Advance Statistics Project

Business Report

Pavithra Devi
PGPDS
Great Learning

Index

Contents

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?
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.
7.1 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.?
7.2 Before the hypotheses may be tested, state the required assumptions. Are the assumptions fulfilled? Comment separately on both alloy types.?
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?
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?
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?

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

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?

Probability of an event is defined as the ratio of two numbers m and n.
P(A)=m/n
Where, m = Number of favourable outcomes
n = Total number of possible outcomes.

Using the formula, we get:
P(that a randomly chosen player would suffer an injury)= Players injured/Total number of players
=145/235
=0.617

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

Here, the event is mutually exclusive(meaning: event A or B occurs and both cannot simultaneously occur).
P(A union B)= P(A) +P(B)

P(Player is a forward)=m/n=94/235=0.4
P(Player is a Winger)=m/n=29/235=0.12340426
P(Player is a forward union Player is a Winger)= P(Player is a Forward) +P(Player is a Winger)
= 0.4+0.1234
=0.52

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

P(A)=m/n
P(Striker and Foot Injury) = Number of players in the striker position who have a foot injury / Number of players in the striker position
=45/77= 0.58

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

P(Injured player is a striker)=Number of strikers injured/Total injured players
=45/145
=0.31

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

P(Injured players who are Forward)=number of injured players who are forward/Total number of injured players.
=56/145
=0.386

P(Injured players who are Midfielders)=number of injured players who are midfielders/Total number of injured players.
=24/145
=0.165

P(Injured player is a Forward or Attacking midfielder)= P(Injured players who are Forward)+P(Injured players who are Midfielder)
=0.386+0.165
=0.55

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:

Solution:
2.1 Probabilities of a fire, a mechanical failure, and a human error, respectively:

Let's denote:
- $P(F)$ = Probability of a fire
- $P(M)$ = Probability of a mechanical failure
- $P(H)$ = Probability of a human error

We are given the following information:
- Probability of radiation leak given a fire: $P(R|F) = 20\% = 0.20$
- Probability of radiation leak given a mechanical failure: $P(R|M) = 50\% = 0.50$
- Probability of radiation leak given a human error: $P(R|H) = 10\% = 0.10$

Also, we know that:
- Probability of radiation leak occurring simultaneously with a fire: $P(R \cap F) = 0.1\% = 0.001$
- Probability of radiation leak occurring simultaneously with a mechanical failure: $P(R \cap M) = 0.15\% = 0.0015$
- Probability of radiation leak occurring simultaneously with a human error: $P(R \cap H) = 0.12\% = 0.0012$

Using this information, we can find the probabilities:

$P(F) = 1 - P(R \cap F) = 1 - 0.001 = 0.999$
$P(M) = 1 - P(R \cap M) = 1 - 0.0015 = 0.9985$
$P(H) = 1 - P(R \cap H) = 1 - 0.0012 = 0.9988$

[Note: These values represent the probabilities of no radiation leak occurring along with each type of accident, making them the complementary probabilities. Subtracting the probabilities of radiation leaks occurring simultaneously with each type of accident from 1 gives us the desired probabilities]

2.2 Probability of a radiation leak (P(R)):
Since two or more types of accidents cannot occur simultaneously, we can assume that the probability of an accident happening is the sum of the probabilities of each type of accident:

P(R) = Probability of a radiation leak

$P(R) = P(R|F) * P(F) + P(R|M) * P(M) + P(R|H) * P(H)$
$P(R) = 0.20 * 0.999 + 0.50 * 0.9985 + 0.10 * 0.9988$
$P(R) = 0.1998 + 0.49925 + 0.09988$
$P(R) = 0.79893$

The probability of a radiation leak is approximately 79.89%.

2.3 Probability that the radiation leak has been caused by:

A Fire (P(F|R)):

Using Bayes' theorem:
$P(F|R) = P(R|F) * P(F) / P(R)$
$P(F|R) = 0.20 * 0.999 / 0.79893$
$P(F|R) = 0.2497$

A Mechanical Failure (P(M|R)):

$P(M|R) = P(R|M) * P(M) / P(R)$
$P(M|R) = 0.50 * 0.9985 / 0.79893$
$P(M|R) = 0.6276$

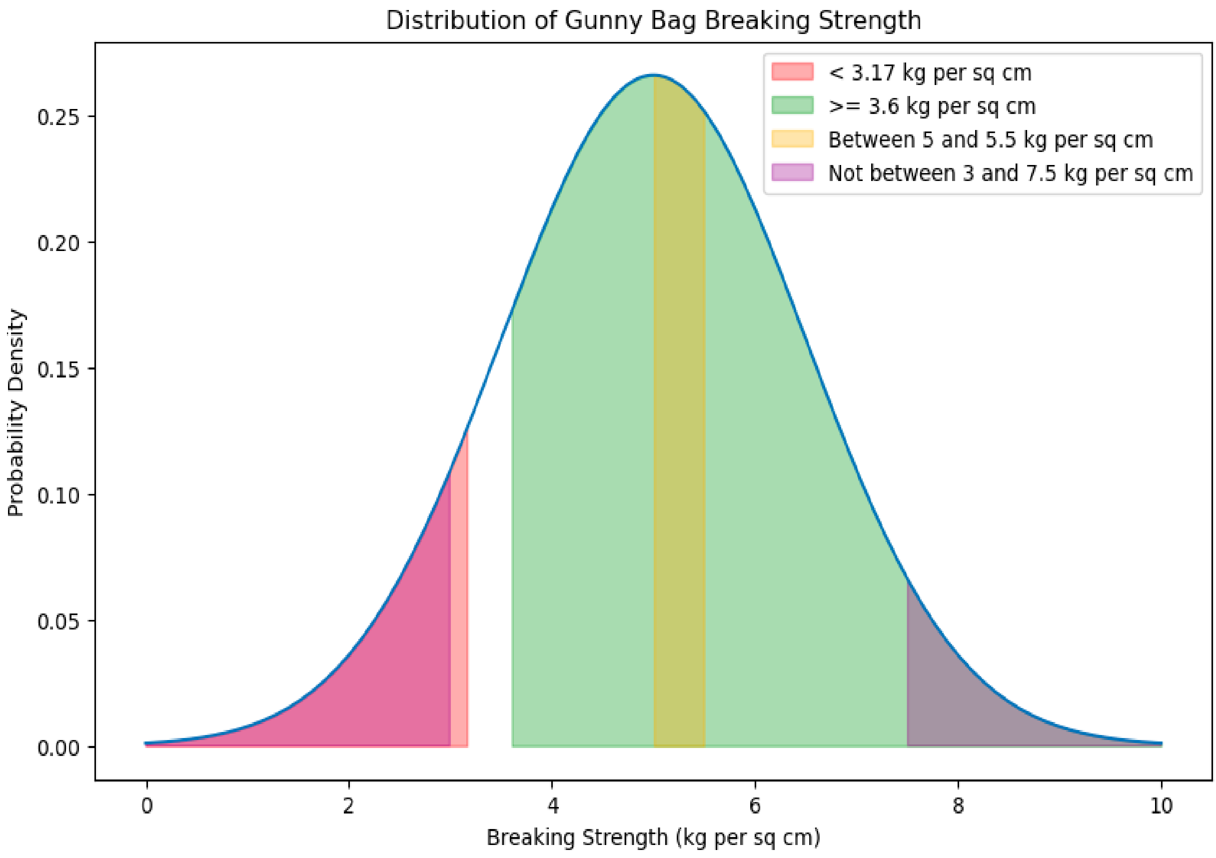
A Human Error (P(H|R)):

$P(H|R) = P(R|H) * P(H) / P(R)$
$P(H|R) = 0.10 * 0.9988 / 0.79893$
$P(H|R) =0.1245$

So, the probability that the radiation leak has been caused by:
- A Fire =24.97%
- A Mechanical Failure = 62.76%
- A Human Error =12.45%

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)

Solution:



3.1 What proportion of the gunny bags have a breaking strength less than 3.17 kg per sq cm?
Proportion of gunny bags with breaking strength less than 3.17 kg per sq cm: 0.1112
It can be interpreted as 11.1% gunny bags have a breaking strength les s 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.?
Proportion of gunny bags with breaking strength at least 3.6 kg per sq cm: 0.8247

3.3 What proportion of the gunny bags have a breaking strength between 5 and 5.5 kg per sq cm.?
Proportion of gunny bags with breaking strength between 5 and 5.5 k g per sq cm: 0.1306

3.4 What proportion of the gunny bags have a breaking strength NOT between 3 and 7.5 kg per sq cm.?
Proportion of gunny bags with breaking strength not between 3 and 7. 5 kg per sq cm: 0.1390

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.

Solution:
4.1 What is the probability that a randomly chosen student gets a grade below 85 on this exam?
Probability that a randomly chosen student gets a grade below 85: 0.8267 or 82.6%

4.2 What is the probability that a randomly selected student scores between 65 and 87?
Probability that a randomly selected student scores between 65 and 87: 0.8013 or 80.1%

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

Solution:

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

We will perform a one-sample t-test since we have the sample mean and standard deviation and are comparing it to a known value (150).

Since the significance level (alpha) is given as 0.05 (5%), we will use this to determine the critical region for the t-test. If the p-value is less than or equal to alpha, we will reject the null hypothesis in favor of the alternative hypothesis.

Calculate the p-value and make the decision:
p-value= 4.171286997419652e-05
Since the p-value is less than alpha, we will reject the null hypothesis in favour of the alternative hypothesis.

The mean hardness of the unpolished stones is significantly lower tha n 150.

5.2 Is the mean hardness of the polished and unpolished stones the same?
No, mean hardness of the polished and unpolished stones are different.

Unpolished	134.110527
Treated and Polished	147.788117

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.

Solution:
To determine whether the program at Aquarius health club is successful in terms of body conditioning, specifically in terms of push-ups, we can use a paired t-test. The paired t-test compares the mean difference between two related samples to a specified value (in this case, 5 push-ups).

1. Formulate the hypotheses:
   - Null Hypothesis (H0): The mean difference in push-ups after the program is less than or equal to 5 (no significant improvement).
   - Alternative Hypothesis (H1): The mean difference in push-ups after the program is greater than 5 (significant improvement).

2. Set the significance level: The significance level (alpha) is given as 0.05, which means we will reject the null hypothesis if the p-value is less than 0.05.

3. Perform the hypothesis test:
   - Assuming the provided sample data represents the number of push-ups before and after the program, we need to calculate the differences and perform the paired t-test.

4. Interpret the results:
   - If the p-value is less than the significance level (0.05), we reject the null hypothesis in favor of the alternative hypothesis. It would indicate that there is a significant improvement in the number of push-ups after the program.
   - If the p-value is greater than or equal to the significance level, we fail to reject the null hypothesis, suggesting that there is not enough evidence to conclude that the program leads to a significant improvement in the number of push-ups.

Here, the p-value is less than the significance level (0.05), we reject the null hypothesis in favor of the alternative hypothesis.

Therefore, it can be concluded that the program at Aquarius health club is successful in terms of body conditioning, specifically in terms of push-ups.

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.

Solution:

7.1 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.?

To test whether there is any difference among the dentists on the implant hardness, we can conduct separate tests for the two types of alloys. Let's define the null and alternative hypotheses for each case:

For Alloy 1:
- Null Hypothesis (H0): There is no significant difference among the dentists in implant hardness for Alloy 1.
- Alternative Hypothesis (H1): There is a significant difference among the dentists in implant hardness for Alloy 1.

For Alloy 2:
- Null Hypothesis (H0): There is no significant difference among the dentists in implant hardness for Alloy 2.
- Alternative Hypothesis (H1): There is a significant difference among the dentists in implant hardness for Alloy 2.

The p-values for Alloy 1 and Alloy 2 will indicate whether there is a significant difference among the dentists in implant hardness for each alloy. If the p-value is less than the significance level (typically 0.05), we reject the null hypothesis and conclude that there is a significant difference among the dentists.

Conclusion stated in answer 7.3

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

Assumptions for Alloy Type 1 and 2:
1. Independence: The responses (implant hardness) should be independent of each other.
2. Normality: The distribution of responses for each dentist should be approximately normal. This assumption is important for parametric tests such as t-tests or ANOVA.
3. Homogeneity of Variance: The variance of responses should be equal among the dentists. This assumption is important for parametric tests like ANOVA.

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?

Conclusion (at $\alpha = 0.05$):
p-Value for Dentist factor (Alloy 1): 0.1165
P-Value for Dentist factor (Alloy 2): 0.718
Since, p-value is greater than the level of significance i.e., 0.05,we fail to reject the null hypothesis.
For Alloy 1, there is no significant difference in implant hardness among different dentists.
For Alloy 2, there is no significant difference in implant hardness among different dentists.

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 Alloy Type 1:
1. Null Hypothesis (H0): There is no difference among the methods on the hardness of dental implants for Alloy Type 1.
   Alternative Hypothesis (H1): There is a difference among the methods on the hardness of dental implants for Alloy Type 1.

For Alloy Type 2:
1. Null Hypothesis (H0): There is no difference among the methods on the hardness of dental implants for Alloy Type 2.
   Alternative Hypothesis (H1): There is a difference among the methods on the hardness of dental implants for Alloy Type 2.

p-Value for Method factor (Alloy 1): 0.004163412167505505
p-Value for Method factor (Alloy 2): 5.415871051443187e-06
Since, p-value is lesser than the level of significance i.e., 0.05,we reje ct the null hypothesis.
For Alloy 1, there are significant differences in implant hardness amo ng different methods.
For Alloy 2, there are significant differences in implant hardness amo ng different methods.

Yes, it is possible to identify which pairs of methods differ.

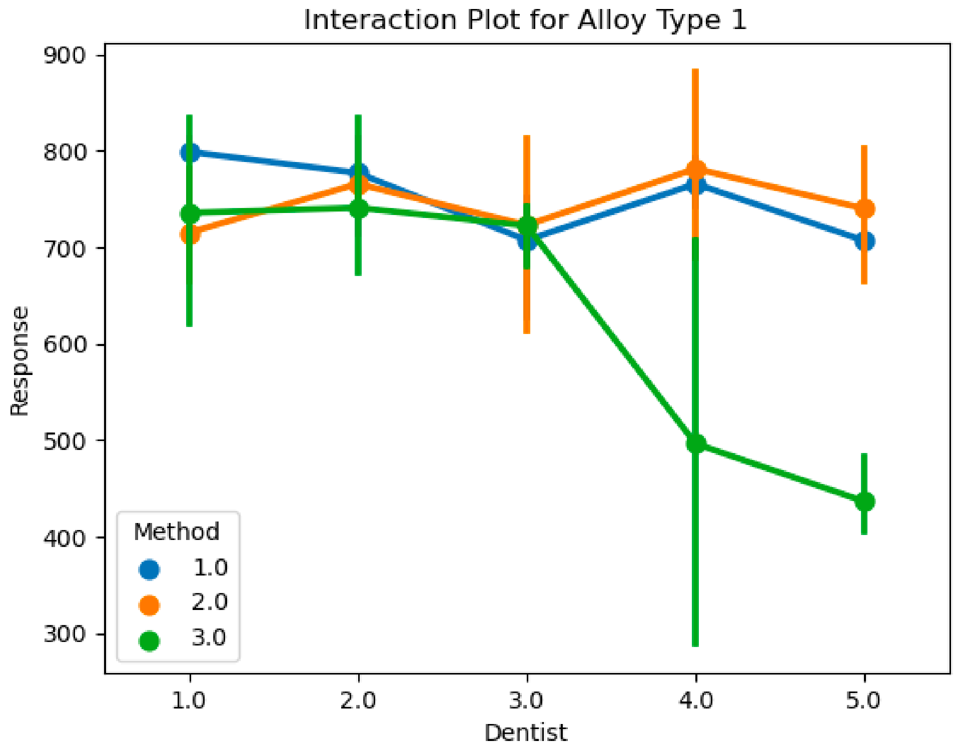
7.5 Now test whether there is any difference among the temperature levels on the hardness of dental implant, separately for the two types of alloys. What are your conclusions? If the null hypothesis is rejected, is it possible to identify which levels of temperatures differ?

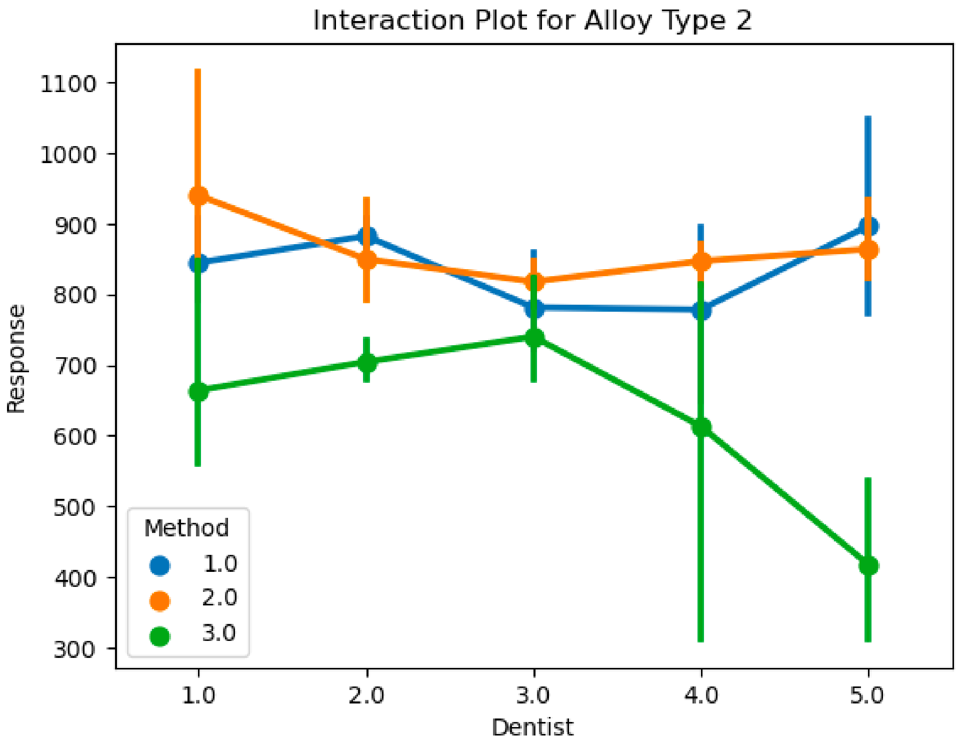
p-Value for Temp factor (Alloy 1): 0.4136182281695441
p-Value for Temp factor (Alloy 2): 0.06724592833336664
Since, p-value is greater than the level of significance i.e., 0.05,we fail to reject the null hypothesis

For Alloy 1, there is no significant difference in implant hardness among different temperatures.
For Alloy 2, there is no significant difference in implant hardness among different temperatures.

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

The lines cross each other for Alloy 1, it indicates a significant interaction effect. In this case, the effect of Dentist on implant hardness varies across different Methods, indicating that the two factors interact.





The method 1 and 2 lines cross each other for Alloy 2, it indicates a significant interaction effect. In this case, the effect of Dentist on implant hardness varies across different Methods, indicating that the two factors interact for method 1 and 2…whereas method 3 is points are nearly parallel, it suggests no significant interaction effect between Dentist and Method.

7.7 Now consider the effect of both factors, dentist, and method, separately on each alloy. What do you conclude?

To summarize the conclusions from the two-way ANOVA results for Alloy Type 1 and Alloy Type 2:

Alloy Type 1:

1. Dentist Effect: The p-value associated with the main effect of Dentist is 0.011484, which is less than the chosen significance level of 0.05. This indicates that there are significant differences in implant hardness among different dentists for Alloy Type 1.

2. Method Effect: The p-value associated with the main effect of Method is 0.000284, which is much smaller than the significance

level (0.05). This indicates that there are significant differences in implant hardness among different methods for Alloy Type 1.

3. Interaction Effect: The p-value associated with the interaction effect between Dentist and Method is 0.006793, which is less than the significance level (0.05). This indicates that there is a significant interaction effect between Dentist and Method on implant hardness for Alloy Type 1.

Alloy Type 2:

1. Dentist Effect: The p-value associated with the main effect of Dentist is 0.371833, which is greater than the chosen significance level of 0.05. This suggests that there are no significant differences in implant hardness among different dentists for Alloy Type 2.

2. Method Effect: The p-value associated with the main effect of Method is 0.000004, which is much smaller than the significance level (0.05). This indicates that there are significant differences in implant hardness among different methods for Alloy Type 2.

3. Interaction Effect: The p-value associated with the interaction effect between Dentist and Method is 0.093234, which is greater than the significance level (0.05). This suggests that the interaction effect between Dentist and Method is marginally not significant at the chosen significance level for Alloy Type 2.

Overall Conclusion:

1. For Alloy Type 1, both Dentist and Method have significant main effects on implant hardness, and there is a significant interaction effect between Dentist and Method.

2. For Alloy Type 2, the main effect of Method is significant, indicating that different implantation methods significantly influence implant hardness. However, the main effect of Dentist is not

significant for Alloy Type 2, suggesting that there are no significant differences in implant hardness among different dentists.

3. The interaction effect for Alloy Type 2 is marginally not significant.

Is it possible to identify which dentists are different, which methods are different, and which interaction levels are different?

For Alloy 1:
Dentist 4 and 5 it is different.
Method 3 is different.
Interaction between dentist 4 and 5 for method 3 is different.

For Alloy 2:
Dentist 1,2, 4 and 5 it is different.
Method 3 is different.
Interaction between method 1and 2 with 3 method is different.

Thank you.