INFERENTIAL STATISTICS

[PROJECT REPORT]

PAPPOPPULA AKHILESH

PGP-DS 2025

Table of 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?

- Based on the provided table,
- The total number of players is = 235
- The total number of injured players = 145
- Probability of an injury = Total number of players / Total number of injured players
- Therefore, the probability that a randomly chosen player will suffer an injury is approximately **61.7%.**

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

- Based on the provided table,
- The total number of players is = 235
- Number of forwards = 94
- Number of wingers = 29
- The total number of forwards and wingers is = 123

- Probability of forward or winger = Total number of forwards and wingers / Total number of players.
- Therefore, the probability that a player is either a forward or a winger is approximately **52.3%.**

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

- Based on the provided table,
- The total number of players is = 235
- The total number of Injured strikers = 45
- Probability of striker and Injury = Total number of Injured strikers / Total number of players
- Therefore, the probability that a randomly chosen player in the striker position has a foot injury is approximately **19.1%.**

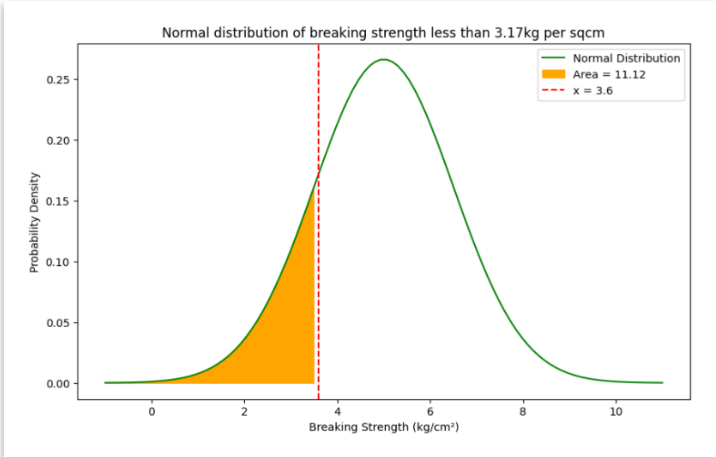
1.4 What is the probability that a randomly chosen injured player is a striker?

- Based on the provided table,
- The total number of injured players = 145
- The total number of injured strikers = 45
- Probability = The total number of injured strikers / The total number of injured players
- Therefore, the probability that a randomly chosen injured player is a striker is **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;

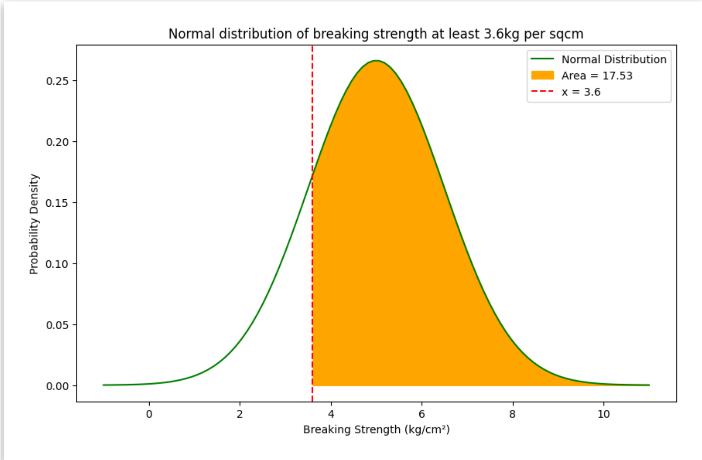
2.1 What proportion of the gunny bags have a breaking strength of less than 3.17 kg per sq cm?



(FIGURE 2.1)

- Approximately **11.2%** of gunny bags have a breaking strength of less than 3.17 kg per square centimetre.

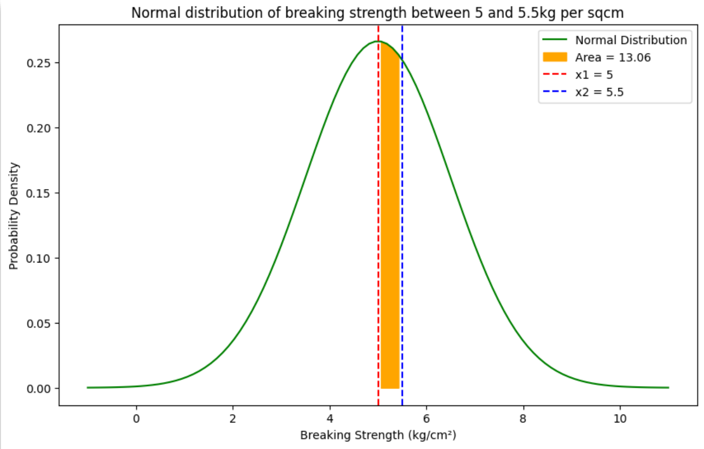
2.2 What proportion of the gunny bags have a breaking strength of at least 3.6 kg per sq cm.?



(FIGURE 2.2)

- Approximately 17.53% of gunny bags have a breaking strength of at least 3.6 kg per square centimetre.

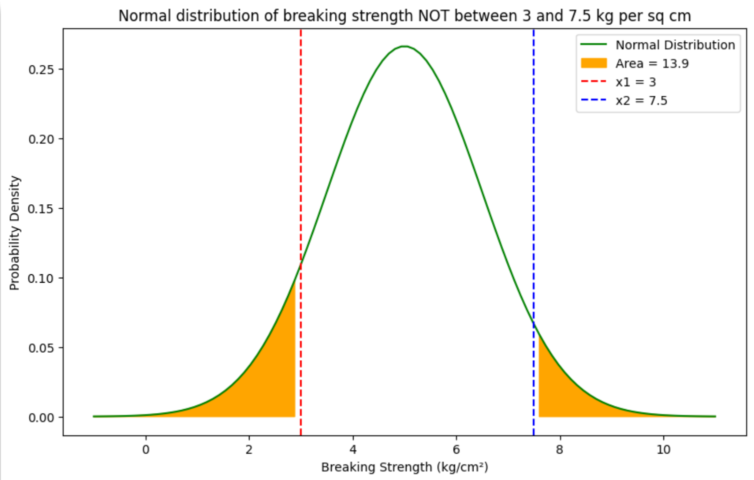
2.3 What proportion of the gunny bags have a breaking strength between 5 and 5.5 kg per sq cm.?



(FIGURE 2.3)

- Approximately 13.06% of 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.?



(FIGURE 2.4)

- Approximately 13.9% of gunny bags have a breaking strength that is not between 3 and 7.5 kg per sq cm.

PROBLEM 3:

Zingaro Stone Printing is a company that specialises 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?

Let us write the null and alternative hypothesis.

- H0: $\mu >= 150$ (The mean Brinell's hardness index of unpolished stones is greater than or equal to 150)
- Ha: $\mu < 150$ (The mean Brinell's hardness index of unpolished stones is less than 150)
- Level of significance is 0.05
- The test statistic is: -4.164629601426757
- The p-value is: 4.171286997419652e-05
- Since the p-value from the t-test is significantly lower than the significance level of 0.05, we can reject the null hypothesis.
- Therefore, it is statistically justified to conclude that unpolished stones may not be suitable for printing.

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

Let us write the null and alternative hypothesis.

- H0: $\mu 1 = \mu 2$ (The mean hardness of polished and unpolished stones is the same)
- Ha: $\mu 1 \neq \mu 2$ (The mean hardness of polished and unpolished stones is not the same)
- Level of significance is 0.05
- The test statistic is: 3.2422320501414053
- The p-value is: 0.0014655150194628353
- Since the p-value from the t-test is significantly lower than the significance level of 0.05, we can reject the null hypothesis.
- Therefore, there is sufficient evidence to suggest that the mean hardness of the polished and unpolished stones is not 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 favourite method. The response is the variable of interest.

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

Setting up the Null and Alternative hypothesis for both the Alloys

- H0 = The hardness of implants is the same across all dentists.
- Ha = The hardness of implants is different between any two dentists
- Even though the assumptions of both Shapiro and Levene's tests failed, we are proceeding with the ANOVA test.

	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

(FIGURE 4.1.1: ANOVA TABLE BASED ON ALLOY 1 DATA)

	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

(FIGURE 4.1.1: ANOVA TABLE BASED ON ALLOY 2 DATA)

- According to the ANOVA test for both alloys, the p-value is greater than 0.05. Therefore, we fail to reject the null hypothesis in both cases.

- This means we can conclude that the hardness of implants is the same across all Dentists.

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

Setting up the Null and Alternative hypothesis for both Alloys.

- H0 = The hardness of implants is the same across all the methods.
- Ha = The hardness of implants is different between any two of the methods.
- Even though the assumptions of both Shapiro and Levene's tests failed, we are proceeding with the ANOVA test.

	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

(FIGURE 4.2.1: ANOVA TABLE BASED ON ALLOY 1 DATA)

	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

(FIGURE 4.2.2: ANOVA TABLE BASED ON ALLOY 2 DATA)

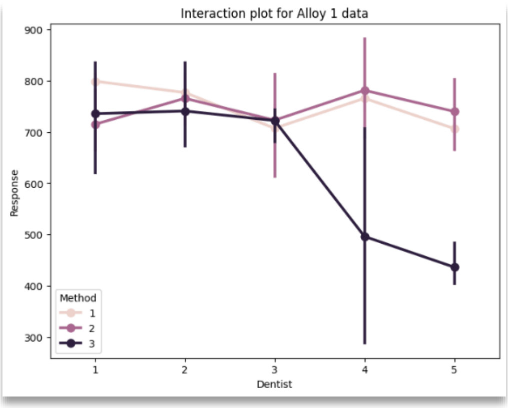
- According to the ANOVA test for both alloys, the p-value is less than 0.05. Therefore, we reject the null hypothesis.
- This means we can conclude that the hardness is different between the two methods.

group1	group2	meandiff	p-adj	lower	upper	reject
1	2	10.4333	0.9415	-64.7584	85.6251	False
1	3	-166.8	0.0	-241.9917	-91.6083	True
2	3	-177.2333	0.0	-252.4251	-102.0416	True

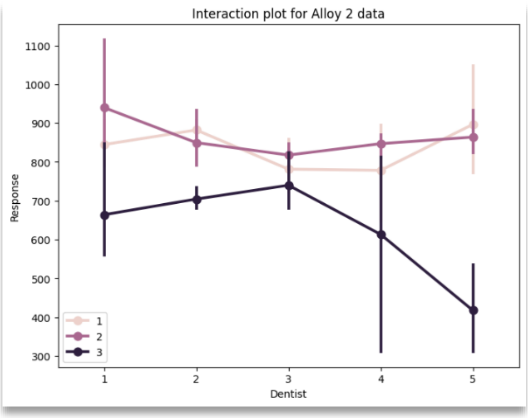
(FIGURE 4.2.3: MULTIPLE COMPARISON TEST TABLE)

- According to the table, Pairs where hardness differs significantly, i.e, (reject = True)
- Method 1 vs Method 3 and Method 2 vs Method 3
- These pairs show a statistically significant difference in implant hardness.

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



(FIGURE 4.4)



(FIGURE 4.4)

Key insights from the Interaction plots:

- Methods 1 and 2 show stable performance across dentists, with only slight variation.
- Method 3 shows a steep decline in hardness from Dentist 3 to Dentist 5, which is not observed in Methods 1 and 2.
- This suggests that Method 3's performance is highly dependent on the dentist, implying a strong interaction effect.

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

Setting up the Null and Alternative hypothesis for both Alloys.

- H0 = There is no interaction effect between the dentist and method on the hardness of implants.
- Ha = There is an interaction effect between the dentist and method on the hardness of implants.
- Even though the assumptions of both Shapiro and Levene’s tests failed, we are proceeding with the ANOVA test.

	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

(FIGURE 4.4.1: ANOVA TABLE BASED ON ALLOY 1 DATA)

	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

(FIGURE 4.4.2: ANOVA TABLE BASED ON ALLOY 2 DATA)

Key Insights:

- For all alloys, the p-value for the interaction term (C(Dentist):C(Method)) is less than 0.05. Therefore, we reject the null hypothesis.
- This means that there is a significant interaction effect between the dentist and the method on the hardness of the implants.
- The way the hardness varies depending on the method is different for different dentists.