Business report

DATA SCIENCE AND BUSINESS ANALYTICS

Content:

Problem 1

1.1 What is the probability that a randomly chosen player would suffer an injury?
1.2 What is the probability that a player is a forward or a winger?
1.3 What is the probability that a randomly chosen player plays in a striker position and has a foot injury?
1.4 What is the probability that a randomly chosen injured player is a striker?

Problem 2

2.1 What proportion of the gunny bags have a breaking strength less than 3.17 kg per sq cm?
2.2 What proportion of the gunny bags have a breaking strength at least 3.6 kg per sq cm.?
2.3 What proportion of the gunny bags have a 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.?

Problem 3

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?"

- State the null and alternate hypotheses - Conduct the hypothesis test and compute the p-value - Write down conclusions from the test results Note: Consider the level of significance as 5%."

3.2 Is the mean hardness of the polished and unpolished stones the same?
- State the null and alternate hypotheses. - Conduct the hypothesis test. - Write down conclusions from the test results. Note: Consider the level of significance as 5%.

Problem 4

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

"- State the null and alternate hypotheses - Check the assumptions of the hypothesis test. - Conduct the hypothesis test and compute the p-value - Write down conclusions from the test results - In case the implant hardness differs, identify for which pairs it differs Note: 1. Both types of alloys cannot be considered together. You must conduct the analysis separately for the two types of alloys. 2. Even if the assumptions of the test fail, kindly proceed with the test."

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

"- State the null and alternate hypotheses - Check the assumptions of the hypothesis test. - Conduct the hypothesis test and compute the p-value - Write down conclusions from the test results - In case the implant hardness differs, identify for which pairs it differs Note: 1. Both types of alloys cannot be considered together. You must conduct the analysis separately for the two types of alloys. 2. Even if the assumptions of the test fail, kindly proceed with the test."

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

"- Create Interaction Plot - Inferences from the plot Note: Both types of alloys cannot be considered together. You must conduct the analysis separately for the two types of alloys."

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

"- State the null and alternate hypotheses - Check the assumptions of the hypothesis test. - Conduct the hypothesis test and compute the p-value - Write down conclusions from the test results - Identify which dentists and methods combinations are different, and which interaction levels are different. Note: 1. Both types of alloys cannot be considered together. You must conduct the analysis separately for the two types of alloys. 2. Even if the assumptions of the test fail, kindly proceed with the test."

Problem 1

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

Finding the probability that a randomly chosen player would suffer an injury:

Probability = 145/235

*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?

Finding the probability that a player is a forward:

Probability = 123/235

* 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 a striker position and has a foot injury?

Finding the probability that a randomly chosen player plays a striker position and has a foot injury

Probability = 45/235

* The probability that a randomly chosen player plays a striker position and has a foot injury is 19.15%

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

Finding the probability that a randomly chosen injured player is a striker

Probability = 45/145

* The probability that a randomly chosen injured player is a striker is 31.03%

Problem 2

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

Finding what proportion of the gunny bags have a breaking strength of less than 3.17 kg per sq cm using norm.cdf function:

* 11.12% is the probability that the time of effect of a dose is less than 3.17

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

Finding what probability of the gunny bags has a breaking strength of at least 3.6kg per sq cm using 1-norm.cdf function:

* 82.47% is the probabilities of the gunny bags have a breaking strength of at least 3.6 kg per sq cm.

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

Finding what proportion of the gunny bags have a breaking strength between 5 and 5.5 kg per sq cm using norm.cdf :

* 13.06% is the probability of the gunny bags have a 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.?

Finding the proportion of the gunny bags has a breaking strength not between 3 and 7.5 kg per sq cm using norm.cdf:

* 13.9% is the probabilities of the gunny bags have a breaking strength not between 3 and 7.5kg per sq cm

Problem 3

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?

Step 1: Define null and alternative hypotheses

Let's frame the null and alternative hypothesis based on the above claim can be formulated as:

H0: The mean hardness of unpolished stones is greater than 150.

Ha: The mean hardness of treated and polished is less than 150.

Let μ be the mean hardness of printing stone

Mathematically, the above formulated hypotheses can be written as:

$H0 : \mu>=150$

$Ha : \mu<150$

CHECKING ASSUMPTIONS OF HYPOTHESIS TEST

- The data is approximately normally distributed.
- Observations are independent.
- The sample size is more than 30.
- Unknown population standard deviation.

Step 2: Select Appropriate test

This is a case of a one-tailed test for the significance of a single mean. As the population standard deviation is unknown, a one-sample t-test will be appropriate to test the formulated hypotheses.

In a one sample t-test, we compare a sample mean with a hypothesized population mean to check if the sample mean differs significantly from the population mean.

We are provided that the sample is randomly selected and we assume that it has come from a normally distributed population.

Step 3: Decide the significance level

Let's set the level of significance (alpha) as 0.05

Step 4: Calculate the p-value

We will use the ttest_1samp() function from the scipy.stats library to perform a one-sample t-test.

The ttest_1samp() function takes the sample observations, hypothesized population mean and the direction of the alternative hypothesis as input and returns the test statistic and the p-value for the test.

The sample observations are the values of hardness of the printing stone in the dataset.

The hypothesized population mean, as provided in the problem statement and defined in the formulated hypotheses is 150.

As it is a one-tailed test towards the right, we will set the argument alternative = 'lesser'

Finding whether Zingaro's belief that the unpolished stones may not be suitable for printing is correct or not.

The type of test (one-tailed) has an effect on the p-value. The alternative argument is set to 'less' as our alternative hypothesis states that mean hardness of zingaro printing stone is less than 150.

The p_value is 0.0195(1.95)

Step 5: Compare the p-value with alpha

As the p-value 0.0195(1.95) is less than the level of significance 0.05 (5), we reject the null hypothesis.

Step 6: Draw Inference

Since the p-value is less than the 5% significance level, we reject the null hypothesis. Hence, we have enough statistical evidence to say that the mean hardness of the printing stone is lesser than 150 which means that the unpolished stones are not suitable for printing.

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

Step 1: Define null and alternative hypotheses

Let's formulate the null and alternative hypothesis for the above problem

H_0: The mean hardness of the polished and unpolished stones are the same.

H_a: The mean hardness of the polished and unpolished stones are not same.

Mathematically, the above formulated hypotheses can be written as:

H0:μ=150

Ha:μ !=150

CHECKING ASSUMPTIONS OF HYPOTHESIS TEST:

- The data is approximately normally distributed.
- Observations are independent.
- The sample size is more than 30.
- Unknown population standard deviation.

Step 2: Select Appropriate test

This is an independent two sample t-test .The null hypothesis Ho is that the mean hardness of polished stones equals the mean hardness of unpolished stones, and the alternative hypothesis * Ha is that they are not equal (two-tailed test)

Step 3: Decide the significance level

Let's set the level of significance (alpha) as 0.05

Step 4: Calculate the p-value

Two-sample t-test

Testing whether the mean hardness of the polished and unpolished stones the same .

The p_value is 0.0014655150194628353(0.0146)

The type of test (two-tailed) has an effect on the p_value. The alternative argument is set to 'two-equal' as our alternative hypothesis states that the hardness of polished stones are not equal to the unpolished stones.

Step 5: Compare the p-value with alpha

As the p-value 5.600076564450542e-16 is less than the level of significance, we reject the null hypothesis.

Step 6: Draw inference

Since the p-value is less than the 5% significance level, we reject the null hypothesis. Hence, we have enough statistical evidence to say that the mean hardness of polished and unpolished stones are not same.

Actionable Insights and Conclusions

- From the sample data, we observed that

- We have enough statistical evidence to say that the unpolished stones are not suitable for printing.

- The mean hardness of polished and unpolished stones are not same.

- There is enough statistical evidence (p-value of 4.171286997419652e-05 at 5% level of significance) to conclude that the mean hardness of the printing stone is lesser than 150 which shows that the unpolished stones are unsuitable for printing.

- There is a difference in the mean hardness between polished and unpolished stones (p-value of 0.00146 at 5% level of significance) to conclude that the mean hardness of the polished and unpolished stones are not same.

greatlearning
Learning for Life

Problem 4

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

Step 1: Define null and alternative hypotheses

Let's formulate the null and alternative hypothesis for the above problem

H_0: There is no significant difference in the mean hardness across dentists.

H_a : There is a significant difference in the mean hardness across dentists.

CHECKING ASSUMPTIONS OF HYPOTHESIS TESTS

- The populations are normally distributed
- Samples are independent simple random samples
- Population variances are equal

Step 2: Select Appropriate test

To examine the variability in hardness based on dentists:

- perform a One way ANOVA, where the factor is the dentist (categorical variable)
- Investigating the means for each dentist and calculate the F-statistic.

Step 3: Decide the significance level

Let's set the level of significance (alpha) as 0.05

Step 4: Calculating the p_value

```
ANOVA Table for Dentist(Alloy 1):
              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

ANOVA Table for Dentist(Alloy 2):
               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 ALLOY 1:

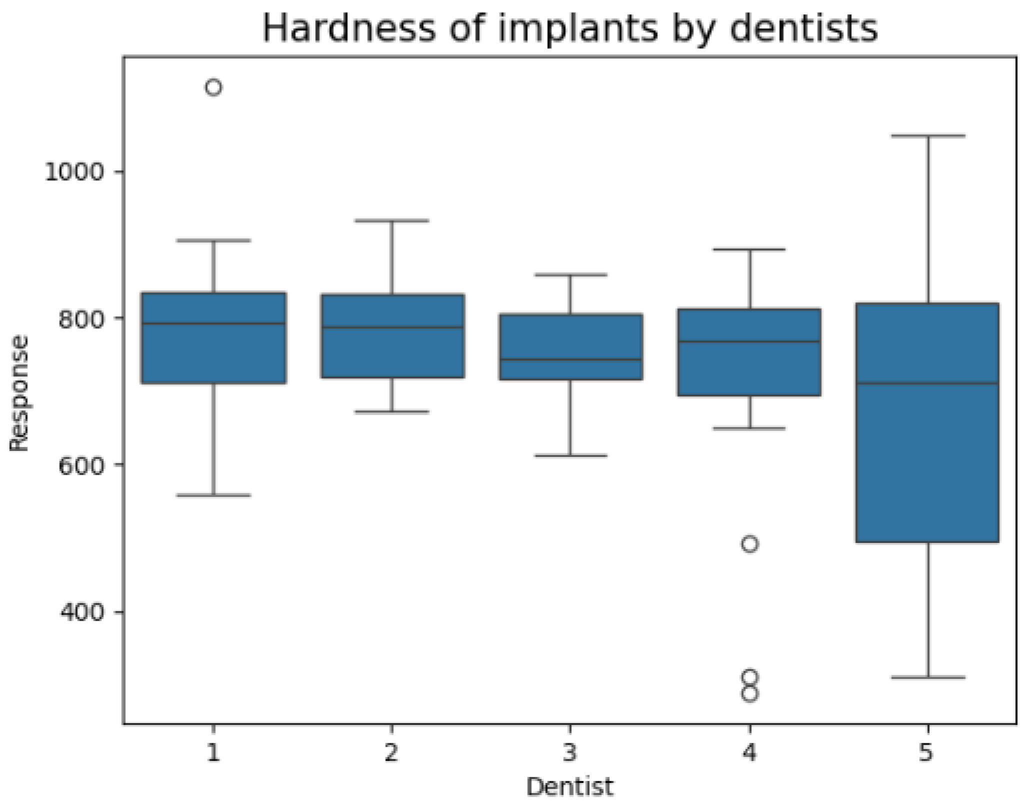
```
Multiple Comparison of Means - Tukey HSD, FWER=0.05
========================================================
group1 group2  meandiff p-adj    lower    upper   reject
--------------------------------------------------------
     1      2   11.3333 0.9996 -145.0423  167.709  False
     1      3  -32.3333 0.9757  -188.709 124.0423  False
     1      4  -68.7778 0.7189 -225.1535  87.5979  False
     1      5 -122.2222 0.1889 -278.5979  34.1535  False
     2      3  -43.6667 0.9298 -200.0423  112.709  False
     2      4  -80.1111 0.5916 -236.4868  76.2646  False
     2      5 -133.5556 0.1258 -289.9312  22.8201  False
     3      4  -36.4444 0.9626 -192.8201 119.9312  False
     3      5  -89.8889 0.4805 -246.2646  66.4868  False
     4      5  -53.4444 0.8643 -209.8201 102.9312  False
```

FOR ALLOY 2:

```
Multiple Comparison of Means - Tukey HSD, FWER=0.05
=======================================================
group1 group2 meandiff p-adj    lower    upper   reject
-------------------------------------------------------
     1      2  -4.1111    1.0 -225.5687 217.3465  False
     1      3 -36.5556 0.9895 -258.0131  184.902  False
     1      4    -70.0 0.8941 -291.4576 151.4576  False
     1      5 -90.1111 0.7724 -311.5687 131.3465  False
     2      3 -32.4444 0.9933  -253.902 189.0131  False
     2      4 -65.8889 0.9132 -287.3465 155.5687  False
     2      5    -86.0 0.8008 -307.4576 135.4576  False
     3      4 -33.4444 0.9925  -254.902 188.0131  False
     3      5 -53.5556 0.9574 -275.0131  167.902  False
     4      5 -20.1111  0.999 -241.5687 201.3465  False
```

Step 5: Draw inference

Since the p-value is less than the 5% significance level, we do not reject the null hypothesis. Hence, we do not have enough statistical evidence to say that the mean hardness of across dentist are not same.



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

Step 1: Define null and alternative hypotheses

Let's formulate the null and alternative hypothesis for the above problem

H_0 :There is no significant difference in the mean hardness across methods.

H_a: There is a significant difference in the mean hardness across methods.

CHECKING ASSUMPTIONS OF HYPOTHESIS TESTS

- The populations are normally distributed
- Samples are independent simple random samples
- Population variances are equal

Step 2: Select Appropriate test

To examine the variability in hardness based on methods:

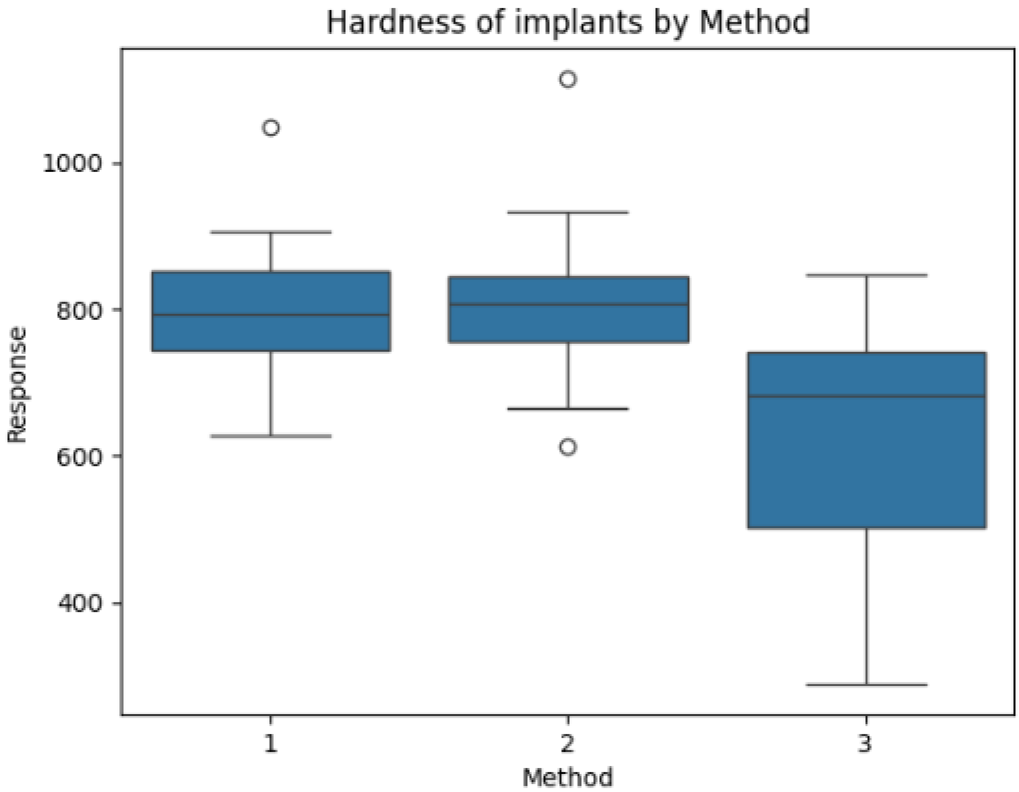
- Perform a One-way ANOVA , where the factor is the method of implant.

- Analyze the practical implications of method effectiveness.

Step 3: Decide the significance level

Let's set the level of significance (alpha) as 0.05

Step 4 : Calculating p_value




```
ANOVA Table for Method(Alloy 1):
             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

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
```

```
ANOVA Table for Method(Alloy 2):
             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


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
```

ANOVA for Method effect:

	df	sum_sq	mean_sq	F	PR(>F)
C(Method)	2.0	5.934275e+05	2.967137e+05	3.428508e+28	0.0
C(Response * Method)	60.0	1.297668e+06	2.162780e+04	2.499079e+27	0.0
Residual	29.0	2.509750e-22	8.654310e-24	NaN	NaN

Step 5: Draw inference

Since the p-value is less than the 5% significance level, we reject the null hypothesis. Hence, we have enough statistical evidence to say that the mean hardness of across methods are not same.

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

Step 1: Define null and alternative hypotheses

Let's formulate the null and alternative hypothesis for the above problem

H_0 :There is no significant interaction effect between dentist and method for each alloy.

H_a :There is a significant interaction effect between dentist and method for each alloy.

CHECKING ASSUMPTIONS OF HYPOTHESIS TESTS

- The populations are normally distributed

- Samples are independent simple random samples

- Population variances are equal

Step 2: Select Appropriate test

The analyze the interaction effect:

- Perform a Two-way ANOVA(factors:dentist and method) with an interaction term:

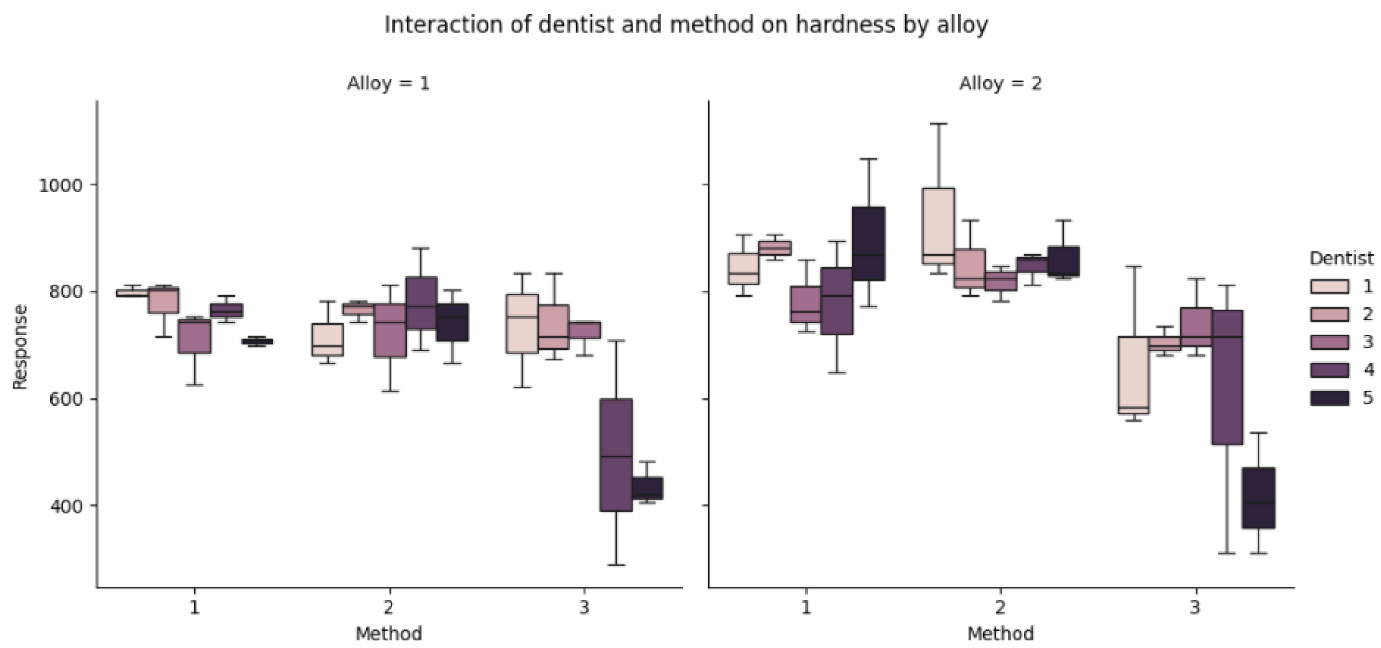
    - model : hardness=dentist+method+(dentist*method)

Step 3: Decide the significance level

Let's set the level of significane (alpha) as 0.05

Step 4: Calculate the p-value

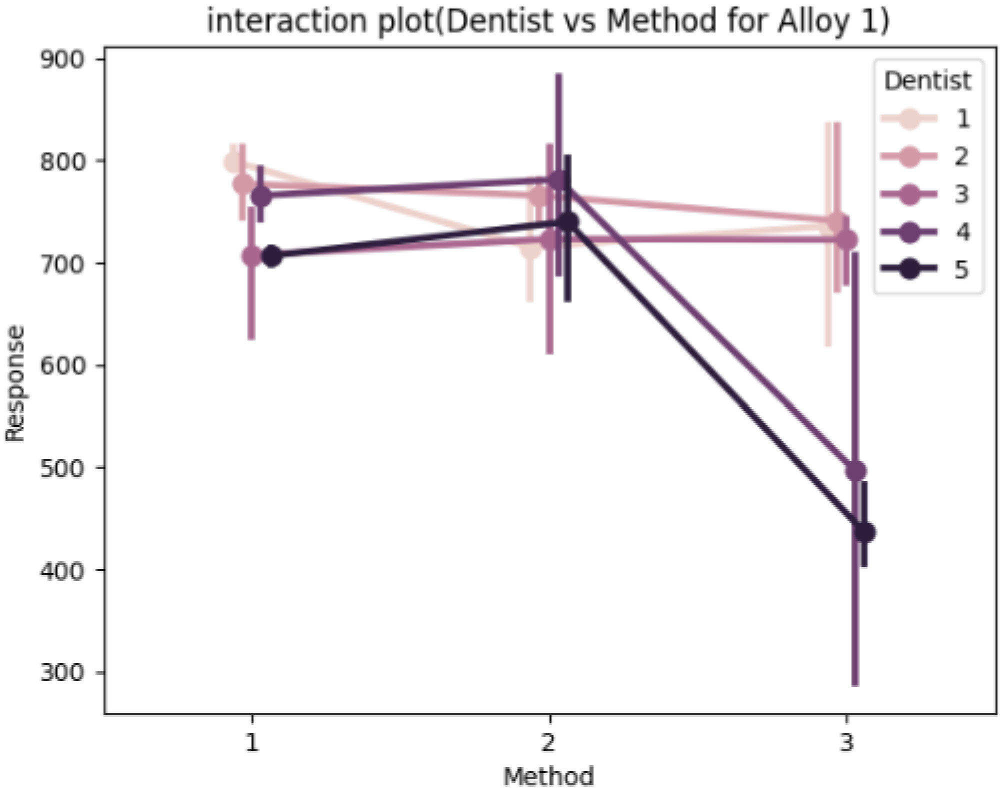
Finding the interaction effect between dentist and method for each alloy



```
anova for interaction effect                                           df         sum_sq
C(Dentist)                        4.0  157794.555556   39448.638889   4.009791
C(Method)                         2.0  593427.488889  296713.744444  30.159727
C(Alloy)                          1.0  105815.511111  105815.511111  10.755710
C(Dentist):C(Method)              8.0  306471.844444   38308.980556   3.893950
C(Dentist):C(Alloy)               4.0    5687.044444    1421.761111   0.144516
C(Method):C(Alloy)                2.0   54685.088889   27342.544444   2.779257
C(Dentist):C(Method):C(Alloy)     8.0   76929.355556    9616.169444   0.977444
Residual                         60.0  590284.666667    9838.077778        NaN

                                     PR(>F)
C(Dentist)                       6.003466e-03
C(Method)                        8.599275e-10
C(Alloy)                         1.733434e-03
C(Dentist):C(Method)             9.385714e-04
C(Dentist):C(Alloy)              9.647377e-01
C(Method):C(Alloy)               7.009137e-02
C(Dentist):C(Method):C(Alloy)    4.623734e-01

Residual                                  NaN

Two-way ANOVA (Dentist x Method)for Alloy 1 :
                        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


Two-way ANOVA (Dentist x Method)for Alloy 2 :
                        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
```

Note : The p-value in the two of the treatments is greater than α(0.05) in alloy2.

Let us check whether there is any interaction effect between the treatments.



Step 6: Drawing conclusion

Since the p-value is less than the 5% significance level, we reject the null hypothesis. Hence, we have enough statistical evidence to say that there is a significant interaction effect between dentist ,response and method for the first alloy and for the second alloy p_value is more than significance level for dentist this means there is an interaction effect between method and response.

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

Step 1: Define null and alternative hypotheses

Let's formulate the null and alternative hypothesis for the above problem
H_0: There is no significant variation of the hardness of implants vary depending on dentists and methods together.

H_a: There is a significant variation of the hardness of implants vary depending on dentists and methods together.

CHECKING ASSUMPTIONS OF HYPOTHESIS TESTS

- The populations are normally distributed
- Samples are independent simple random samples
- Population variances are equal

Step 2: Select Appropriate test:

To assess the combined effect of dentists and methods:

- Use the Two-way ANOVA results , focusing on the main effects of dentists and method (ignoring interaction initially).
- Assess whether the main effects are significant:
  - If both main effects are significant: Both factors independently influence hardness.
  - If interaction is significant: The combined effect of dentists and methods must be interpreted carefully.

Step 3: Decide the significance level

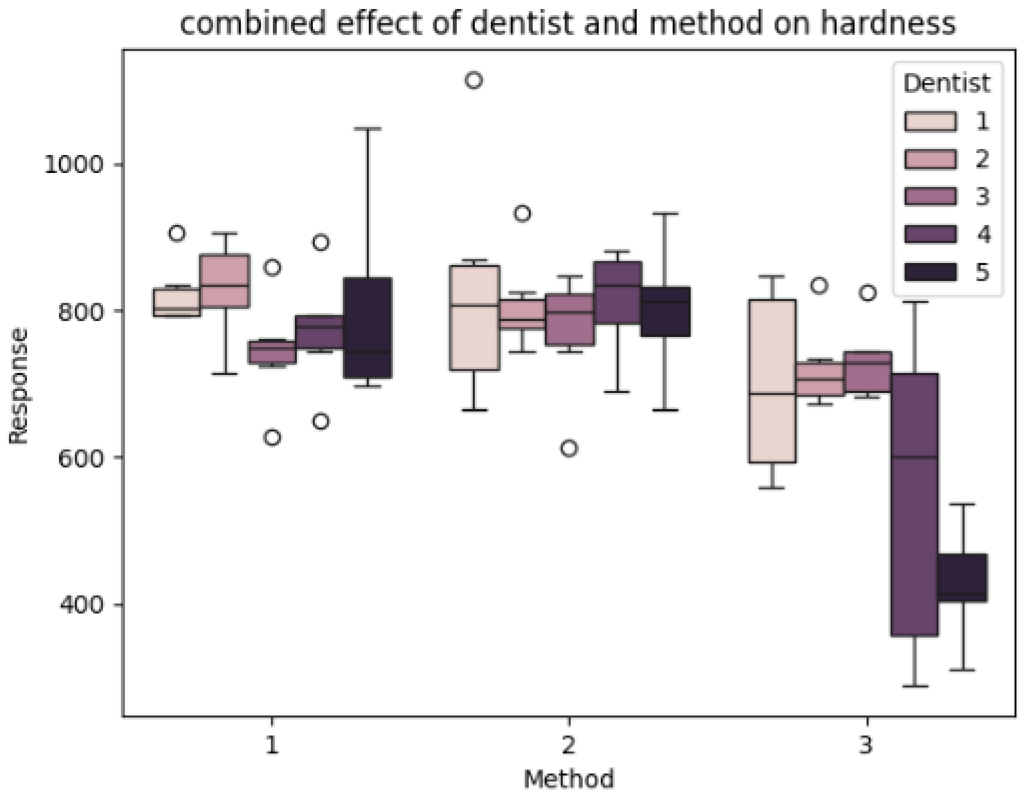
Let's set the level of significance (alpha) as 0.05

Step 4: Calculate the p-value

```
ANOVA for dentists and methods together
                                                                df        sum_sq
C(Dentist)             4.0  157794.555556   39448.638889   3.550086
C(Method)              2.0  593427.488889  296713.744444  26.702047
C(Dentist):C(Method)   8.0  306471.844444   38308.980556   3.447526
Residual              75.0  833401.666667   11112.022222        NaN

                           PR(>F)
C(Dentist)            1.045384e-02
C(Method)             1.750208e-09
C(Dentist):C(Method)  1.969515e-03
Residual                       NaN

combined ANOVA (Dentist + Method + Interaction)for Alloy 1:
                        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
```

combined ANOVA (Dentist + Method + Interaction)for Alloy 2:

	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



Step 5: Drawing inferences

Since the p-value is less than the 5% significance level, we reject the null hypothesis for first alloy and for second there is not enough evidence to reject the null hypothesis since denist and dentist with method has p_value more than significance level. Hence, we have enough statistical evidence to say that there is a significant variation of the hardness of implants vary depending on dentists and methods together in alloy 1 and for alloy 2 there is no significance variation of the hardness of implants vary depending on dentist and dentist with method.