Inferential Statistics Project
BUSINESS REPORT

PGP DSA 22 JUNE 2025

ANJU SAINI

CONTENT

Problem 1

Problem 2

Problem 3

Problem 4

Graphs

1.1) Introduction:-

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.

1.2) Structure :-

	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**

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

To calculate the probability that a randomly chosen player would suffer an injury,

We need to divide the number of injured players by the total number of players.

Number of injured players = 145

Total number of players = 235

Probability of a randomly chosen player suffering an injury = Number of injured

players / Total number of players

Probability = 145 / 235 ≈ 0.617

or 61.7%

So, the probability that a randomly chosen player would suffer an injury is

approximately 0.617 or 61.7%.

1.4) What is the probability that a player is a forward or a winger?

The total number of players in the forward position is 94, and the total number of

Players in the winger position are 29.

As these are mutually exclusive events, a player can be a forward or winger.

We can calculate the probability as:

Probability of being a forward or a winger = (Number of forwards + Number of wingers) / Total number of players

Probability of being a forward or a winger = (94 + 29) / 235

Probability of being a forward or a winger = 123 / 235

Probability of being a forward or a winger ≈ 0.523 or 52.3%

Therefore, the probability that a player is a forward or a winger is approximately

0.523 or 52.3%.

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

To calculate this probability, we need to consider the number of players who

play in the striker position and have a foot injury, divided by the total

number of players.

Number of players in striker position = 77

Number of injured players who play in striker position = 45

Total number of players = 235

Probability of being a striker with a foot injury = Number of injured strikers / Total number of players

Plugging in the values:

Probability of being a striker with a foot injury =

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%.

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

To find the probability that a randomly chosen injured player is a striker, we need to divide the number of injured players in the striker position by the total number of injured players:

Probability of being a striker given injury = (Number of injured players in striker position) / (Number of injured players)

Probability of being a striker given injury = 45 / 145 ≈ 0.310

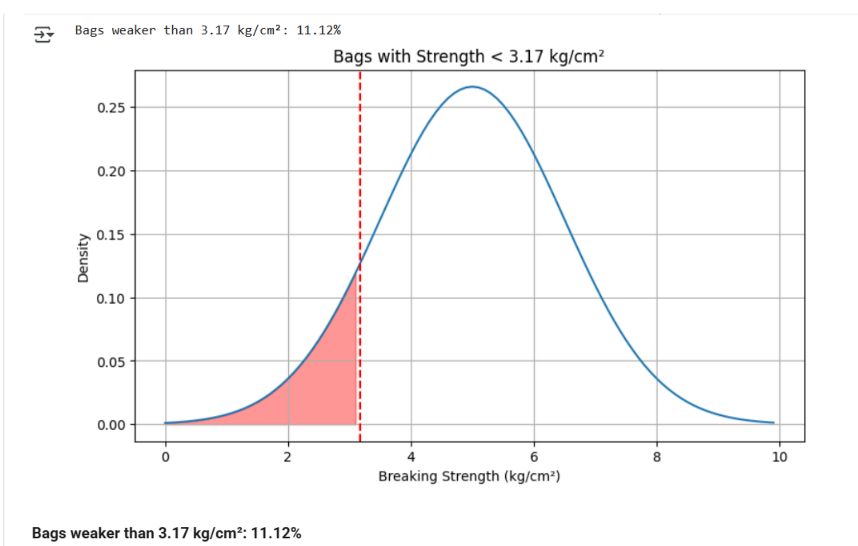
Therefore, the probability that a randomly chosen injured player is a striker is approximately 0.310 or 31.0%.

2.1) Introduction:-

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;

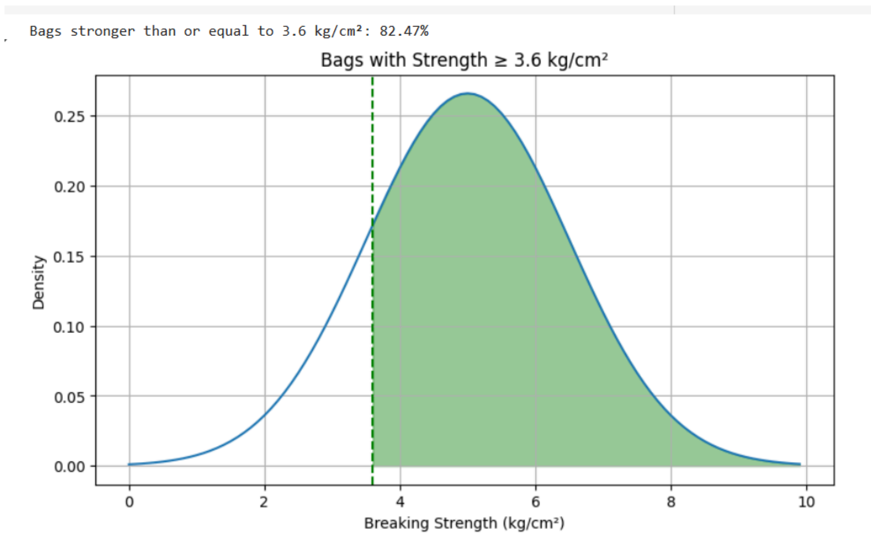
2.2 What proportion of the gunny bags have a breaking strength less than 3.17 kg per sq cm?

By looking at the below graph we can observe the proportion of gunny bags having a breaking strength less than
3.27 kg per sq. cm.is 0.1112 = 11.12%.



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

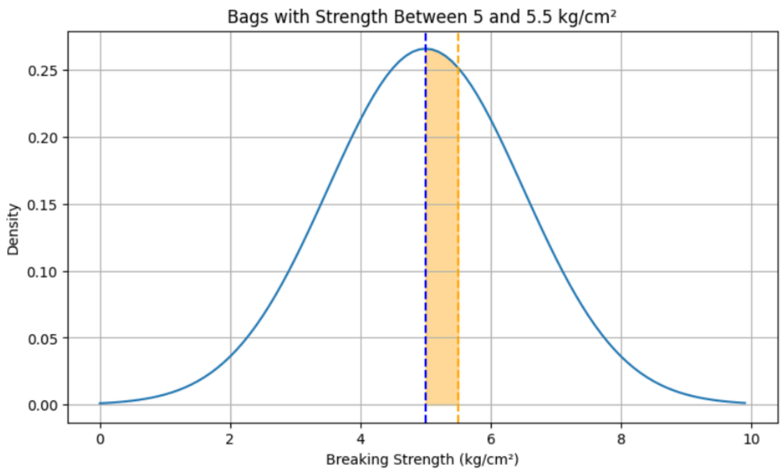
By looking at the below graph we can easily observe the proportion of gunny bags having a breaking strength at least
3.6 kg per sq. cm. is 0.8247 = 82.47%



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

From the below graph we can have a look at the proportion of the gunny bags having
a breaking strength between 5 and 5.5 kg per sq cm. = 0.1306 = 13.06%

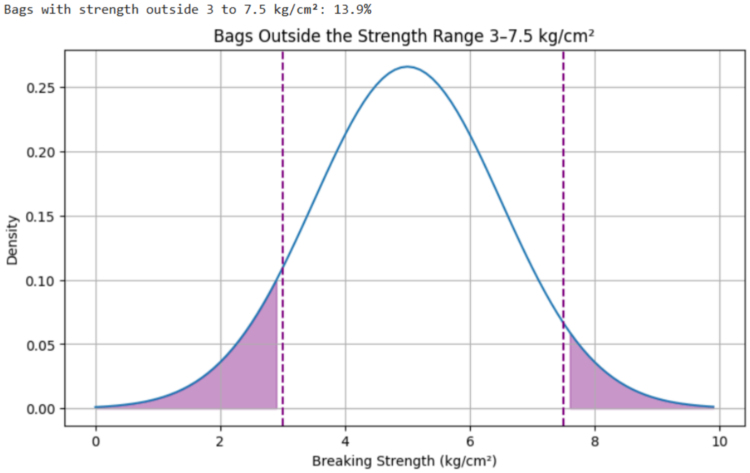
Bags between 5 and 5.5 kg/cm²: 13.06%



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

By having a look at the below graph we can observe the proportion if the gunny bags
having a breaking strength Not between 3 and 7.5 kg per sq cm. = 0.1390 = 13.90%

Bags with strength outside 3 to 7.5 kg/cm²: 13.9%



3.1) Introduction:-

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.

3.2) Objective:-

This report summarizes the findings of a statistical analysis conducted on a recent batch of polished and unpolished stones to determine their suitability for printing. The analysis focused on the Brinell hardness index, with a minimum threshold of 150 required for optimum printing.

3.3) Data Overview:-

Understanding the data and Structure

- There are 75 rows and 2 columns in the dataset.

```
(75, 2)
```

There are 75 rows and 2 columns in the dataset.

- Data Type of Different columns

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 75 entries, 0 to 74
Data columns (total 2 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   Unpolished            75 non-null     float64
 1   Treated and Polished  75 non-null     float64
dtypes: float64(2)
memory usage: 1.3 KB
```

Both The columns Unpolished and Treat and Polished are Float Type

3.4) Statistical Summary

	count	mean	std	min	25%	50%	75%	max
Unpolished	75.0	134.110527	33.041804	48.406838	115.329753	135.597121	158.215098	200.161313
Treated and Polished	75.0	147.788117	15.587355	107.524167	138.268300	145.721322	157.373318	192.272856

Max value of its is 200.16 and 1922.27 and min value for unpolished is 48.40 and 107.52 for treate and polished

Observation

Max value of its is 200.16 and 1922.27 and min value for unpolished is 48.40 and 107.52 for treat and polished

6) Checking Missing Data

	0
Unpolished	0
Treated and Polished	0

dtype: int64

Based on the given response, There is no any missing data in given fields

7)Key Insights

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?

Solution :-

Given Hypotheses:

- **H_0 (Null):** $\mu \geq 150$
- **H_1 (Alternate):** $\mu < 150$

This matches the correct setup for a **left-tailed one-sample test**.

Test Chosen:

- The sample size is **n = 75**, which is **greater than 30**, so using a **z-test** (or large-sample t-test) is acceptable.

Given Values:

- Sample Mean (Unpolished) = **134.11**
- Sample Std Dev (Unpolished) = **33.04**
- $\alpha = 0.05$
- p-value ≈ **0.00** (very small)

Conclusion:

- Since **p-value < 0.05**, **reject H_0**
- So yes, **Zingaro is justified** — unpolished stones likely do **not** meet the required hardness of 150.

Hence,Null Hypothesis Should Be Rejected .And Alternative Hypothesis Should Be Accepted . So Zingaro is justified in that Unpolished Stone may be not suitable For Printing

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

Solution:

Given Hypotheses:

- **H_0:** $\mu_1 = \mu_2$ (means are equal)
- **H_1:** $\mu_1 \neq \mu_2$ (means are different)

Correct Test Chosen:

- Since you're comparing **two independent samples**, a **two-sample t-test** is appropriate.

Given Values:

- Mean (Unpolished) = 134.11, Std = 33.04
- Mean (Polished) = 147.79, Std = 15.59
- Sample size = 75 each
- p-value = **0.0007328** < **0.05**

Conclusion:

- Since p-value < 0.05, **reject H_0**
- Therefore, the means are **significantly different**.

Hence,we can reject null Hypothesis and conclude the mean hardness of the polished and unpolished stones are not the same.

4.1) Introduction:-

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.

4.2) Objective:-

- Assess the effect of the dentist on implant hardness.
- Evaluate the impact of the method of implantation.
- Identify any interaction effects between the dentist and the method across alloys.
- Understand combined effects for making informed recommendations.

4.3) Data Overview:

```
(90, 5)
```

There are 90 rows and 5 columns in the dataset.

There are 90 rows and 5 columns in the dataset.

Data Info:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 90 entries, 0 to 89
Data columns (total 5 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Dentist   90 non-null     int64
 1   Method    90 non-null     int64
 2   Alloy     90 non-null     int64
 3   Temp      90 non-null     int64
 4   Response  90 non-null     int64
dtypes: int64(5)
memory usage: 3.6 KB
```

All The columns are integer type and contain numbers . there is no null value in any of the columns.

4.4) Statistical Summary:-

	count	mean	std	min	25%	50%	75%	max
Dentist	90.0	3.000000	1.422136	1.0	2.0	3.0	4.0	5.0
Method	90.0	2.000000	0.821071	1.0	1.0	2.0	3.0	3.0
Alloy	90.0	1.500000	0.502801	1.0	1.0	1.5	2.0	2.0
Temp	90.0	1600.000000	82.107083	1500.0	1500.0	1600.0	1700.0	1700.0
Response	90.0	741.777778	145.767845	289.0	698.0	767.0	824.0	1115.0

Dentist: There are 5 unique dentists (ranging from 1 to 5).

Method: There are 3 unique methods (ranging from 1 to 3).

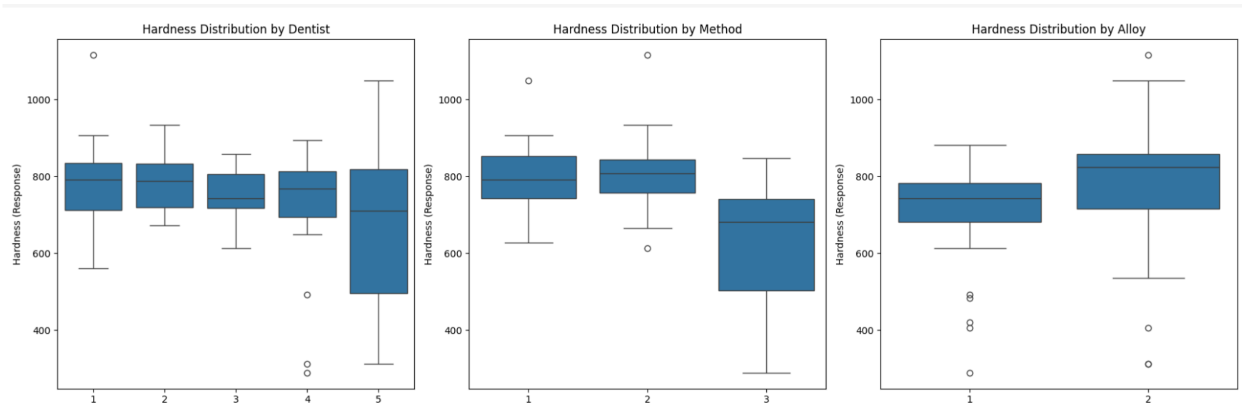
Alloy: There are 2 unique alloys (ranging from 1 to 2).

Temp: There are 3 unique temperatures (1500, 1600, and 1700).

Response: This is the variable of interest, representing hardness. The values range from 289 to 1115, with a mean of approximately 741.78 and a standard deviation of about 145.77.

All columns are of integer type and have no missing values.

4.5.1) Hardness Distribution by Dentist, Method, and Alloy (Boxplots)



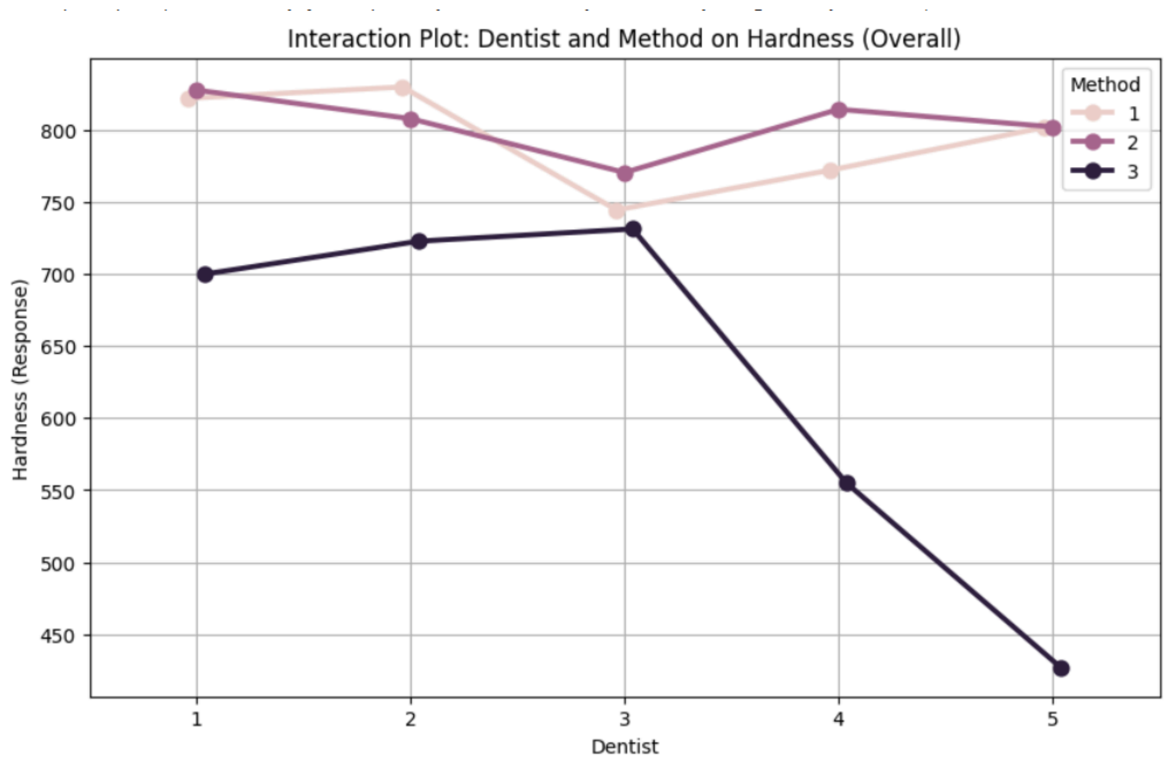
These boxplots give us a quick, visual way to understand how the hardness of implants changes across the three main factors we're looking at: Dentist, Method, and Alloy.

The Dentist plot shows the range and average hardness achieved by each dentist. It quickly highlights if certain dentists tend to produce harder or softer implants than others.

The Method plot lets us compare the different methods used, showing which methods generally lead to harder or softer implants.

The Alloy plot clearly shows if one type of alloy tends to be harder than the other.

4.5.2) Interaction Plot (Overall) — Dentist vs Method



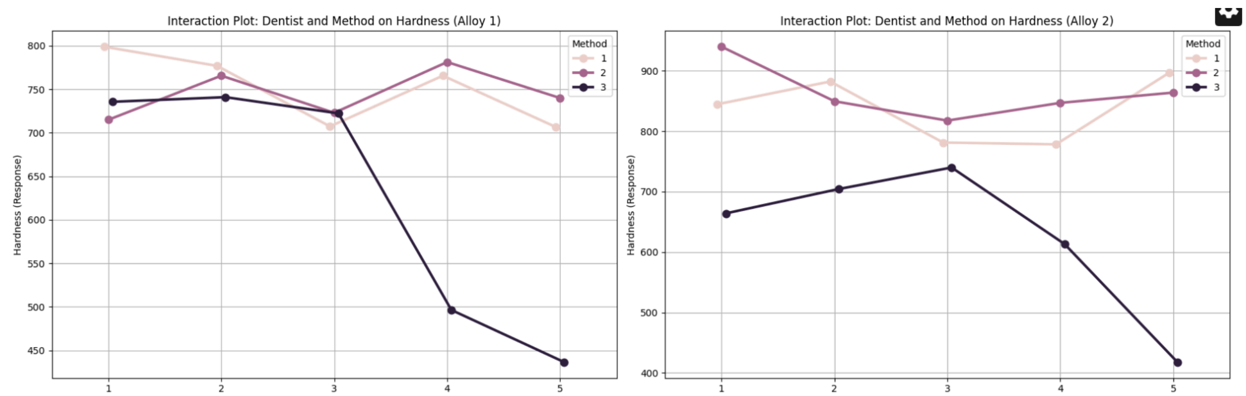
This point plot shows how the combination of Dentist and Method affects implant hardness across both alloys combined.

The lines represent different methods, and their position across the dentists shows how well each method works for each dentist.

If the lines are roughly parallel, it means the method works similarly across all dentists.

If the lines cross, it means certain methods work better for some dentists than others.

4.5.3) Interaction Plots by Alloy (Dentist vs Method)



These point plots break things down even further by looking at the interaction between Dentist and Method within each alloy.

In Alloy 1, you can see lines that cross or diverge sharply, suggesting certain methods work much better with certain dentists. This means getting the best results depends heavily on the specific pairing of method and dentist.

In Alloy 2, the lines run more or less parallel, which suggests the method works similarly across all dentists for that alloy.

This allows for precision when making decisions:

For Alloy 1, it’s crucial to consider both the method and the dentist when trying to achieve the best implant hardness.

For Alloy 2, you can focus more on choosing the best method, regardless of who the dentist is.

4.6) How does the hardness of implants vary depending on dentists?

Solution :

For Alloy 1, we found that the dentist can make a difference. The results show that some dentists achieve harder implants than others. This means the person doing the work matters when using this alloy.

For Alloy 2, it doesn't seem to matter which dentist is doing the work. The results were very similar across all dentists.

4.7) How does the hardness of implants vary depending on the method?

Solution :

For both Alloy 1 and Alloy 2, the method used makes a big difference. We found that Method 3 generally gives weaker results compared to Methods 1 and 2. So, no matter which alloy is used, choosing the right method is very important for getting strong and durable implants.

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

Solution :

For Alloy 1, the best method depends on which dentist is doing the work. In other words, some methods work better for certain dentists than others.
For Alloy 2, the method works roughly the same way no matter which dentist does the procedure. In this case, it doesn't matter who the dentist is — the method itself is what counts.

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

Solution :

When we consider both the dentist and the method at the same time, we see that both matter. The best results aren't just about choosing a good method or relying on a specific dentist — it's about finding the right match between the method and the person doing the work. In short, the combination of the dentist and the method gives the best results.

ThanKyou