AS PROJECT REPORT

CONTENTS

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?

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?

1.5 What is the probability that a randomly chosen injured player is either a forward or an attacking midfielder?

Problem 2

An independent research organization is trying to estimate the probability that an accident at a nuclear power plant will result in radiation leakage. The types of accidents possible at the plant are, fire hazards, mechanical failure, or human error. The research organization also knows that two or more types of accidents cannot occur simultaneously.

According to the studies carried out by the organization, the probability of a radiation leak in case of a fire is 20%, the probability of a radiation leak in case of a mechanical 50%, and the probability of a radiation leak in case of a human error is 10%. The studies also showed the following;

- The probability of a radiation leak occurring simultaneously with a fire is 0.1%.

- The probability of a radiation leak occurring simultaneously with a mechanical failure is 0.15%.
- The probability of a radiation leak occurring simultaneously with a human error is 0.12%.

On the basis of the information available, answer the questions below:

2.1 What are the probabilities of a fire, a mechanical failure, and a human error respectively?

2.2 What is the probability of a radiation leak?

2.3 Suppose there has been a radiation leak in the reactor for which the definite cause is not known. What is the probability that it has been caused by:

- A Fire.
- A Mechanical Failure.
- A Human Error. (Page 9)

Problem 3:

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)**

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

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

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

3.4 What proportion of the gunny bags have a breaking strength NOT between 3 and 7.5 kg per sq cm.? (Page 13)

Problem 4:

Grades of the final examination in a training course are found to be normally distributed, with a mean of 77 and a standard deviation of 8.5. Based on the given information answer the questions below.

4.1 What is the probability that a randomly chosen student gets a grade below 85 on this exam?

4.2 What is the probability that a randomly selected student scores between 65 and 87?

4.3 What should be the passing cut-off so that 75% of the students clear the exam?

(Page 18)

Problem 5:

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);

5.1 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. Do you think Zingaro is justified in thinking so?

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

(Page 19)

Problem 6:

Aquarius health club, one of the largest and most popular cross-fit 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 more than 5 push-ups, as compared to when he/she enrolled in the program. Using the sample data provided can you conclude whether the program is successful? (Consider the level of Significance as 5%)

Note that this is a problem of the paired-t-test. Since the claim is that the training will make a difference of more than 5, the null and alternative hypotheses must be formed accordingly.

(Page 21)

Problem 7:

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.

1. Test whether there is any difference among the dentists on the implant hardness. State the null and alternative hypotheses. Note that both types of alloys cannot be considered together. You must state the null and alternative hypotheses separately for the two types of alloys.?
2. Before the hypotheses may be tested, state the required assumptions. Are the assumptions fulfilled? Comment separately on both alloy types.?
3. Irrespective of your conclusion in 2, we will continue with the testing procedure. What do you conclude regarding whether implant hardness depends on dentists? Clearly state your conclusion. If the null hypothesis is rejected, is it possible to identify which pairs of dentists differ?
4. Now test whether there is any difference among the methods on the hardness of dental implant, separately for the two types of alloys. What are your conclusions? If the null hypothesis is rejected, is it possible to identify which pairs of methods differ?
5. Now test whether there is any difference among the temperature levels on the hardness of dental implant, separately for the two types of alloys. What are your conclusions? If the null hypothesis is rejected, is it possible to identify which levels of temperatures differ?
6. Consider the interaction effect of dentist and method and comment on the interaction plot, separately for the two types of alloys?
7. Now consider the effect of both factors, dentist, and method, separately on each alloy. What do you conclude? Is it possible to identify which dentists are different, which methods are different, and which interaction levels are different?

(Page 22)

SOLUTIONS

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?

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?

1.5 What is the probability that a randomly chosen injured player is either a forward or an attacking midfielder?

Answer:

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

Here we are using marginal probability to solve the problem

Total players injured = 145

Total players = 235

$$\Pr(\text{player suffering an injury}) = \text{Total players injured } / \text{Total players Probability}$$

Pr(player suffering an injury) = Total players injured /Total players Probability = 145 / 235 ≈ 0.617

Therefore, the probability that a randomly chosen player would suffer an injury is approximately 0.617 or 61.7%.

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

Here we are using marginal probability to solve the problem

Total forwards = 94

Total wingers = 29

Total players = 235

$$\Pr(\text{ forward or a winger player }) = (\text{Total forwards } + \text{ Total wingers}) \ / \text{ Total players}$$

Pr(forward or a winger player) = (Total forwards + Total wingers) / Total players Probability = (94 + 29) / 235 ≈ 0.523

Therefore, the probability that a player is a forward or a winger is approximately 0.523 or 52.3%.

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

Here we are using conditional probability to solve the problem

Strikers with foot injuries = 45

Total players = 235

$$\Pr\ (\text{striker with a foot injury player }) = (\text{Strikers with foot injuries })/\text{Total players}$$

Pr (striker with a foot injury player) = (Strikers with foot injuries)/ Total players

Probability = 45 / 235 ≈ 0.191

Therefore, the probability that a randomly chosen player plays in a striker position and has a foot injury is approximately 0.191 or 19.1%.

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

Here we are using conditional probability to solve the problem.

Strikers injured = 45

Total players injured = 145

$$\mathrm{Pr(injured\ player\ being\ a\ striker\ )} = \mathrm{Strikers\ injured\ /Total\ players\ injured\ Probability}$$

Pr(injured player being a striker)= Strikers injured / Total players injured Probability = 45 / 145 ≈ 0.310

Therefore, the probability that a randomly chosen injured player is a striker is approximately 0.310 or 31.0%.

1.5 What is the probability that a randomly chosen injured player is either a forward or an attacking midfielder?

Here we are using conditional probability to solve the problem.

Forwards injured = 56

Attacking midfielders injured = 24

Total players injured = 145

$$\begin{aligned}&\mathrm{Pr(injured\ player\ being\ a\ forward\ or\ an\ attacking\ midfielder)}\\&\qquad = \mathrm{(Forwards\ injured\ +\ Attacking\ midfielders\ injured)}\\&\qquad \mathrm{/Total\ players\ injured\ Probability}\end{aligned}$$

Pr(injured player being a forward or an attacking midfielder)= (Forwards injured + Attacking midfielders injured) / Total players injured Probability = (56 + 24) / 145 ≈ 0.662

Therefore, the probability that a randomly chosen injured player is either a forward or an attacking midfielder is approximately 0.662 or 66.2%.

Problem 2

An independent research organization is trying to estimate the probability that an accident at a nuclear power plant will result in radiation leakage. The types of accidents possible at the plant are, fire hazards, mechanical failure, or human error. The research organization also knows that two or more types of accidents cannot occur simultaneously.

According to the studies carried out by the organization, the probability of a radiation leak in case of a fire is 20%, the probability of a radiation leak in case of a mechanical 50%, and the probability of a radiation leak in case of a human error is 10%. The studies also showed the following;

- The probability of a radiation leak occurring simultaneously with a fire is 0.1%.
- The probability of a radiation leak occurring simultaneously with a mechanical failure is 0.15%.
- The probability of a radiation leak occurring simultaneously with a human error is 0.12%.

On the basis of the information available, answer the questions below:

2.1 What are the probabilities of a fire, a mechanical failure, and a human error respectively?

2.2 What is the probability of a radiation leak?

2.3 Suppose there has been a radiation leak in the reactor for which the definite cause is not known. What is the probability that it has been caused by:

- A Fire.
- A Mechanical Failure.
- A Human Error.

2.1 What are the probabilities of a fire, a mechanical failure, and a human error respectively?

Answer:

To solve this problem using Bayes' theorem, let's define the events:

F = Fire hazard

M = Mechanical failure

H = Human error

L = Radiation leak

We are given the following probabilities:

P(L|F) = 0.2 (probability of a radiation leak given a fire)

P(L|M) = 0.5 (probability of a radiation leak given a mechanical failure)

P(L|H) = 0.1 (probability of a radiation leak given a human error)

P(F∩L) = 0.001 (probability of a radiation leak occurring simultaneously with a fire)

P(M∩L) = 0.0015 (probability of a radiation leak occurring simultaneously with a mechanical failure)

P(H∩L) = 0.0012 (probability of a radiation leak occurring simultaneously with a human error).

$$P(F) = \frac{P(R \cap L)}{P\left(\frac{R}{L}\right)}$$

Substituting the given values:

P(F) = 0.001/0.2 = 0.005.

Hence the probability of fire occurring is 0.005.

$$p(M) = \frac{P(M \cap L)}{P\left(\frac{M}{L}\right)}$$

Substituting the given values:

P(M) = 0.0015/0.5 = 0.003.

Hence the probability of mechanical failure occurring is 0.003.

$$P(H) = \frac{P(H \cap L)}{P\left(\frac{H}{L}\right)}$$

Substituting the given values:

P(H) = 0.0012/0.1 = 0.012.

Hence the probability of human error occurring is 0.003.

2.2 What is the probability of a radiation leak?

Answer

P(L) = Probability of a radiation leak

To calculate the probability of a radiation leak, we need to consider all the possible scenarios:

$$\mathrm{P(L)} = \mathrm{P(L|F)} * \mathrm{P(F)} + \mathrm{P(L|M)} * \mathrm{P(M)} + \mathrm{P(L|H)} * \mathrm{P(H)}$$

Substituting the given values:

P(L) = 0.2 * P(F) + 0.5 * P(M) + 0.1 * P(H)

P(L)=0.2*0.005+0.50*0.003+0.10*0.012 = 0.0037.

Hence the probability of radiation leak is 0.0037.

2.3 Suppose there has been a radiation leak in the reactor for which the definite cause is not known. What is the probability that it has been caused by:

- A Fire.
- A Mechanical Failure.
- A Human Error.

Answer

Now, let's calculate the conditional probabilities of each cause given the radiation leak, using Bayes' theorem:

$$P(F|L) = P(L|F) * P(F) / P(L)$$

$$P(M|L) = P(L|M) * P(M) / P(L)$$

$$P(H|L) = P(L|H) * P(H) / P(L)$$

Substituting the given values and using the previously calculated value of P(L), we can compute the probabilities of the causes.

There has been a radiation leak in the reactor for which the definite cause is not known. What is the probability that it has been caused by:

- A Fire.

$$P(F|L) = P(L|F) * P(F) / P(L)$$

Substituting the given values:

0.2*0.005/0.0037 = 0.270.

Hence the probability of radiation leak happening because of fire is 0.270.

There has been a radiation leak in the reactor for which the definite cause is not known. What is the probability that it has been caused by:

- A Mechanical Failure.

$$P(M|L) = P(L|M) * P(M) / P(L)$$

Substituting the given values:

0.5*0.03/0.037 = 0.405.

Hence the probability of radiation leak happening because of mechanical failure is 0.405.

There has been a radiation leak in the reactor for which the definite cause is not known. What is the probability that it has been caused by:

- A Human Error.

$$P(H|L) = P(L|H) * P(H) / P(L)$$

Substituting the given values:

0.1*0.012/0.0037 = 0.3243.

Hence the probability of radiation leak happening because of human error is 0.3243.

Problem 3:

The breaking strength of gunny bags used for packaging cement is normally distributed with a mean of 5 kg per sq. centimetre and a standard deviation of 1.5 kg per sq. centimetre. 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)**

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

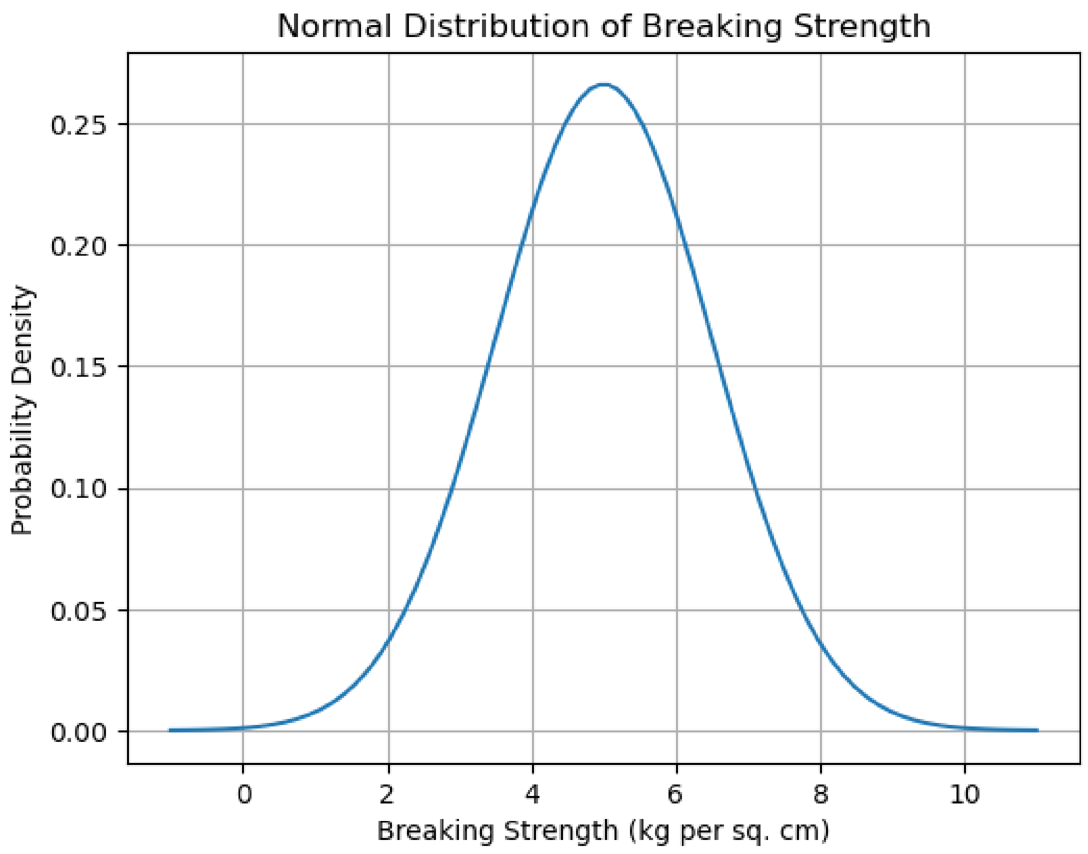
3.2 What proportion of the gunny bags have a breaking strength at least 3.6 kg per sq cm.?

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

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

Answer:

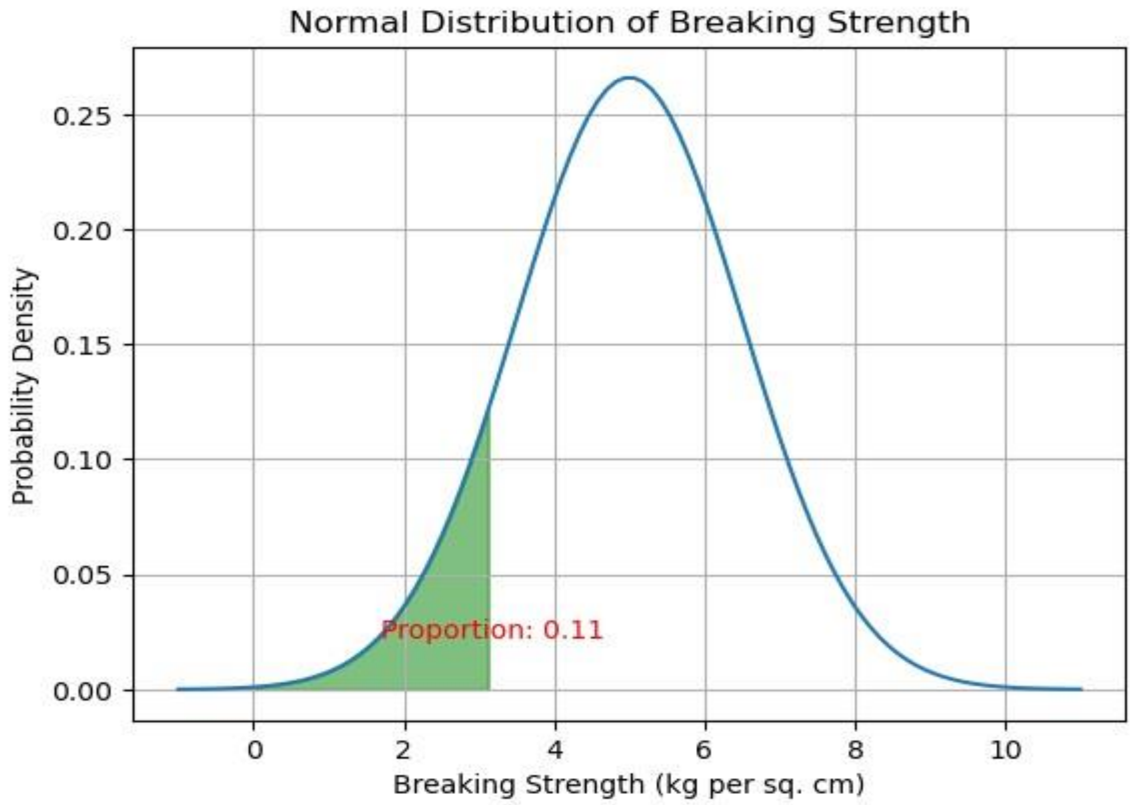
Here is the graph for the breaking strength of gunny bags used for packaging cement that is normally distributed with a mean of 5 kg per sq. centimetre and a standard deviation of 1.5 kg per sq. centimetre



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

Answer:

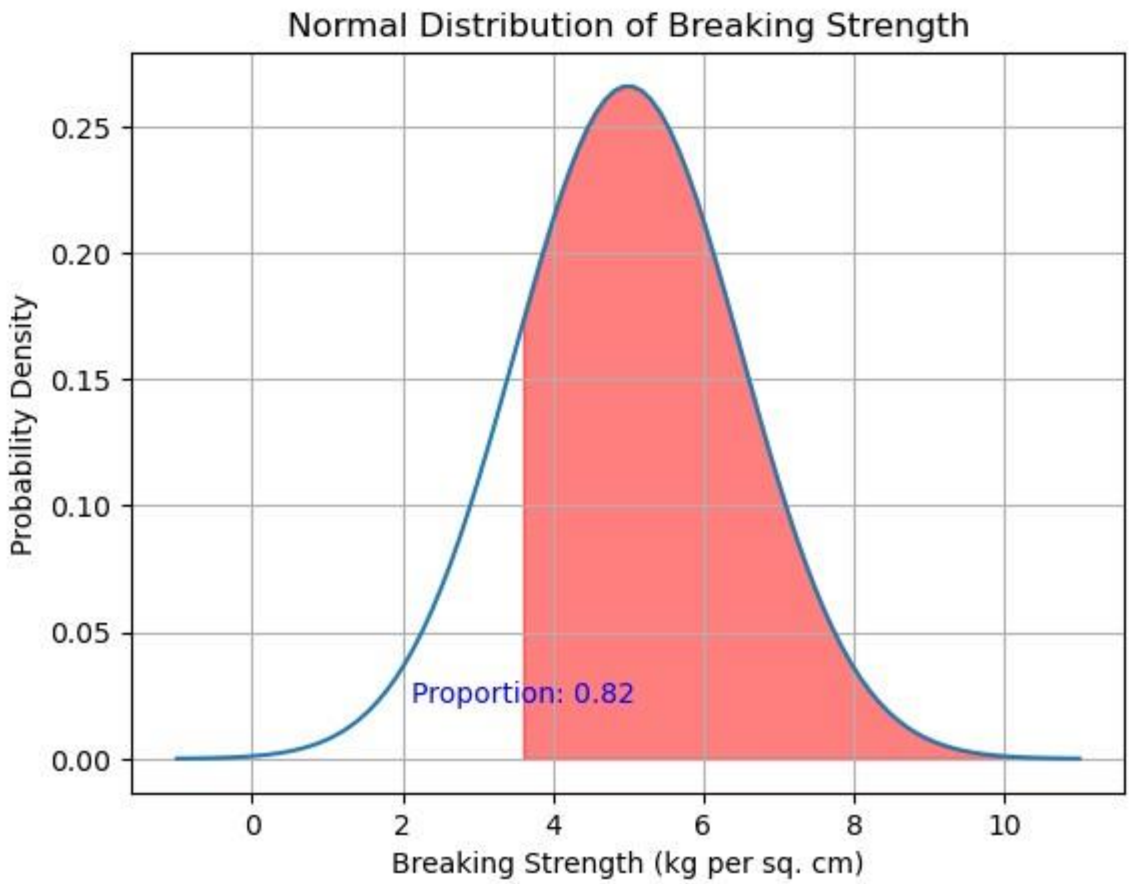
The proportion of the gunny bags have a breaking strength less than 3.17 kg per sq cm is 0.111



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

Answer:

The proportion of the gunny bags having a breaking strength at least 3.6 kg per sq c m is 0.82

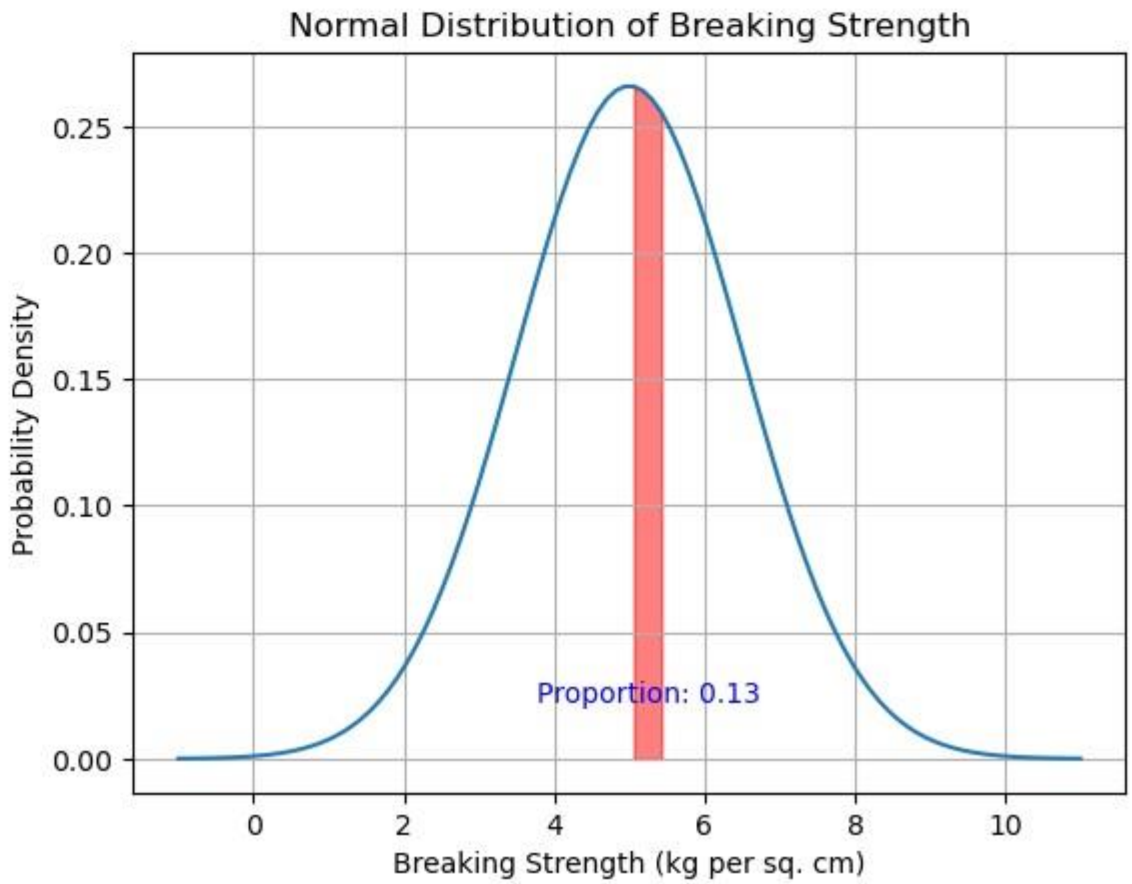


.?

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

Answer

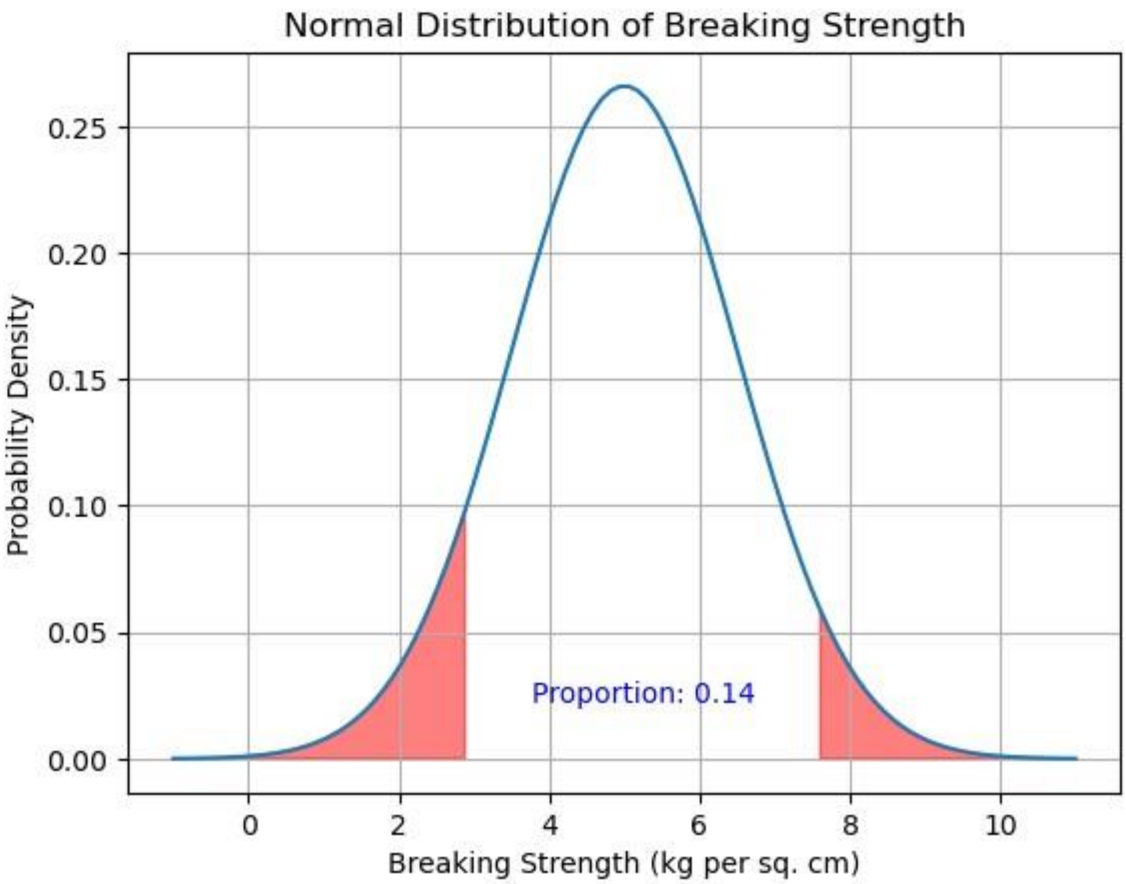
The proportion of the gunny bags having a breaking strength between 5 and 5.5 kg per is 0.13



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

Answer

The proportion of the gunny bags having a breaking strength NOT between 3 and 7.5 kg per sq cm is 0.14


Normal Distribution of Breaking Strength
Probability Density
0.25
0.20
0.15
0.10
0.05
0.00
Proportion: 0.14
0
2
4
6
8
10
Breaking Strength (kg per sq. cm)

Problem 4:

Grades of the final examination in a training course are found to be normally distributed, with a mean of 77 and a standard deviation of 8.5. Based on the given information answer the questions below.

4.1 What is the probability that a randomly chosen student gets a grade below 85 on this exam?

4.2 What is the probability that a randomly selected student scores between 65 and 87?

4.3 What should be the passing cut-off so that 75% of the students clear the exam?

Answers:

Solution

$$probability\_below\_85 = norm.cdf(85, loc = mean, scale = std\_dev)$$

The probability of getting a grade below 85 is: 0.8267

4.2 What is the probability that a randomly selected student scores between 65 and 87?

$$probability\_between = norm.cdf(87, loc = mean, scale = std\_dev) - norm.cdf(65, loc = mean, scale = std\_dev)$$

The probability of scoring between 65 and 87 is: 0.8013

4.3 What should be the passing cut-off so that 75% of the students clear the exam?

$$passing\_cutoff = norm.ppf(0.75, loc = mean, scale = std\_dev)$$

The passing cut-off for 75% of the students is: 82.73.

Problem 5:

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);

5.1 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. Do you think Zingaro is justified in thinking so?

Answer

First we can define the null and alternative hypothesis :

- Null Hypothesis (H0): The mean hardness of unpolished stones is greater than or equal to 150 ($\mu \geq 150$).
- Alternative Hypothesis (H1): The mean hardness of unpolished stones is less than 150 ($\mu < 150$).

We can perform one-sample t-test. Using a significance level of 0.05, we can compare the calcuated t_statistic and p_value to determine if there is enough evidence to reject the null hypothesis or not.

Here after performing one-sample t-test we get the results below:

T-Statistic: -4.164629601426757
P-Value: 4.171286997419652e-05.

Since P-Value <0.05 we can reject the Null Hypothesis.

Hence Zingaro is justified in thinking that unpolished stones may not be suitable for Printing as the mean hardness of unpolished stones is less than 150 ($\mu < 150$).

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

To determine whether the mean hardness of the polished and unpolished stones is the same, we can perform a hypothesis test. Let's denote the mean hardness of polished stones as $\mu 1$ and the mean hardness of unpolished stones as $\mu 2$.

- Null Hypothesis (H0): Mean hardness of polished stones is equal to the Mean hardness of unpolished stones ($\mu 1 = \mu 2$).

- Alternative Hypothesis (H1): Mean hardness of polished stones is not equal to the mean hardness of unpolished stones ($\mu 1 \neq \mu 2$).

We can perform a two-sample t-test to compare the means of two independent samples. The test will determine if there is enough evidence to reject the null hypothesis and conclude that the mean hardness of the polished and unpolished stones is different.

Here after performing a two-sample t-test we get the following results:

T-Statistic: -3.2422320501414053
P-Value: 0.0007327575097314177.

Since P-Value <0.05 we can reject the Null Hypothesis.

Hence we can say that the mean hardness of polished and unpolished stones are Different.

6. Aquarius health club, one of the largest and most popular cross-fit 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 more than 5 push-ups, as compared to when he/she enrolled in the program. Using the sample data provided can you conclude whether the program is successful? (Consider the level of Significance as 5%) Note that this is a problem of the paired-t-test. Since the claim is that the training will make a difference of more than 5, the null and alternative hypotheses must be formed accordingly.

This is a problem of the paired-t-test. Since the claim is that the training will make a difference of more than 5, the null and alternative hypotheses must be formed accordingly.

First we can define the null and alternative hypothesis:

- Null Hypothesis (H0): The mean push-ups the candidate is able to do is less than or equal to 5.
  ($\mu <= 5$).
- Alternative Hypothesis (H1): The mean push-ups the candidate is able to do is more than 5. ($\mu > 5$).

The output we get after performing paired-t-test is

T-Statistic: 19.322619811082458
P-Value: 2.2920419252511966e-35.

Here the level of significance is 0.05.

Since the P-Value is 2.2920419252511966e-35 < 0.05 (which is much lesser than the level of Significance as 5%) we reject the null hypothesis.

Hence the Alternative Hypothesis (H1): The mean push-ups the candidate is able to do is more than 5. ($\mu > 5$) stands true and the program is successful.

Problem 7:

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.

1. Test whether there is any difference among the dentists on the implant hardness. State the null and alternative hypotheses. Note that both types of alloys cannot be considered together. You must state the null and alternative hypotheses separately for the two types of alloys.?

We can solve this using ANNOVA

For Alloy 1:

Null hypothesis (H0): The mean implant hardness is the same among the dentists for Alloy 1 (μ1 = μ2)/There is no significant difference among the dentists on the hardness of dental implants for Alloy1.

Alternative hypothesis (HA): There is a difference in mean implant hardness among the dentists for Alloy 1 (μ1 ≠ μ2).

Here we get the output as below:

Here we get the output as :

	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

Here from the output we can see that the p-value is greater than 0.05 (assuming a 5% significance level).

Therefore, we fail to reject the null hypothesis in this case for alloy 1.

This suggests that there is not enough evidence to conclude a significant difference in mean implant hardness among the dentists for alloy1.

Hence we can say the Null Hypothesis :The mean implant hardness is the same among the dentists for Alloy 1 (μ1 = μ2) stands true.

For Alloy 2:

Null hypothesis (H0): The mean implant hardness is the same among the dentists for Alloy 2 (μ1 = μ2)/There is no significant difference among the dentists on the hardness of dental implants for Alloy2.

Alternative hypothesis (HA): There is a difference in mean implant hardness among the dentists for Alloy 2 ($\mu1 \neq \mu2$).

Here we get the output as below:

	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

Here from the output we can see that the p-value is greater than 0.05 (assuming a 5% significance level).

Therefore, we fail to reject the null hypothesis in this case for alloy 2.

This suggests that there is not enough evidence to conclude a significant difference in mean implant hardness among the dentists for alloy2.

Hence we can say the Null Hypothesis :The mean implant hardness is the same among the dentists for Alloy 2 ($\mu1 = \mu2$) stands true.

So for either of the alloys here is not enough evidence to conclude a significant difference in mean implant hardness among the dentists.Hence we can say the Null Hypothesis :The mean implant hardness is the same among the dentists for both the alloys ($\mu1 = \mu2$) stands true.

7.2 Before the hypotheses may be tested, state the required assumptions. Are the assumptions fulfilled? Comment separately on both alloy types.?

Answer

1. Independent Sample - Sample should be selected randomly(Equally likely events). There should not be any pattern in the selection of sample
2. Normal Distribution - Distribution of each group should be normal
3. Homogenous Group - Variance between the group should be the same.

7.3 Irrespective of your conclusion in 7.2, we will continue with the testing procedure. What do you conclude regarding whether implant hardness depends on dentists? Clearly state your conclusion. If the null hypothesis is rejected, is it possible to identify which pairs of dentists differ?

Answer

For Alloy 1:

Null hypothesis (H0): The mean implant hardness is the same among the dentists for Alloy 1.

Alternative hypothesis (HA): There is a difference in mean implant hardness among the dentists for Alloy 1.

The p-value for Alloy 1 is 0.116567, which is greater than the significance level of 0.05. Therefore, we fail to reject the null hypothesis. This suggests that there is not enough evidence to conclude a significant difference in mean implant hardness among the dentists for Alloy 1.

For Alloy 2:

Null hypothesis (H0): The mean implant hardness is the same among the dentists for Alloy 2.

Alternative hypothesis (HA): There is a difference in mean implant hardness among the dentists for Alloy 2.

The p-value for Alloy 2 is 0.718031, which is also greater than the significance level of 0.05. Therefore, we fail to reject the null hypothesis for Alloy 2 as well.

This indicates that there is not enough evidence to conclude a significant difference in mean implant hardness among the dentists for Alloy 2.

In conclusion, based on the ANOVA results, there is no significant evidence to suggest that implant hardness depends on the dentists, neither for Alloy 1 nor for Alloy 2. Therefore, we do not have enough information to identify specific pairs of dentists that differ in terms of implant hardness. If null hypothesis is rejected it is possible to identify which pairs of dentists differ through Tukey Honest Significant Difference method.

7.4 Now test whether there is any difference among the methods on the hardness of dental implant, separately for the two types of alloys. What are your conclusions? If the null hypothesis is rejected, is it possible to identify which pairs of methods differ?

For Alloy1

Null hypothesis (H0): There is no significant difference among the methods on the hardness of dental implants./The mean implant hardness is the same among the methods for Alloy 1 ($\mu 1 = \mu 2$) .

Alternative hypothesis (HA): There is a significant difference among the methods on the hardness of dental implants for Alloy1.

ANOVA results 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

Based on the ANOVA results for Alloy 1:

The null hypothesis (H0) states that there is no significant difference among the methods on the hardness of dental implants for Alloy 1.

The alternative hypothesis (HA) states that there is a significant difference among the methods on the hardness of dental implants for Alloy 1.

The p-value associated with the F-statistic is 0.004163, which is less than the significance level of 0.05. Therefore, we reject the null hypothesis.

In conclusion, there is sufficient evidence to suggest that there is a significant difference among the methods on the hardness of dental implants for Alloy 1.

For Alloy2

ANOVA results for 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

The null hypothesis (H0) states that there is no significant difference among the methods on the hardness of dental implants for Alloy 2.

The alternative hypothesis (HA) states that there is a significant difference among the methods on the hardness of dental implants for Alloy 2.

The p-value associated with the F-statistic is 0.000005, which is less than the significance level of 0.05. Therefore, we reject the null hypothesis.

In conclusion, there is sufficient evidence to suggest that there is a significant difference among the methods on the hardness of dental implants for Alloy 2.

Based on Tukey HSD test for the comparisons between different methods on the hardness of dental implants, separately for Alloy 1 and Alloy 2.

For Alloy 1:

The comparison between Method 1 and Method 2 does not show a significant difference in mean hardness ($p > 0.05$). The comparison between Method 1 and Method 3 shows a significant difference in mean hardness ($p < 0.05$), indicating that Method 3 differs significantly from Method 1 in terms of implant hardness. The comparison between Method 2 and Method 3 also shows a significant difference in mean hardness ($p < 0.05$), indicating that Method 3 differs significantly from Method 2 in terms of implant hardness.

For Alloy 2:

The comparison between Method 1 and Method 2 does not show a significant difference in mean hardness ($p > 0.05$). The comparison between Method 1 and Method 3 shows a significant difference in mean hardness ($p < 0.05$), indicating that Method 3 differs significantly from Method 1 in terms of implant hardness. The comparison between Method 2 and Method 3 also shows a significant difference in mean hardness ($p < 0.05$), indicating that Method 3 differs significantly from Method 2 in terms of implant hardness. In summary, for both Alloy 1 and Alloy 2, there are significant differences in mean hardness between Method 3 and the other two methods (Method 1 and Method 2). However, there is no significant difference in mean hardness between Method 1 and Method 2 for both.

7.5 Now test whether there is any difference among the temperature levels on the hardness of dental implant, separately for the two types of alloys. What are your conclusions? If the null hypothesis is rejected, is it possible to identify which levels of temperatures differ?

<u>For Alloy 1:</u>

Null hypothesis (H0): There is no significant difference among the temperature levels on the hardness of dental implants./The mean temperature level is the same among the methods for Alloy 1 ($\mu1 = \mu2$)

Alternative hypothesis (HA): There is a significant difference among the temperature level on the hardness of dental implants for Alloy1.

ANOVA results for Alloy 1:

	df	sum_sq	mean_sq	F	PR(>F)
C(Temp)	2.0	10154.444444	5077.222222	0.335224	0.717074
Residual	42.0	636122.800000	15145.780952	NaN	NaN

Based on the ANOVA results for Alloy 1:

The null hypothesis (H0) states that there is no significant difference among the temperature levels on the hardness of dental implants for Alloy 1.

The alternative hypothesis (HA) states that there is a significant difference among the temperature levels on the hardness of dental implants for Alloy 1.

The ANOVA table shows that the p-value associated with the F-statistic is 0.717074, which is greater than the significance level of 0.05. Therefore, we fail to reject the null hypothesis.

This means that there is not enough evidence to suggest a significant difference among the temperature levels on the hardness of dental implants for Alloy 1. The mean temperature level is likely the same among the methods for Alloy 1.

In other words, based on the available data, we do not have sufficient evidence to conclude that the temperature levels have a significant effect on the hardness of dental implants for Alloy 1.

<u>For Alloy 2:</u>

ANOVA results for Alloy 2:

	df	sum_sq	mean_sq	F	PR(>F)

C(Temp)	2.0	9.374893e+04	46874.466667	1.883492	0.164678
Residual	42.0	1.045254e+06	24886.996825	NaN	NaN

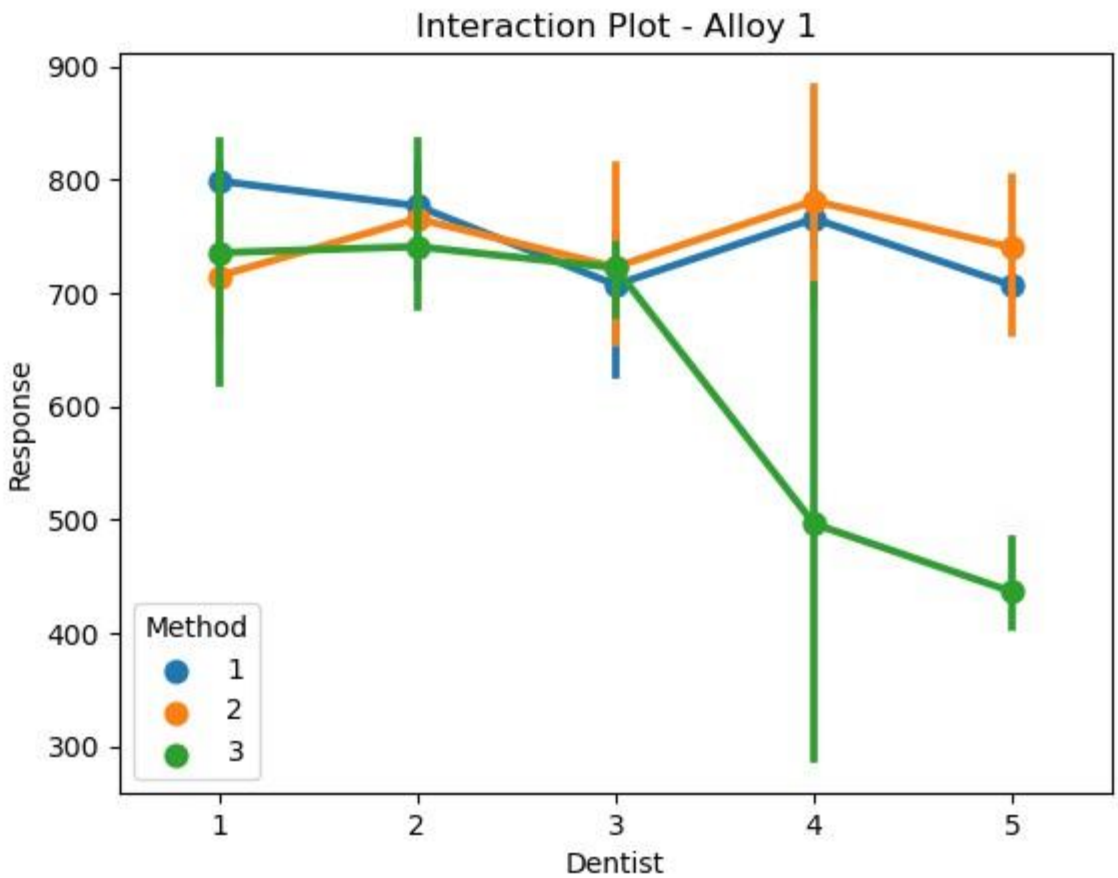
Based on the ANOVA results for Alloy 2:

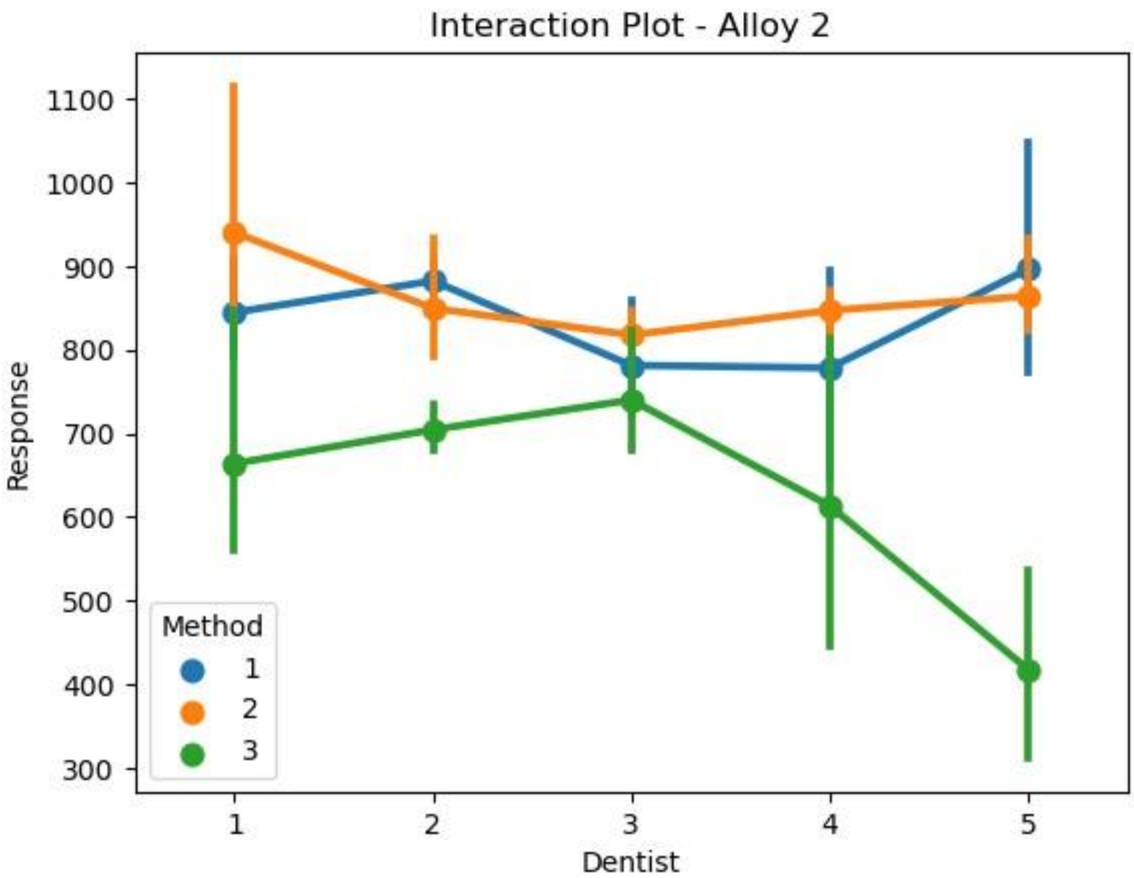
The null hypothesis (H0) states that there is no significant difference among the temperature levels on the hardness of dental implants for Alloy 2. The alternative hypothesis (HA) states that there is a significant difference among the temperature levels on the hardness of dental implants for Alloy 2. The ANOVA table shows that the p-value associated with the F-statistic is 0.164678, which is greater than the significance level of 0.05. Therefore, we fail to reject the null hypothesis.

This means that there is not enough evidence to suggest a significant difference among the temperature levels on the hardness of dental implants for Alloy 2. The mean temperature level is likely the same among the methods for Alloy 2.

In other words, based on the available data, we do not have sufficient evidence to conclude that the temperature levels have a significant effect on the hardness of dental implants for Alloy 2. If null hypothesis is rejected in both cases it is possible to identify which pairs of dentists differ through Tukey Honest Significant Difference method.

7.6 Consider the interaction effect of dentist and method and comment on the interaction plot, separately for the two types of alloys?





Here for alloy 1 we can see that Dentist 4 and Dentist 5 is having much variance in the mean of response variable which indicate the implant hardness for Alloy1. Even though from annova analysis we failed to reject the null hypothesis which suggested that there is not enough evidence to conclude a significant difference in mean implant hardness among the dentists for alloy1 this can be due to noise in the data which can affect the mean.

Similarly for methods among alloy 1 we can see that Method 3 differs from Method 1 and Method 2, we have proved that also using Tukey HSD test.

For Alloy 2 Dentist 1, Dentist 4 and Dentist 5 is having much variance in the mean of response variable which indicate the implant hardness for Alloy2. Even though from annova analysis we failed to reject the null hypothesis which suggested that there is not enough evidence to conclude a significant difference in mean implant hardness among the dentists for alloy2 this can be due to noise in the data which can affect the mean.

Similarly for methods among alloy 1 we can see that Method 3 differs from Method 1 and Method 2, we have proved that also using Tukey HSD test.

7.7 Now consider the effect of both factors, dentist, and method, separately on each alloy. What do you conclude? Is it possible to identify which dentists are different, which methods are different, and which interaction levels are different?

For Alloy1

ANOVA results 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

Based on a significance level of 0.05, we can restate the conclusions from the ANOVA results as follows:

For Alloy 1:

C(Dentist): The factor "Dentist" has a significant effect on the hardness of dental implant (p-value = 0.0115 < 0.05). Therefore, we reject the null hypothesis and conclude that there is a significant difference among the dentists in terms of the hardness of dental implant.

C(Method): The factor "Method" has a significant effect on the hardness of dental implant (p-value = 0.0003 < 0.05). Thus, we reject the null hypothesis and conclude that there is a significant difference among the methods in terms of the hardness of dental implant.

C(Dentist):C(Method) (Interaction effect): The interaction between "Dentist" and "Method" is significant (p-value = 0.0068 < 0.05). Hence, we reject the null hypothesis and conclude that the effect of the method on the hardness of dental implant depends on the dentist, and vice versa.

For Alloy 2:

ANOVA results 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

C(Dentist): The factor "Dentist" does not have a significant effect on the hardness of dental implant (p-value = 0.3718 > 0.05). Therefore, we fail to reject the null hypothesis and conclude that there is no significant difference among the dentists in terms of the hardness of dental implant.

C(Method): The factor "Method" has a significant effect on the hardness of dental implant (p-value = 0.0000 < 0.05). Hence, we reject the null hypothesis and conclude that there is a significant difference among the methods in terms of the hardness of dental implant.

C(Dentist):C(Method) (Interaction effect): The interaction between "Dentist" and "Method" is not significant (p-value = 0.0932 > 0.05). Therefore, we fail to reject the null hypothesis and conclude that the effect of the method on the hardness of dental implant does not depend significantly on the dentist, and vice versa. In summary, for Alloy 1, both the dentist and method have significant effects on the hardness of dental implant, and their interaction is also significant. However, for Alloy 2, only the method has a significant effect, and the interaction effect is not significant.

It is possible to identify which dentists are different, which methods are different, and which interaction levels are different by using Tukey HSD test as we have done for earlier questions.