ADVANCED STATISTICS

BUSINESS REPORT

THANUSRI

05/11/2023

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**

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

Sol:

The probability that a randomly chosen player would suffer an injury is 0.617

Number of player who would suffer an injury = 145

Total number of players = 235

The probability that a randomly chosen player would suffer an injury = 145 / 235 = 0.6170

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

Sol:

The probability that a player is a forward or a winger is 0.5234

Number of player who are a forward or a winger = 94 + 29 = 123

Total number of players = 235

The probability that a randomly chosen player is a forward or a winger = 123 / 235 = 0.5234

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

Sol:

The probability that a randomly chosen player plays in a striker position is 45 and has a foot injury =45/77, i.e., 0.58%

Number of player who plays in a striker position and has a foot injury = 45

Total number of players = 77

The probability that a randomly chosen player plays in a striker position and has a foot injury is 45/77=0.58

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

Sol:

The probability that a randomly chosen injured player is a striker is 0.3103

Number of injured player who are striker = 45

Number of injured players = 145

The probability that a randomly chosen injured player is a striker = 45 / 145 = 0.3103

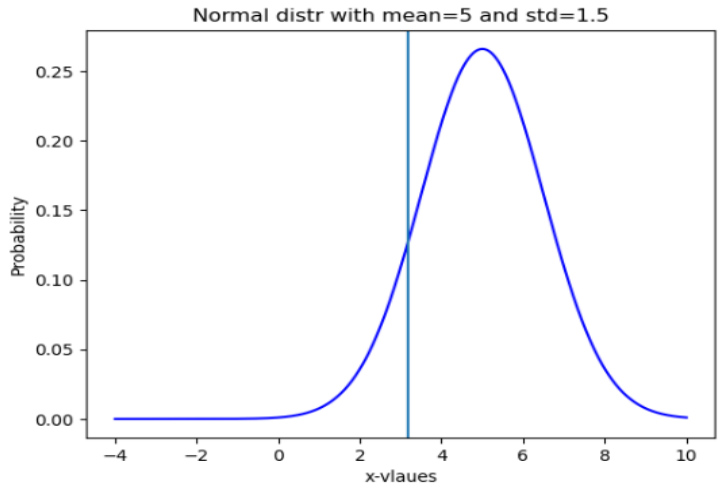
Problem 2

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 of less than 3.17 kg per sq cm?

Sol:

M (Mean) = 5
σ (Standard Deviation) = 1.5
X (Gunny Bag Strength) = 3.17
Z Value (Z) = (X -μ)/ σ = -1.22
CDF Value = 0.1112
Z = (3.17-5)/1.5 = -1.22P(z < -1.22) = 0.1112

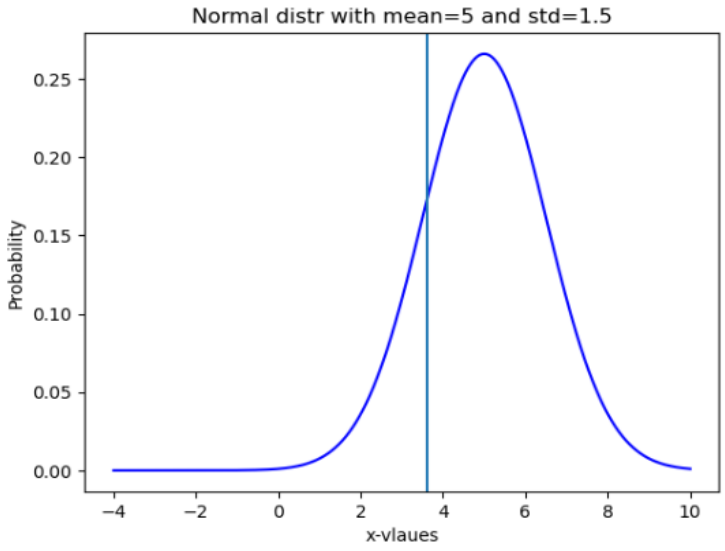


It can be interpreted as 11.1% gunny bags have a breaking strength less than 3.17 kg per sq. cm.

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

Sol:

Z = (3.6 – 5)/1.5 = -0.9333
P(Z > -0.9333) = 0.8247

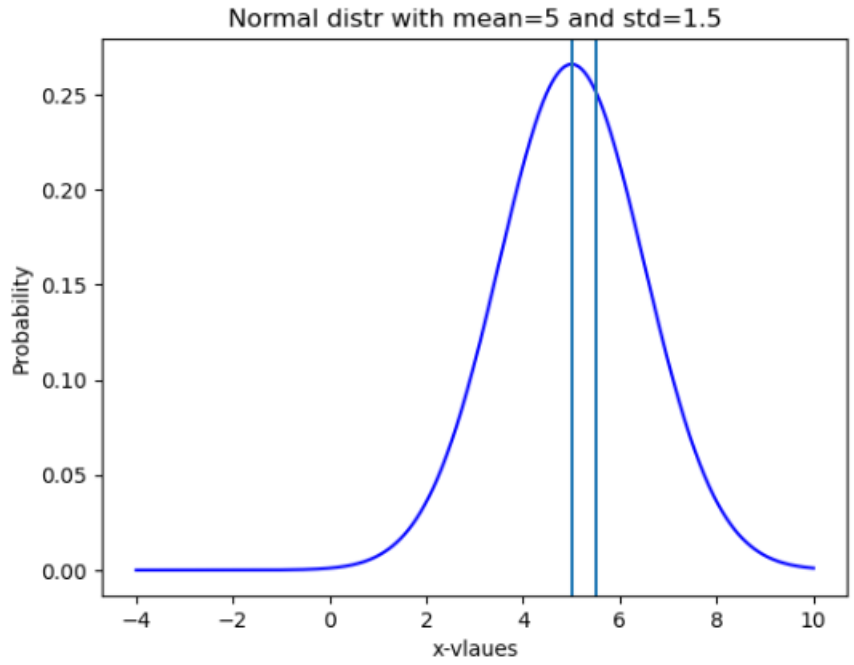


Thus we conclude that the 82.46% of gunny bags have the breaking strength at least 3.6 kg/sq.cm

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

Sol:

Z = (5.5 –5)/1.5 = 0.3333
Z = (5 – 5)/1.5 = 0
P(0 < z < 0.3333) = P(Z < 0.3333) – P(Z < 0) = 0.1306



Thus we conclude that 13.06% of gunny bags have breaking strength between 5 and5.5 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.?

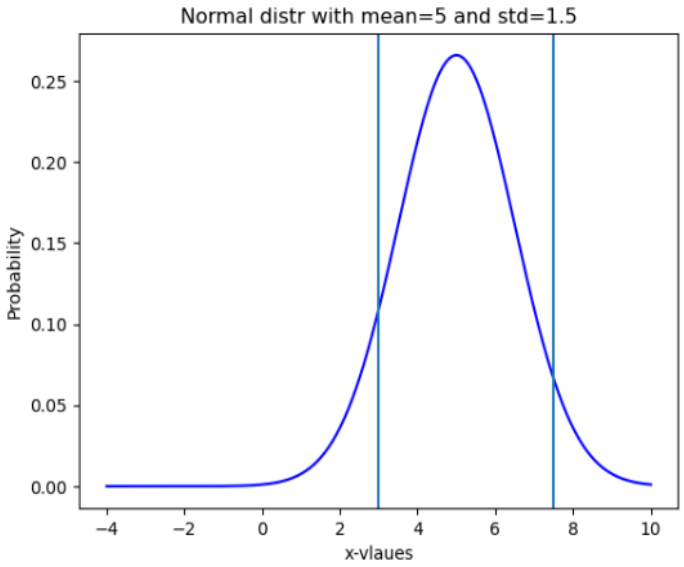
Sol:

Z3 = (3-5)/1.5 = -1.3333

Z4 = (7.5-5)/1.5 =1.6666

P(Z4)- P(Z3)= 0.86099

1-0.86099=0.13901



Thus we conclude that the proportion of gunny bags having strength not between 3

and 7.5 per sq cm is 13.9%

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

hypotheses are :
Null hypothesis : Unpolished stone is suitable for printing which means Brinell's hardness index is atleast or equal to 150
Alternative hypothesis : Unpolished stone is not suitable for printing which means Brinell's hardness index is less than 150.

•H0 :μ=150
•HA :μ< 150,

Where μ be the population mean of unpolished stones
Assuming Level of significance α = 0.05
Sample Size = 75
Degrees of Freedom =75-1=74,
since the sample size for both samples are the same .

Approach :
Here, independent one sample one-tailed t test (> or < sign is there)is to be conducted.
so hypothesis for unpolished stones is being done here.So ,Test statistic

p-value is 4.171286997419652e-05

Decision rule :
The p-value approach states that reject null hypothesis if p-value < significance level . In this case, the p-value which is not higher than the level of significance (0.05). So, the null hypothesis is rejected.

Conclusion :
Since, the null hypothesis is rejected. At 5% significance level, there is enough evidence to make conclusion that the unpolished stones are not suitable for printing and his claim is justified.

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

Sol:

Hypotheses are :
Null hypothesis: Mean hardness of Polished and Unpolished stones are the same.
Alternative hypothesis: Mean hardness of Polished and Unpolished stones are not same.

•H0 :μ Unpolished =μ Treated and Polished
•HA:μ Unpolished ≠ μ Treated and Polished

Where μ Unpolished be the population mean of unpolished stones and μ Treated and Polished be the population mean of treated and polishedstones.

Approach:
The not equal to sign in the alternative hypothesis indicate that the test is two-tailed and here we will use two independent samples t test .So ,
Test statistic

p-value is 0.00146551

Decision rule :
The p-value approach states that reject null hypothesis if p-value < significance level . In this case, the p-value is 0.002 which is not higher than the significance level (0.05). Hence, the null hypothesis is rejected.

Conclusion:
Since, the null hypothesis is rejected . At 5% significance level, there is insufficient evidence to conclude that the mean hardness of the polished and unpolished stones are the same .

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.

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

Hypothesis Test for Shapiro Wilk Test for Alloy 1

- Null Hypothesis (for Alloy 1): The population from which the sample is drawn follows a normal distribution.
- Alternative Hypothesis (for Alloy 1): The population from which the sample is drawn does not follow a normal distribution.

```
1.0: ShapiroResult(statistic=0.9113541841506958, pvalue=0.3254688084125519)
2.0: ShapiroResult(statistic=0.9642462134361267, pvalue=0.8415456414222717)
3.0: ShapiroResult(statistic=0.8721169233322144, pvalue=0.12953516840934753)
4.0: ShapiroResult(statistic=0.8368974328041077, pvalue=0.05333680287003517)
5.0: ShapiroResult(statistic=0.8534296751022339, pvalue=0.08127813786268234)
```

From the Wilk test: The Test Statistic value is 0.83 and P-value is 0.05

Conclusion: We can conclude that P-value is 0.053 (<<0.05), so we have enough evidence to Reject Null Hypothesisand consider that the sample drawn for Alloy 1 does not follow a normal distribution

Hypothesis Test for Shapiro Wilk Test for Alloy 2

- Null Hypothesis (for Alloy 2): The population from which the sample is drawn follows a normal distribution.
- Alternative Hypothesis (for Alloy 2): The population from which the sample is drawn does not follow a normal distribution.

```
1: ShapiroResult(statistic=0.9039731621742249, pvalue=0.27593979239463806)
2: ShapiroResult(statistic=0.9392004013061523, pvalue=0.5735077857971191)
3: ShapiroResult(statistic=0.9340971112251282, pvalue=0.5213080644607544)
4: ShapiroResult(statistic=0.7613219022750854, pvalue=0.007332688197493553)
5: ShapiroResult(statistic=0.9131584167480469, pvalue=0.33861100673675537)
```

From the Wilk test: The Test Statistic value is 0.76 and P-value is 0.007

Conclusion: • We can conclude that the P-value is 0.007 (<<0.05). So we have enough evidence to Reject Null Hypothesis and consider that the sample drawn for Alloy 2 does not follow a normal distribution

Define the null and alternative hypothesis

Hypothesis 1
Ho: The mean hardness of dental implant is same for all 5 dentists provided the alloy 1 is used.
Ha: for at least 1 dentist the mean Hardness of Dental implant is different when using Alloy1.

	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

P-value is 0.11 which is greater than alpha i.e., 0.05. Hence, we Fail to reject null hypothesis and consider there is no difference in means among the dentists in terms of implant hardness for Alloy 1.

Hypothesis 2

Ho: The mean Hardness of Dental implant is same for all 5 dentists provided the Alloy 2 is used.
Ha: for at least 1 dentist the mean Hardness of Dental implant is different when using 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

P-value is 0.71 which is greater than alpha i.e., 0.05. Hence, we Fail to reject null hypothesis and consider there is no difference in means among the dentists in terms of implant hardness for Alloy 2

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
1.0	2.0	3.6111	nan	nan	nan	False
1.0	3.0	-34.4444	nan	nan	nan	False
1.0	4.0	-69.3889	nan	nan	nan	False
1.0	5.0	-106.1667	nan	nan	nan	False
1.0	nan	nan	nan	nan	nan	False
2.0	3.0	-38.0556	nan	nan	nan	False
2.0	4.0	-73.0	nan	nan	nan	False
2.0	5.0	-109.7778	nan	nan	nan	False
2.0	nan	nan	nan	nan	nan	False
3.0	4.0	-34.9444	nan	nan	nan	False
3.0	5.0	-71.7222	nan	nan	nan	False
3.0	nan	nan	nan	nan	nan	False
4.0	5.0	-36.7778	nan	nan	nan	False
4.0	nan	nan	nan	nan	nan	False
5.0	nan	nan	nan	nan	nan	False

based on above Tukey HSD test at a FWER of 0.05 there are no statistically significant differences in means in between any of the groups we've compared in above analysis

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

Hypothesis Test for Shapiro Wilk Test for Alloy 1
- Null Hypothesis (for Alloy 1): The population from which the sample is drawn follows a normal distribution.
- Alternative Hypothesis (for Alloy 1): The population from which the sample is drawn does not follow a normal distribution.

```
1.0: ShapiroResult(statistic=0.9183822870254517, pvalue=0.18198540806770325)
2.0: ShapiroResult(statistic=0.9732585549354553, pvalue=0.9030335545539856)
3.0: ShapiroResult(statistic=0.9114548563957214, pvalue=0.14254699647426605)
```

From the Wilk test: The Test Statistic value is 0.97 and P-value is 0.90

Hypothesis Test for Shapiro Wilk Test for Alloy 2
- Null Hypothesis (for Alloy 2): The population from which the sample is drawn follows a normal distribution.
- Alternative Hypothesis (for Alloy 2): The population from which the sample is drawn does not follow a normal distribution.

```
1.0: ShapiroResult(statistic=0.963810384273529, pvalue=0.7582374811172485)
2.0: ShapiroResult(statistic=0.755793035030365, pvalue=0.001051110913977027)
3.0: ShapiroResult(statistic=0.9021322131156921, pvalue=0.1025901660323143)
```

From the Wilk test: The Test Statistic value is 0.75 and P-value is 0.001

Hypothesis Test for Alloy 1 data
- Null hypothesis (for Alloy 1): There is no difference in means among the Methods in terms of implant hardness for Alloy1.
- 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)
C(Method)	2.0	148472.177778	74236.088889	6.263327	0.004163
Residual	42.0	497805.066667	11852.501587	NaN	NaN

P-value is 0.004 which is lesser than alpha i.e., 0.05. Hence, we have enough evidence to reject null hypothesis and consider there is a difference in means among the Methods in terms of implant hardness for Alloy 1.

Hypothesis Test for Alloy 2 data
- Null hypothesis (for Alloy 2): There is no difference in means among the Methods in terms of implant hardness for Alloy2.

• Alternative hypothesis (for Alloy2): There is a difference in means among the Methods in terms of implant hardness for Alloy2.

	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

P-value is 0.000005 which is lesser than alpha i.e., 0.05. Hence, we have enough evidence to reject null hypothesis and consider there is a difference in means among the Methods in terms of implant hardness for Alloy 2.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
1.0	2.0	10.4333	nan	nan	nan	False
1.0	3.0	-166.8	nan	nan	nan	False
1.0	nan	nan	nan	nan	nan	False
2.0	3.0	-177.2333	nan	nan	nan	False
2.0	nan	nan	nan	nan	nan	False
3.0	nan	nan	nan	nan	nan	False

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

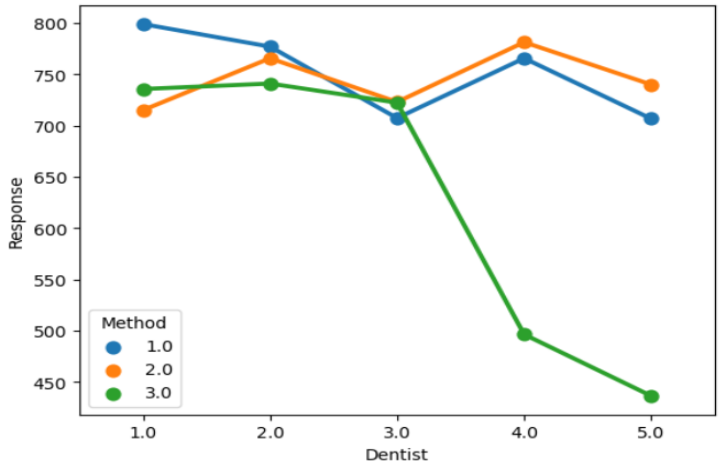
Define null and alternative hypothesis

• Null hypothesis: There is no difference among the Interaction effect between Dentist and Method levels in terms of implant hardness

• Alternative hypothesis: There is a difference among the Interaction effect between Dentist and Method levels in terms of implant hardness

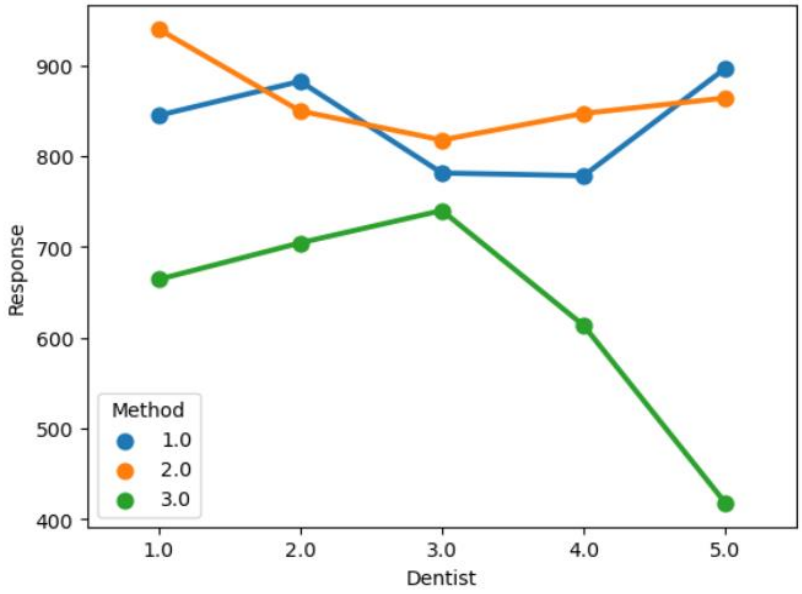
	df	sum_sq	mean_sq	F	PR(>F)
C(Dentist):C(Method)	14.0	441097.244444	31506.946032	4.606728	0.000221
Residual	30.0	205180.000000	6839.333333	NaN	NaN

P-value is 0.000221 which is lesser than alpha i.e., 0.05. Hence, we have enough evidence to 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.



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

P-value is 0.000482 which is lesser than alpha i.e., 0.05. Hence, we have enough evidence to 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.



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

Hypothesis Test alloy1

- Null hypothesis: There is no difference among the factors Dentist and Method levels in terms of implant hardness.
- Alternative hypothesis: There is a difference among the factors Dentist and Method levels in terms of implant hardness

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

P value for Method is 0.0022 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

Hypothesis Test for Alloy 2

• 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)
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

P value for Method is 0.000008 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 2

Calculate two-way anova for interaction between dentist and 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

P value for Dentist, Method, and Interaction variable (Dentist: Method) are all less than 0.05. Hence, we can conclude that to reject null hypothesis and consider there is at least one pair of variables means are different for Alloy 1

Calculate two-way anova for interaction between dentist and 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