Inferential Statistics Project

G S Jaiguruam
Data Scientist

CONTENTS

Problem 1

Statement: 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.

Data Table:

	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?

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

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

The probability that a player is a forward or a winger is 0.523.

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

The probability that a player is a striker and has foot injury is 0.118.

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

The probability that a randomly chosen injured player is a striker is 0.31.

Problem 2

Statement: 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?

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

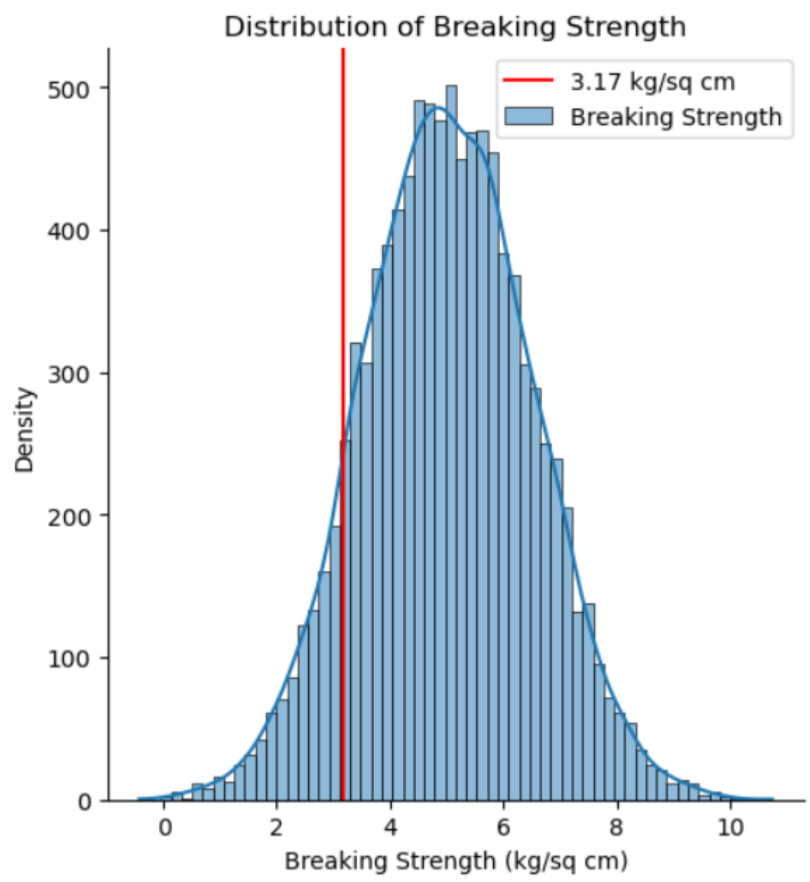

Fig 1

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

```
The proportion of gunny bags with a breaking strength of at least 3.6 kg per sq cm
is 0.825.
```

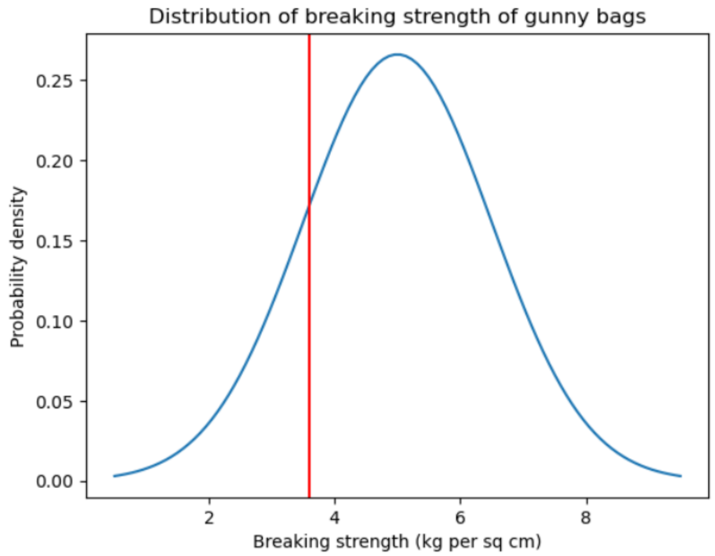

Fig 2

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

```
The probability that a randomly selected gunny bag can withstand between 5 and 5.5
 kg per sq. centimeter is 0.131.
```

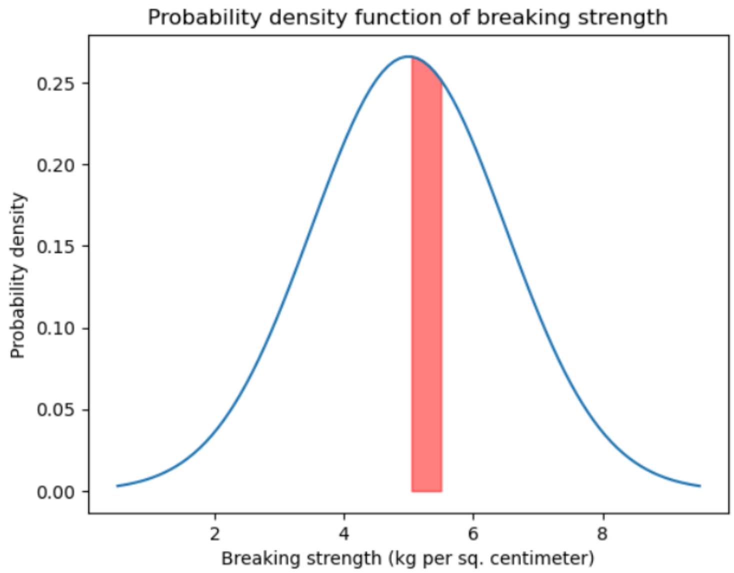

Fig 3

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

```
The proportion of gunny bags with a breaking strength NOT between 3 and 7.5 kg per
 sq cm is 0.139.
```

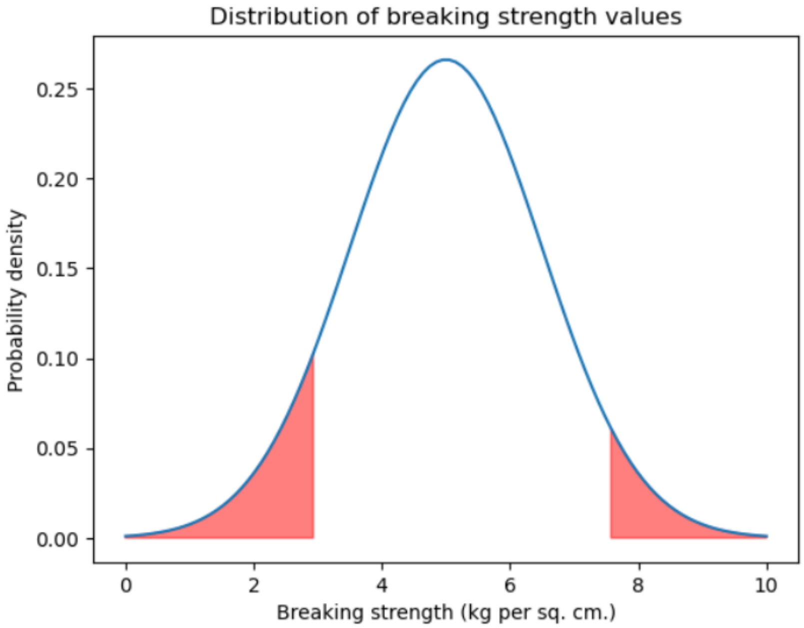

Fig 4

Problem 3

Statement: 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?

```
Hypotheses:
Null hypothesis:
    H0: mu_polished - mu_unpolished = 150
Alternative hypothesis:
    Ha: mu_polished - mu_unpolished < 150
```

Conducted the hypothesis test using z_score method and compute the p-value as 0.0 which is less then the assumed significance level of 5%.

Hence,

We reject the null hypothesis and conclude that there is evidence to suggest that the unpolished stones may not be suitable for printing.

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

```
Hypotheses:
Null hypothesis:
    H0: mu_polished = mu_unpolished
Alternative hypothesis:
    Ha: mu_polished ≠ mu_unpolished
```

Conducted the hypothesis test using z_score method and compute the p-value as 0.0 which is less than the assumed significance level of 5%.

Hence,

Reject the null hypothesis. There is significant evidence to suggest that the mean hardness of the polished and unpolished stones is not the same.

Problem 4

Statement: 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.

Note – Provide data was group into to group based on the alloy type provided in the datasheet

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

```
Hypotheses:
Null Hypothesis (H0):
    H0: mu1 = mu2 = mu3 = mu4 = mu5
    The mean hardness of implants is the same across all dentists.
Alternative Hypothesis (Ha):
    Ha: At least one of the means is different.
    At least one pair of dentists has a different mean hardness of implants.

Assumptions of Hypotheses:
Normality: The hardness measurements for implants from each dentist are normal
ly distributed.
Homogeneity of variances: The variances of hardness measurements for implants
are approximately equal across all dentists.

We check the assumptions and Concluded the below points:
The Shapiro-Wilk test results suggest that the normality assumption is general
ly met for both alloy1 and alloy2 data across all dentists.
The Levene's test results suggest that the variances in both alloy1_data and a
lloy2_data are homogeneous for all dentists.
Assumption are met to conduct the ANOVA test.

We conducted a one-way ANOVA test and computed p-Value as 0.72, so we fail to r
eject the null hypothesis.

The results of the one-way ANOVA tests suggest that the mean hardness of both
alloys varies depending on dentists.
This means that the type of alloy used and the dentist who performs the procedure
both have a significant impact on the hardness of the dental implants.

To identify for which pairs it differs we did pairwise comparison and We concluded
that:
For Alloy 1:
The pairwise comparison table shows that the mean hardness of the alloy1_data
is significantly different between the following pairs of dentists: Dentist 1
and Dentist 2 Dentist 1 and Dentist 3 Dentist 1 and Dentist 4 Dentist 2 and
Dentist 4 Dentist 3 and Dentist 4.
```

Explanation:
This means that the mean hardness of the alloy1_data is significantly higher for Dentist 1 compared to Dentists 2, 3, and 4. It is also significantly higher for Dentist 2 compared to Dentist 4, and for Dentist 3 compared to Dentist 4.

For Alloy 2:
The pairwise comparison table shows that the mean hardness of the alloy2_data is significantly different between the following pairs of dentists: Dentist 1 and Dentist 2 Dentist 1 and Dentist 3 Dentist 1 and Dentist 4 Dentist 2 and Dentist 4 Dentist 3 and Dentist 4.
Explanation:
This means that the mean hardness of the alloy2_data is significantly higher for Dentist 1 compared to Dentists 2, 3, and 4. It is also significantly higher for Dentist 2 compared to Dentist 4, and for Dentist 3 compared to Dentist 4.

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

Hypotheses:
Null Hypothesis (H0):
    H0: mu1 = mu2 = mu3
    The mean hardness of implants is the same across all methods.
Alternative Hypothesis (Ha):
    Ha: At least one of the means is different.
    At least one pair of methods has a different mean hardness of implants.

Assumptions of Hypotheses:
Normality: The hardness measurements for implants within each method group should follow a normal distribution
Homogeneity of variances: The variances of hardness measurements for implants should be approximately equal across all method groups.

We check the assumptions and Concluded the below points:
The Shapiro-Wilk test results suggest that the data in both alloy1_data and al loy2_data is normally distributed for all methods.
The Levene's test results suggest that the variances in both alloy1_data and a lloy2_data are homogeneous for all methods.
Assumption are met to conduct the ANOVA test.

We conducted one-way ANOVA test and computed the p-values, which are below the conventional significance level of 0.05 for both alloy1 and alloy 2 group set, indicating that there is strong evidence against the null hypothesis.
The results of the one-way ANOVA tests suggest that the mean hardness of both a lloys varies depending on methods. This means that the type of alloy used and t he method used to perform the procedure both have a significant impact on the h ardness of the dental implants.

To identify for which pairs it differs, we did pairwise comparison and We concluded that:
The pairwise comparison tables show that the mean hardness of both alloys is significantly different between the following pairs of methods:

- Alloy 1:
  - Method 1 and Method 3
  - Method 2 and Method 3
- Alloy 2:
  - Method 1 and Method 3
  - Method 2 and Method 3

This means that the results of the one-way ANOVA tests are valid and that the differences in mean hardness between methods are not due to chance.

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

Interaction plot between dentist and method on hardness of implant represented by response: -

For Alloy 1:

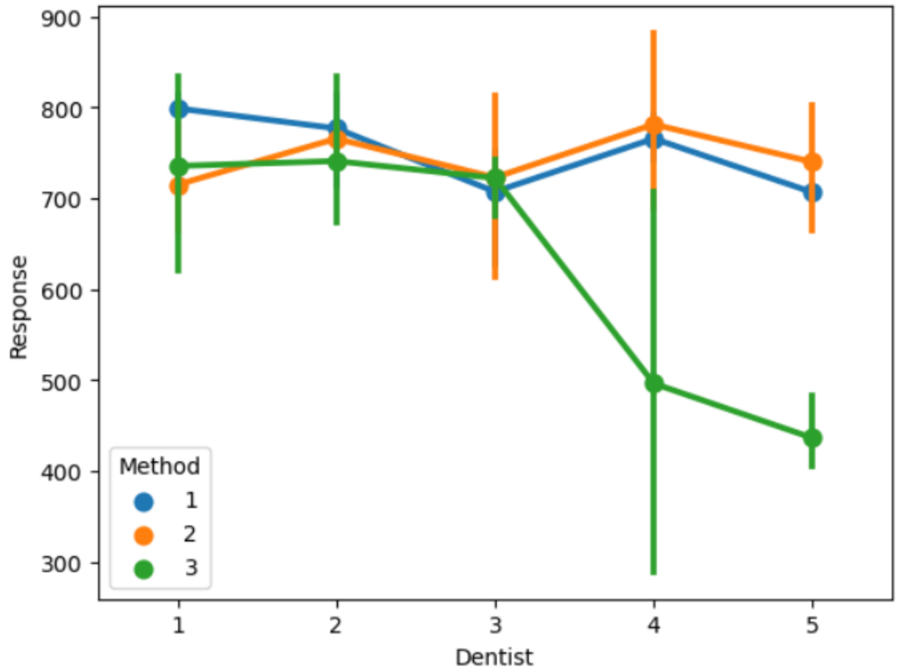

Fig 5

For Alloy 2:

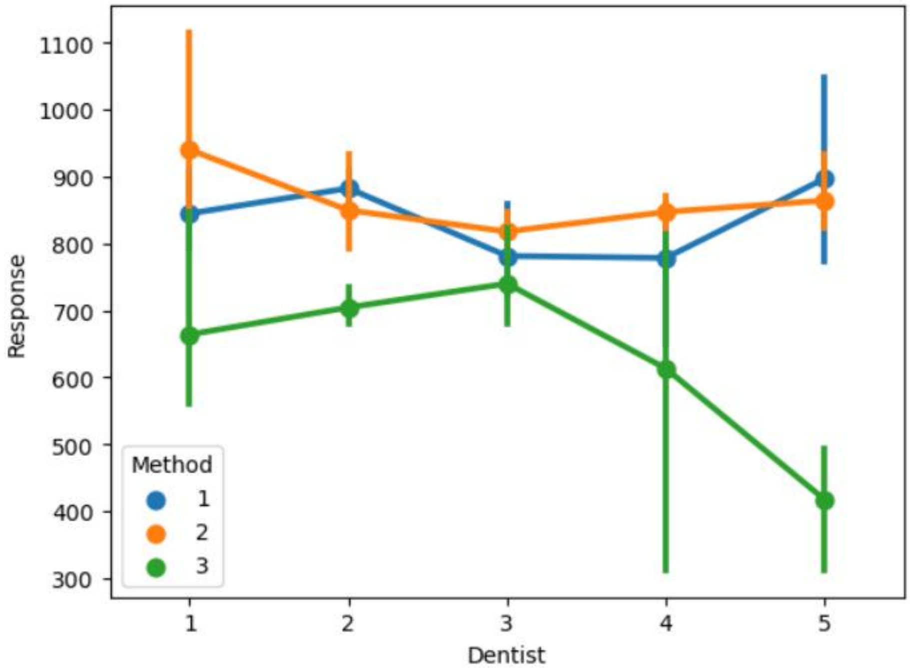

Fig 6

Inferences from the plot:

For Alloy 1
Method 1 and Method 2 seem to have similar responses across all dentists, as their lines are close together and follow a similar trend.
Method 3, Shows a distinct difference, especially with Dentist 4, where the response is significantly lower than the other methods.
For Method 3, there is a noticeable downward trend from Dentist 3 to Dentist 5, suggesting that the response for this method decreases for these dentists.
Method 1 and Method 2 appear to be more consistent across different dentists compared to Method 3, which has a larger variation, especially with Dentist 4.
The highest average response for any method appears to be with Dentist 1 using Method 1, while the lowest average response is with Dentist 5 using Method 3.

For Alloy 2:
Method 1 and Method 2 seem to have similar responses across all dentists, as their lines are close together and follow a similar trend.
Method 3 generally has a lower response level compared to the other two methods.
There is a noticeable downward trend in the response levels for Method 3 from Dentist 3 to Dentist 5, suggesting that this method is less effective or less preferred as we move to higher dentist numbers.
The highest response level recorded is for Method 2 with Dentist 2, while the lowest is for Method 3 with Dentist 5.

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

Hypotheses:
Null Hypothesis (H0):
H0: mu111 = mu112 = mu113 = mu121 = mu122 = mu123 = mu131 = mu132 = mu133
There is no interaction effect between dentists and methods on the hardness of implants.
Alternative Hypothesis (Ha):
Ha: At least one of the means is different.
There is an interaction effect between dentists and methods on the hardness of implants.

Assumptions of the hypothesis test:
Normality: The response variable should be normally distributed for each combination of the independent variables (dentists and methods).
Homogeneity of variances: The variances of the response variable should be equal for each combination of the independent variables.

We Checked the assumption and concluded below points:
Shapiro-Wilk test suggested that for most combinations of dentist and method, the data appear to be normally distributed.
In both alloys, the Levene's test suggests that the assumption of homogeneity of variances are met across different combinations of dentists. Therefore, the variances of hardness values do not significantly differ between dentists for both Alloy 1 and Alloy 2.

We conducted two-way ANOVA test and computed the p-values, which are below the conventional significance level of 0.05 for both alloy1 and alloy 2. So, Reject the null hypothesis. Both Alloy 1 and Alloy 2 show significant variation in mean hardness depending on both dentists and methods, as indicate d by the low p-values from the two-way ANOVA tests.

To Identify which dentists and methods combinations are different, and which intera ction levels are different we did pairwise comparison and concluded that:

For alloy 1
Based on the pairwise comparison analysis for Alloy 1, we reject the null hypothesis of equality of means for several combinations of dentists and methods.
There is significant evidence to suggest that the mean hardness of dental implants vary depending on the specific combination of dentists and methods used.
For alloy 2
There are pairwise comparisons with p-values below 0.05, indicating significant differences in mean hardness between various combinations of dentists and methods.
Hence, we reject the null hypothesis of equality of means for Alloy 2 as well.
The analysis suggests that the mean hardness of dental implants for Alloy 2 also varies significantly depending on the specific combination of dentists and met hods used.

List of figures