AS – Project Guided

Advanced Statistics

ANIL KUMAR PGBPDSBA AUG’2023

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**

Based on the above data, answer the following questions.

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?

Solution :

Total Players = Players Injured + Players Not Injured

=145+90

=235

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

Ans 1.1 – The Probabilities of a randomly chosen player would suffer and injury is

Total no. of players = 235

Total no. of injured players = 145

Therefore, the Probability that a randomly chosen player would suffer an injury

= 145/235 = 0.627

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

Ans 1.2 –

Total no. of players who are forward = 94

Total no. of injured players = 145

Total no. of players who are wingers = 29

Total no. of players = 235

P(Forward) = 94/235 = 0.4

P(Winger) = 29/235 = 0.12

P(Forward or Winger) = P(F) + P(W) = 0.4 + 0.12 = 0.52

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

Ans 1.3 –

Total no. of players who are strikers = 45

Total no. of players = 235

P(Striker who has a foot injury) = 45/235

=0.19

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

Ans 1.4 –

Total no. of players who are strikers = 45

Total no. of injured players = 145

P(Injured player is a striker) = 45/145

=0.31

Problem 2

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)**

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

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

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

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

Q 2.1 What proportion of the gunny bags have a breaking strength of less than 3.17 kg per sq cm?

Ans 2.1 –

Mean = 5 kg ,

Std dev = 1.5 kg

Z = X – mean / std dev

= 3.17 – 5 / 1.5

-1.22

The proportion = stas.norm.cdf(-1.22) = 0.11123 = 11.1 %

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

Ans 2.2 –

Z = X – mean / std dev

= 3.6– 5 / 1.5

= -0.93

The proportion = (-0.93) = 0.823

= 82.3%

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

Ans 2.3 –

Z1 = (5-5) / 1.5

Z2 = (5.5 – 5) / 1.5

Stats.norm.cdf(Z2) – stats.norm.cdf(Z1)

= 0.1305

= 13.5%

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

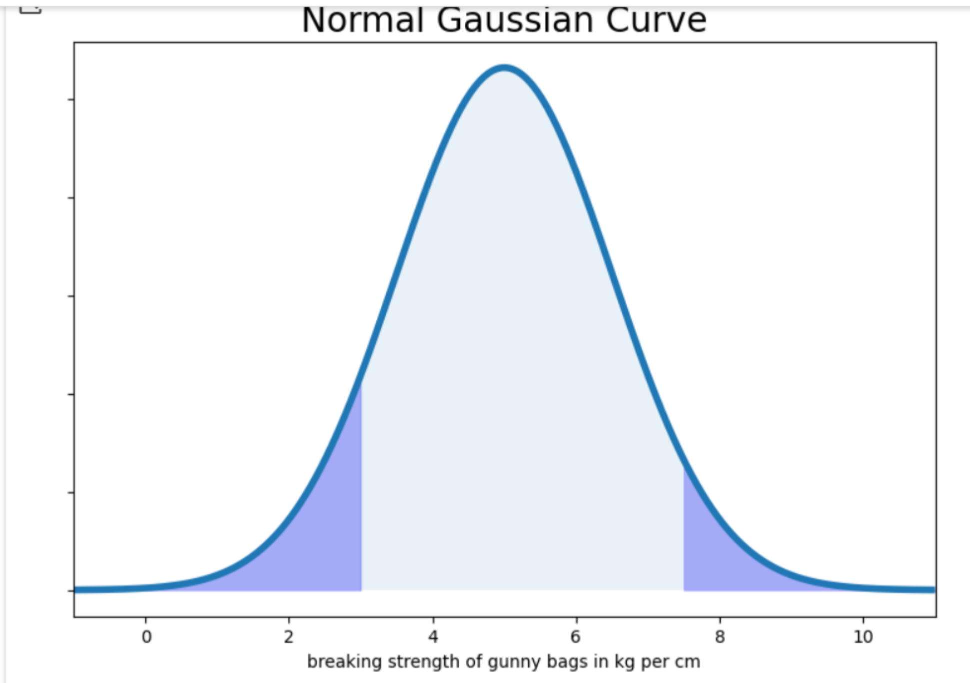
Ans 2.4 –

Z3 = (3 – 5) / 1.5

Z4 = (7.5 – 5) / 1.5

1 - Stats.norm.cdf(Z4) – stats.norm.cdf(Z3)

= 0.1390 = 14%



Problem 3

Zingaro stone printing is a company that specializes in printing images or patterns on polished or unpolished stones. However, for the optimum level of printing of the image, the stone surface has to have a Brinell's hardness index of at least 150. Recently, Zingaro has received a batch of polished and unpolished stones from its clients. Use the data provided to answer the following (assuming a 5% significance level);

3.1 Zingaro has reason to believe that the unpolished stones may not be suitable for printing. Do you think Zingaro is justified in thinking so?

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

Q 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?

Hypothesis:
Ho : $\mu \geq 150$
Null hypothesis: Unpolished stone is suitable for printing
Ha : $\mu < 150$
Alternative Hypothesis: Unpolished stone is not suitable for printing

(left tailed test)

```
u_mean = df['Unpolished '].mean()
u_std = df['Unpolished '].std()
cnt =  df['Unpolished '].count()
t_stat = (u_mean - 150) / (u_std / (cnt**0.5))
p_value = stats.t.sf(abs(t_stat), cnt-1)
```

```
t stat is  -4.164629601426758
```

df= n - 1 = 75 - 1 = 74

$\alpha = 0.05$
Crtical value = 1.6657 (from t-table)
since, absolute t statistics is greater than critical value
so, Null hypothesis should be rejected and alternative hypothesis should be accepted.

So, Zingaro is justified to think that unpolished stone is not suitable for printing.

Q3.2 Is the mean hardness of the polished and unpolished stones the same?

Ans:-

H0 : Mean hardness of the polished and unpolished stones are the same

H1 : Mean hardness of the polished and unpolished stones are not equal At 95% significance.

We use statistics to test if the polished and unpolished stones have the same hardness.
We use a test to see if the average hardness of the polished and unpolished stones is different.
If the result of the test is less than 5%, we can say that the polished and unpolished stones have different hardness.

```
# Two-sample t-test for polished and unpolished stones
p_mean = df['Treated and Polished'].mean()
u_mean = df['Unpolished '].mean()
p_std = df['Treated and Polished'].std()
u_std = df['Unpolished '].std()
n1 = df['Treated and Polished'].count()
n2 = df['Unpolished '].count()
sp = ((n1-1)*p_std**2 + (n2-1)*u_std**2) / (n1+n2-2)
t_stat = (p_mean - u_mean) / (sp * ((1/n1) + (1/n2)))**0.5
p_value = stats.t.sf(abs(t_stat), n1+n2-2)
if p_value < 0.05:
    print("The mean hardness of the polished and unpolished stones is different.")
else:
    print("The mean hardness of the polished and unpolished stones is the same.")
```

Problem 4

Dental implant data: The hardness of metal implants in dental cavities depends on multiple factors, such as the method of implant, the temperature at which the metal is treated, the alloy used as well as the dentists who may favour one method above another and may work better in his/her favorite method. The response is the variable of interest.

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

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

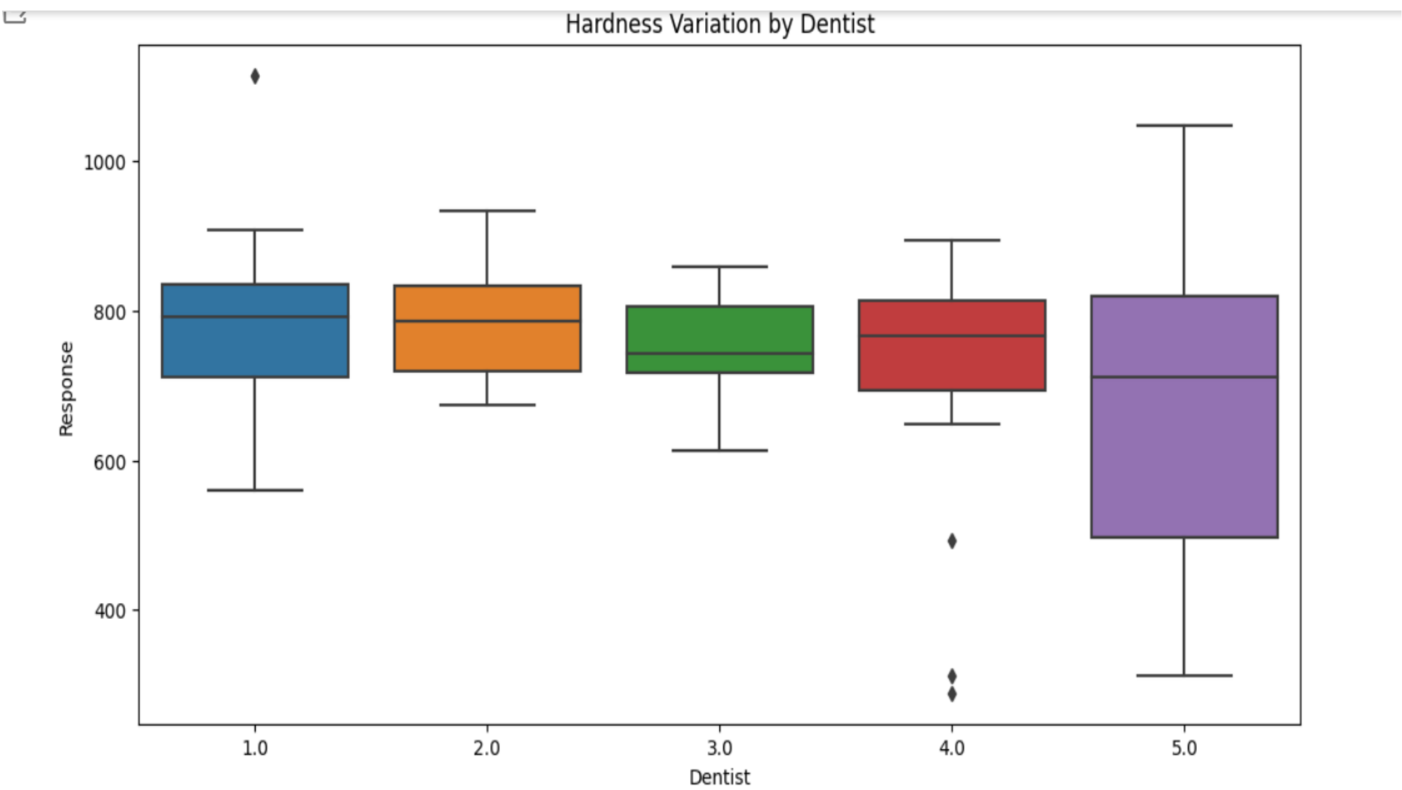
4.3 What is the interaction effect between the dentist and method on the hardness of dental implants for each type of alloy?

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

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

Ans 4.1 –

Null Hypothesis (H0): There is no difference among the dentists/methods/temperature levels in terms of implant hardness.
Alternative Hypothesis (Ha): There is a difference among the dentists/methods/temperature levels in terms of implant hardness.
We have separated Alloy1 and Alloy2 Dataset with name o DF1 and DF2 for
better understanding
H0 or DF1- Dentists on the implant hardness is same for Alloy 1
H0 or DF2- Dentists on the implant hardness is same for Alloy 2
H1 For DF1- Any two Denists the implant hardness is diferent For Alloy 1
H1 for DF2- Any two Dentists the implant hardness is diferent For Alloy 2
As per the ANOVA test, H0 for Both Alloys are accepted , p-value in both the cases are greater than significance 0.05.
Visualization :



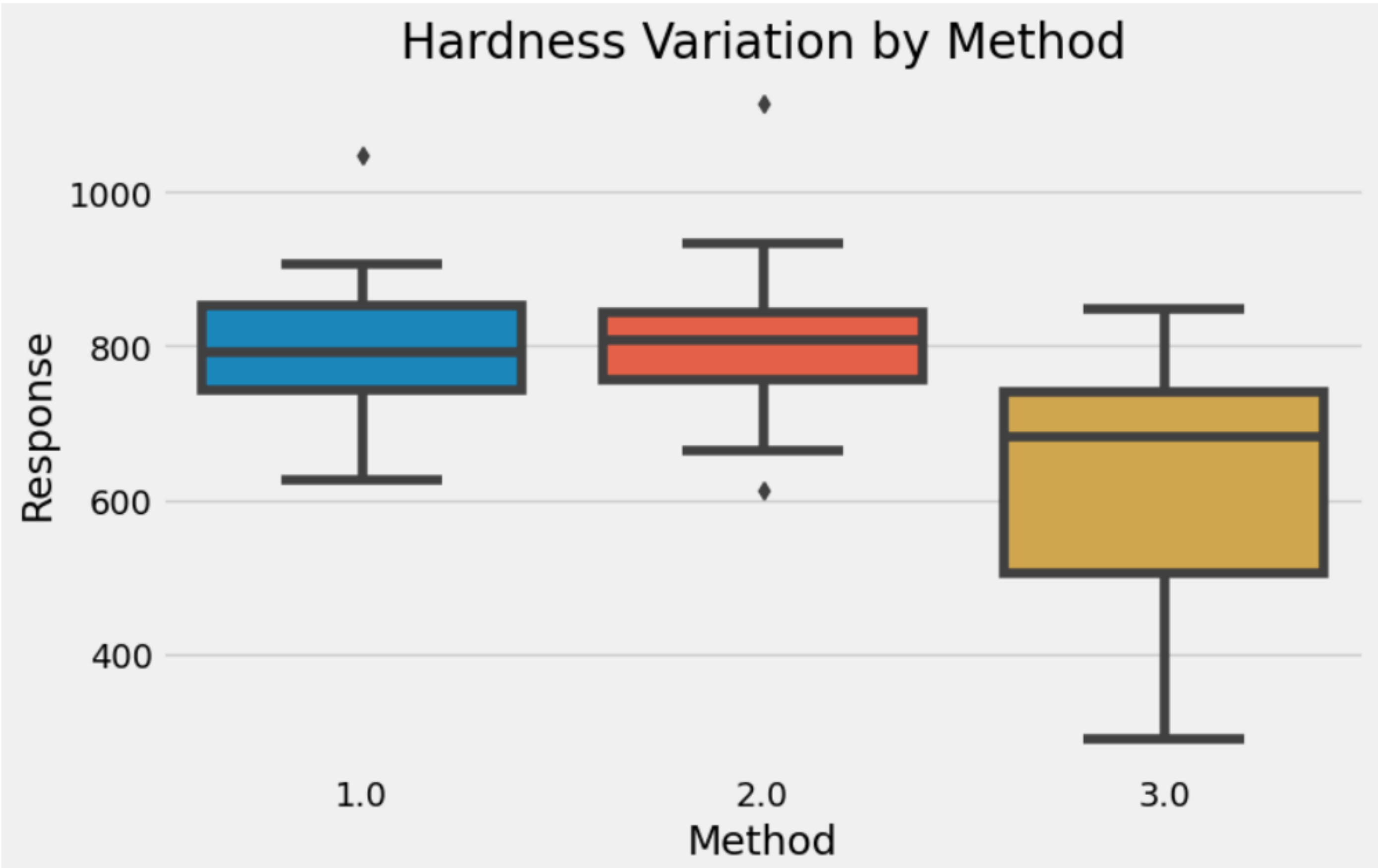
4.2 How does the hardness of implants vary depending on methods?

Ans 4.2

```
model = ols('Response ~ C(Method) * C(Alloy)', data=df_dental).fit()
aov_table=sm.stats.anova_lm(model,type=2)
(aov_table)
```

	df	sum_sq	mean_sq	F	PR(>F)
C(Method)	2.0	5.934275e+05	296713.744444	21.917576	2.189867e-08
C(Alloy)	1.0	1.058155e+05	105815.511111	7.816354	6.415387e-03
C(Method):C(Alloy)	2.0	5.468509e+04	27342.544444	2.019732	1.390848e-01
Residual	84.0	1.137167e+06	13537.707937	NaN	NaN

Since the p- value is less than 0.05 the null hypothesis is rejected and therefore the implant hardness is dependent on the method used



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

Ans 4.3:

We can use ANOVA to assess the interaction effect

The interaction effect between the dentist and method on the hardness of dental implants can be studied by examining how the combination of a specific dentist and method affects the hardness of implants for each type of alloy. This analysis can help determine if certain dentists achieve higher hardness levels with specific methods and alloys, or if there are any synergistic or antagonistic effects between dentists and methods.

```
# we can use Anova to access the interaction effect
model = ols('Response ~ Dentist * Method', data=df_dental).fit()
anova_table = sm.stats.anova_lm(model, typ=2)
print(anova_table)
```

	sum_sq	df	F	PR(>F)
Dentist	1.465472e+05	1.0	10.385174	1.795039e-03
Method	4.173336e+05	1.0	29.574648	4.970469e-07
Dentist:Method	1.136521e+05	1.0	8.054037	5.662740e-03
Residual	1.213563e+06	86.0	NaN	NaN

Q 4.4

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

Ans 4.4

To understand how the hardness of implants varies depending on dentists and methods together, we can conduct a comprehensive analysis that considers both factors simultaneously. This analysis would involve comparing the hardness of implants placed by different dentists using different methods and examining any combined effects. It can provide insights into the combined influence of dentists and methods on implant hardness.