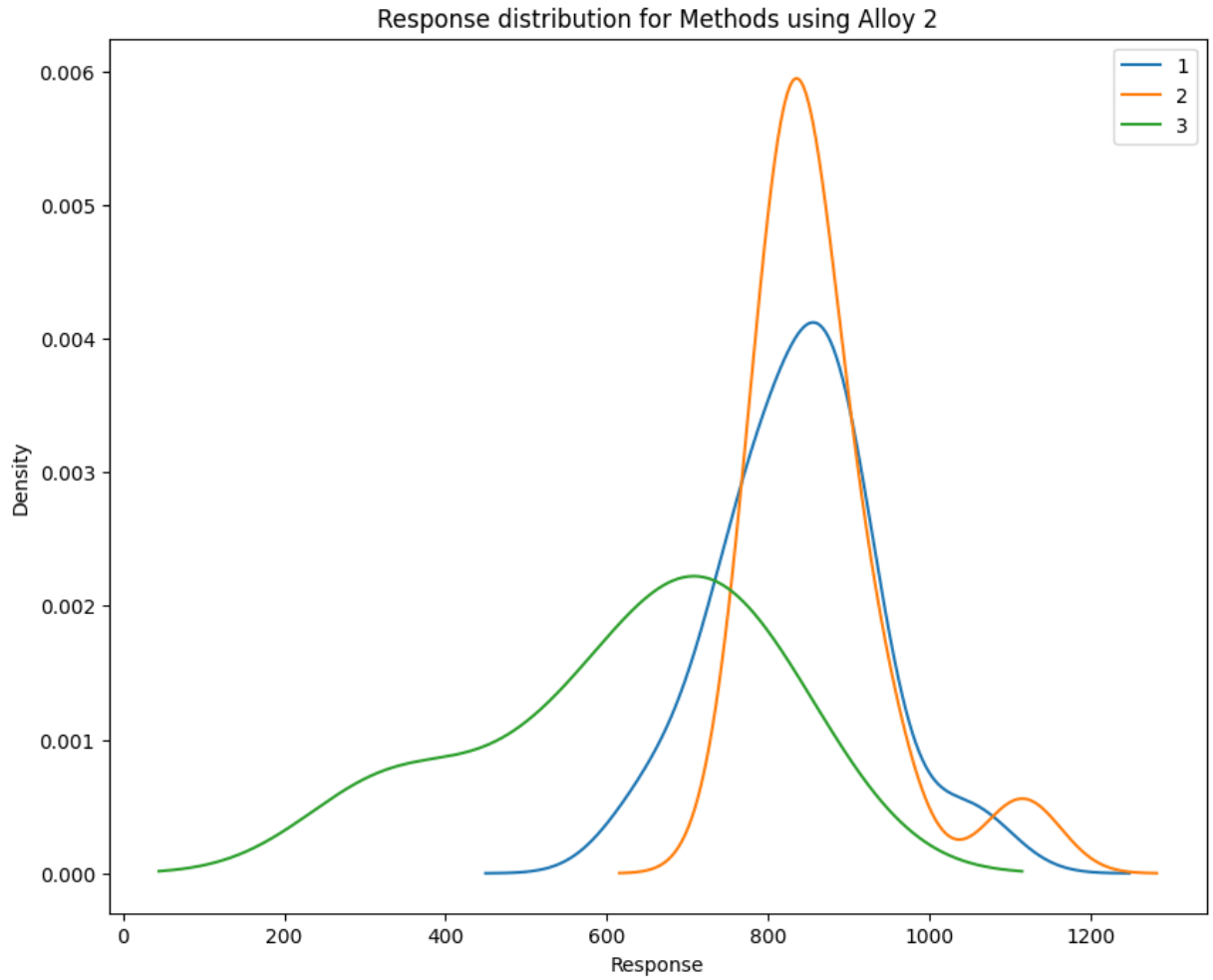

Fig-11 Response Data for Methods when Alloy 2 is used

From the above graph it is quite clear that not all methods data follow normal distribution for the response variable. On finding out the variances, there is a significant difference between Method 3 against the rest of the methods. The sample size is also less than 30. Hence, the assumptions are not correct for Alloy 2 - Methods data.

The assumptions of ANOVA failed for all the cases, but we would still proceed with the test of ANOVA as directed.

Is there any difference among the dentists on the implant hardness?

To take out the factor of alloys we will perform the test for both the alloys separately just like we checked the assumptions separately.

Hypothesis Test

The task is about the difference between the Dentists on implant hardness, so our Null hypothesis will be there is no difference among mean of Response for different dentists. The Alternate hypothesis will be that there is a difference somewhere among the dentists, it could be between a single pair not being the same or all of them are different from each other.

$H_0 \rightarrow \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

$H_a \rightarrow$ There is some difference in at least 1 pair of dentists

Alloy 1 - Dentists

	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

For level of significance 5%, we get the p-value of 0.116567. Hence, we **Failed to Reject the Null Hypothesis**. There is indeed no difference in the mean Responses for metal implants among different dentists when using Alloy 1.

Alloy 2 - Dentists

	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

For level of significance 5%, we get the p-value of 0.718031. Hence, we **Failed to Reject the Null Hypothesis**. There is indeed no difference in the mean Responses for metal implants among different dentists when using Alloy 2.

Is there any difference among the methods on the implant hardness?

To take out the factor of alloys we will perform the test for both the alloys separately just like we checked the assumptions separately.

Hypothesis Testing

The task is about the difference between the Methods on implant hardness, so our Null hypothesis will be there is no difference among mean of Response for different methods. The Alternate hypothesis will be that there is a difference somewhere among the methods, it could be between a single pair not being the same or all of them are different from each other.

$H_0 \rightarrow \mu_1 = \mu_2 = \mu_3$
$H_a \rightarrow$ There is some difference in at least 1 pair of methods

Alloy 1 - Methods

	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

For level of significance 5%, we get the p-value of 0.004163. Hence, we **Reject the Null Hypothesis**. There is a difference in the mean Responses for metal implants among different methods when using Alloy 1.

To find the pair of methods that differ we either use Post ANOVA tests like TukeyHSD or do a Two Sample T-test picking up 2 methods at a time.

Hence we found out that pairs of Methods 1-3 and 2-3 are different. So, we can conclude that using method 3 influences the response for metal implants when alloy 1 is used.

Alloy 2 - Methods

	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

For level of significance 5%, we get the p-value of 0.000005. Hence, we **Reject the Null Hypothesis**. There is a difference in the mean Responses for metal implants among different methods when using Alloy 2.

To find the pair of methods that differ we either use Post ANOVA tests like TukeyHSD or do a Two Sample T-test picking up 2 methods at a time.

Hence we found out that pairs of Methods 1-3 and 2-3 are different. So, we can conclude that using method 3 influences the response for metal implants when alloy 2 is used.

THE END