ADVANCED STATISTICS

PROJECT REPORT

HARI HARAN

8TH JAN, 2023

Great Learning

CONTENTS	PAGE
PROBLEM 4	**11**
4.1 What is the probability that a randomly chosen student gets a grade below 85 on this exam?	11
4.2 What is the probability that a randomly selected student scores between 65 and 87?	12
4.3 What should be the passing cut-off so that 75% of the students clear the exam?	13
PROBLEM 5	**14**
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	14-15
5.2 Is the mean hardness of the polished and unpolished stones the same?	16
PROBLEM 6	**17**
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.	17-18

FIGURES AND TABLES	PAGE

<u>PROBLEM 1</u> :

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.

TABLE 1.1

	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**

Q1.1

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

Solution :

Total_Payers, P (TP) = 235

Total_Payers_not_injured, P (TPNI)= 90

So, the probability of injury to the players who are not injured is,

Probability_of_injury, P(I) = (P(TPNI) / P(TP)) * 100 = (90/235) * 100 = 38. 3 %

The probability that a randomly chosen player would suffer an injury is 38.3%.

Q1.2

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

Solution :

T_forward , P(F) = 94

T_winger, P(W) = 29

T_win_for, P(F&W) = P(F)+ P(W) = 94+29 = 123

Total_players, P(T) = 235

So, The probability of choosing a winger or a forward is,

P (F or W) = (P(F&W) / P(T))*100 = (123/235) * 100 = 52.33 %

The probability that a player is a forward or a winger is 52.33%.

Q1.3
What is the probability that a randomly chosen player plays in a striker position and has a foot injury?
Solution :
T_strikers, P(S) = 77
T_injured_strikers, P(SI) = 45

So the probability that a randomly chosen player plays in a striker position and has a foot injury,
Pb_injured_striker_selection, P(SIS) = (P(SI) / P(S))*100 = (45 /77)* 100 = 58.4 %

The probability that a randomly chosen player plays in a striker position and has a foot injury is 58.4%

Q1.4
What is the probability that a randomly chosen injured player is a striker?
Solution :
Total_injured_players, P(TP) = 145
Total_injured_strikers, P(SI) = 45
So, the probability that a randomly chosen injured player is a strike is,
pb_choosing_injured_striker,

P(CIS) = (P(SI) / P(SI))*100 = (45/145) *100
The probability that a randomly chosen injured player is a striker is 31.03%.

Q1.5
What is the probability that a randomly chosen injured player is either a forward or an attacking midfielder?
Solution :
Total_forward_injured = 56
Total_att_midfielder_injured = 24
Total_players_injured = 145

probability of choosing an injured forward,
Pb_forward_injured, P (FI) = 56/145
probability of choosing an injured attacking midfielder,

Pb_att_midfielder_injured, P(AMI) = 24/145

Pb_fw_md_inj_add = P (FI) + P(AMI)
Pb_fw_md_inj = P (FI) * P(AMI)

So, probability of choosing either forward or attacking midfielder,

Pb_fwd_or_mid = ((P (FI) + P(AMI)) - (P (FI) * P(AMI))) * 100
= ((56/145 + 24 /145) - (56/145 * 24/145)) *100
= 48.78 %

The probability that a randomly chosen injured player is either a forward or an attacking midfielder is 48.78% .

<u>PROBLEM 2</u> :

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%.

Q2.1
What are the probabilities of a fire, a mechanical failure, and a human error respectively?
Solution:
From the above Data,
Probability of a radiation leak in case of fire,
Pb_F = 20/100
Probability of a radiation leak in case of mechanical failure,
Pb_M = 50/100
Probability of a radiation leak in case of human error,
Pb_HE = 10/100
Pb_Total = (Pb_F + Pb_M + Pb_HE)
= 20/100 + 50/100 +10/100
= 0.79
The total probability of a radiation leak due to fire, mechanical failure and human error is 79 %.

Probability of a fire,
Pb_Fire = (Pb_F/Pb_Total) *100 = (20/100 / 0.79)* 100 = 25.0 %
The probability of a fire is 25.0%.

Probability of a Mechanical failure,
Pb_Mech = (Pb_M/Pb_Total) *100 = (50/100 /0.79) * 100 = 62.5 %
The probability of a mechanical failure is 62.5%.

Probability of a human error,
Pb_Human_Error = (Pb_HE/Pb_Total) *100 = (10/100 / 0.79)*100 = 12.5 %
The probability of a human error is 12.5%.

Q2.2

What is the probability of a radiation leak?

Solution:

Probability of a radiation leak occurring simultaneously with a fire is 0.1%, so Pb_F = 0.1

Probability of a radiation leak occurring simultaneously with a mechanical failure is 0.15%, so Pb_M = 0.15

Probability of a radiation leak occurring simultaneously with a human error is 0.12%, so Pb_HE = 0.12

Probability of a radiation leak

Pb_L = (Pb_F + Pb_M + Pb_HE) = 0.1 + 0.15 + 0.12 = 0.37

The probability of a radiation leak is 37.0 %.

Q2.3

Suppose there has been a radiation leak in the reactor for which the definite cause is not known. What is the probability that it has been caused by:

- **A Fire.**

Solution :

The probability of radiation leak due to fire,

Pb_Fire_leak = Pb_F_L/Pb_L = (0.1/0.37) * 100 = 0.2703 x100 = 27.027 %

The probability of a radiation leak due to fire is 27. 03%

- **A Mechanical Failure**.

Solution :

The probability of radiation leak due to mechanical failure,

Pb_Mech_leak = Pb_M_L/Pb_L = (0.15/0.37) * 100 = 0.4054 x100 = 40.54%

The probability of a radiation leak due to fire is 40.54%.

- **A Human Error.**

Solution :

The probability of radiation leak due to human error,

Pb_Human_Error_leak = Pb_F/Pb_L = (0.12/0.37) * 100 = 0.3243 x100 = 32.43 %

The probability of a radiation leak due to fire is 32. 43%

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)

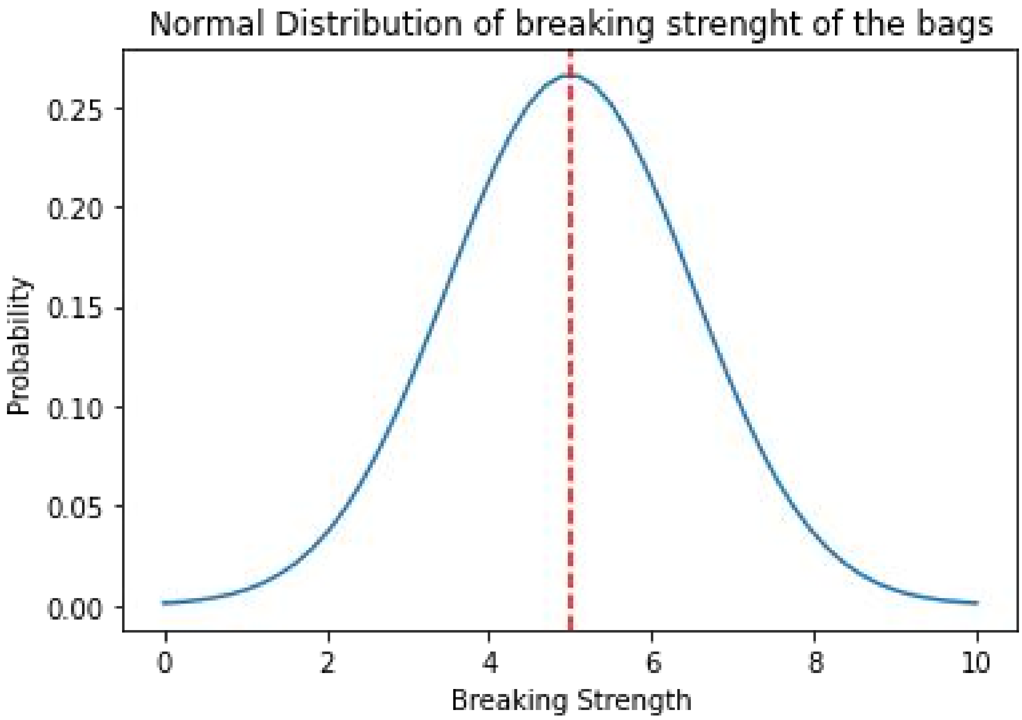

FIGURE 3.1 - NORMAL DISTRIBUTION FOR BREAKING STRENGHT

From the above data and figure Red-Dotted Lines indicates the mean of the population,

Mean, μ (mu) = 5

Standard deviation, σ (sigma) = 1.5

Q3.1

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

Solution :

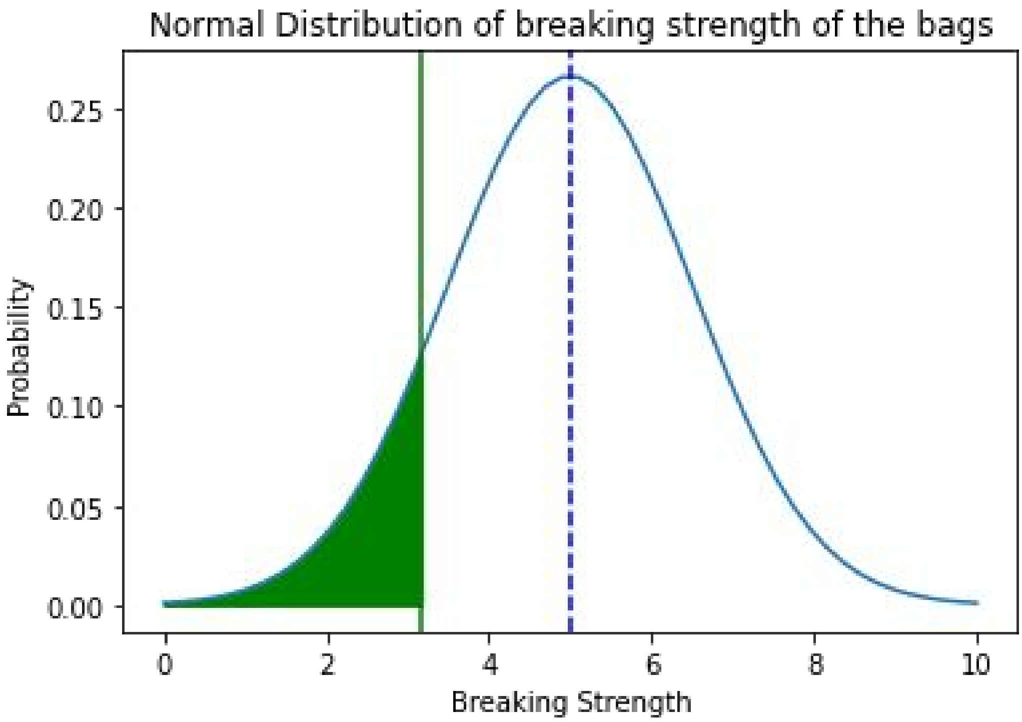

FIGURE 3.1 - BEAKING STRENGTH less than 3.17 Kg/ Cm^2

Using the cumulative probability method, we find that the he proportion of the gunny bags have a breaking strength less than 3.17 kg per sq cm is **0.1112.**

In the above figure, the ***green highlighted area*** under the curve shows the proportion of gunny bags that have breaking strength less than 3.17 Kg/ Cm^2 .

The proportion of the gunny bags have a breaking strength less than 3.17 Kg/ Cm^2 is 11.12%.

Q3.2

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

Solution :

Using the cumulative probability method, we find that the proportion of the gunny bags have a breaking strength of at least 3.6 Kg/Cm^2 is 0.1753

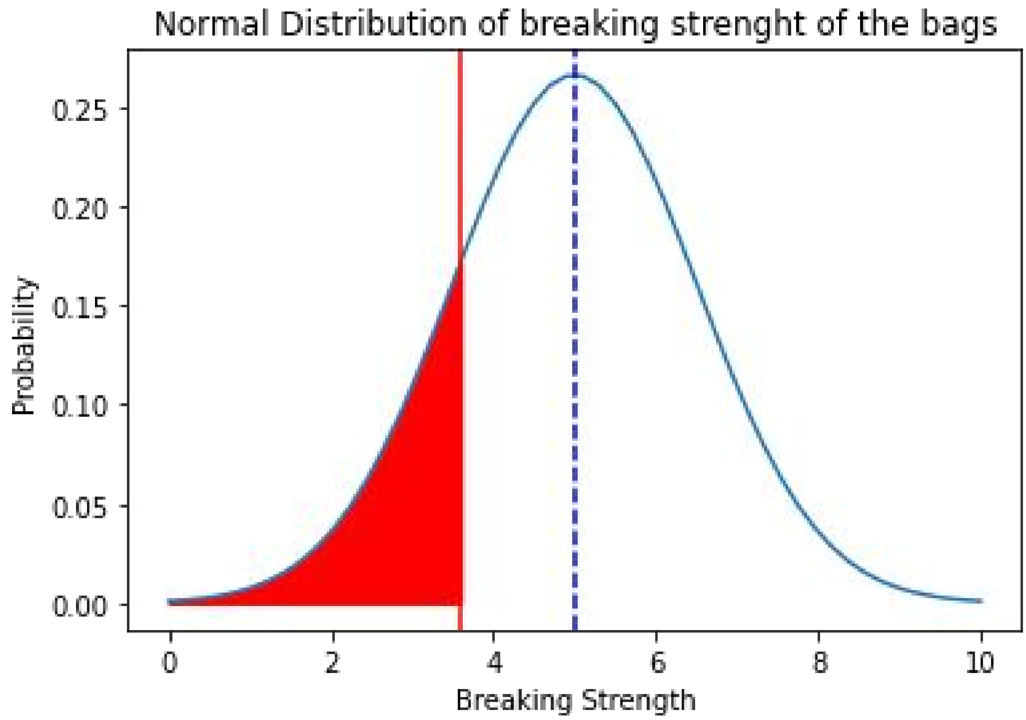

FIGURE 3.2 - BEAKING STRENGTH at least 3.6 Kg/ Cm^2

In the above figure, the ***red highlighted area*** under the curve shows the proportion of gunny bags that have breaking strength at least 3.6 Kg/ Cm^2 .

The proportion of the gunny bags have a breaking strength less than 3.6 Kg/ Cm^2 is 17.53%.

Q3.3

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

Solution :

We will apply Z-score test,

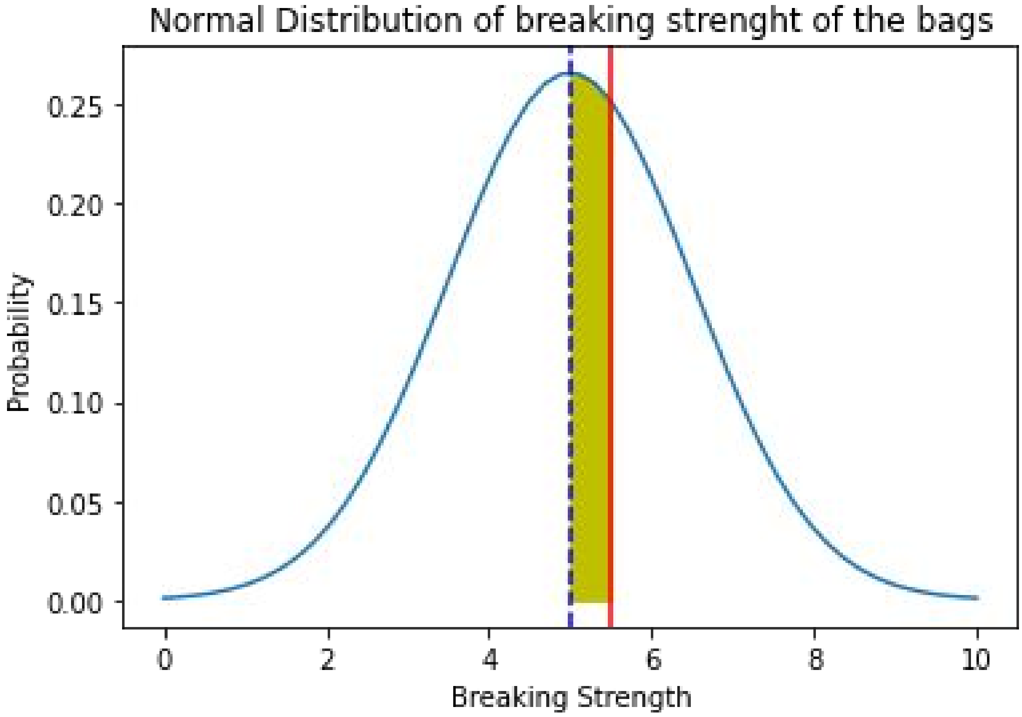

FIGURE 3.3 - BEAKING STRENGTH between 5 Kg/ Cm^2 and 5.5 Kg/ Cm^2

Using Z- score test method, we find that the proportion of the gunny bags have a

breaking strength between 5 Kg/ Cm^2 and 5.5 Kg/ Cm^2 is 1.139

The ***yellow highlighted*** are under the curve shows that only 11.39 % of the bad t hat have breaking strength between 5 Kg/ Cm^2 and 5.5 Kg/ Cm^2..

So, from the above figure we can say that the we find that the proportion of th e gunny bags have a breaking strength between 5 Kg/ Cm^2 and 5.5 Kg/ Cm^2 is on ly 13.9%.

Q3.4

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

Solution :

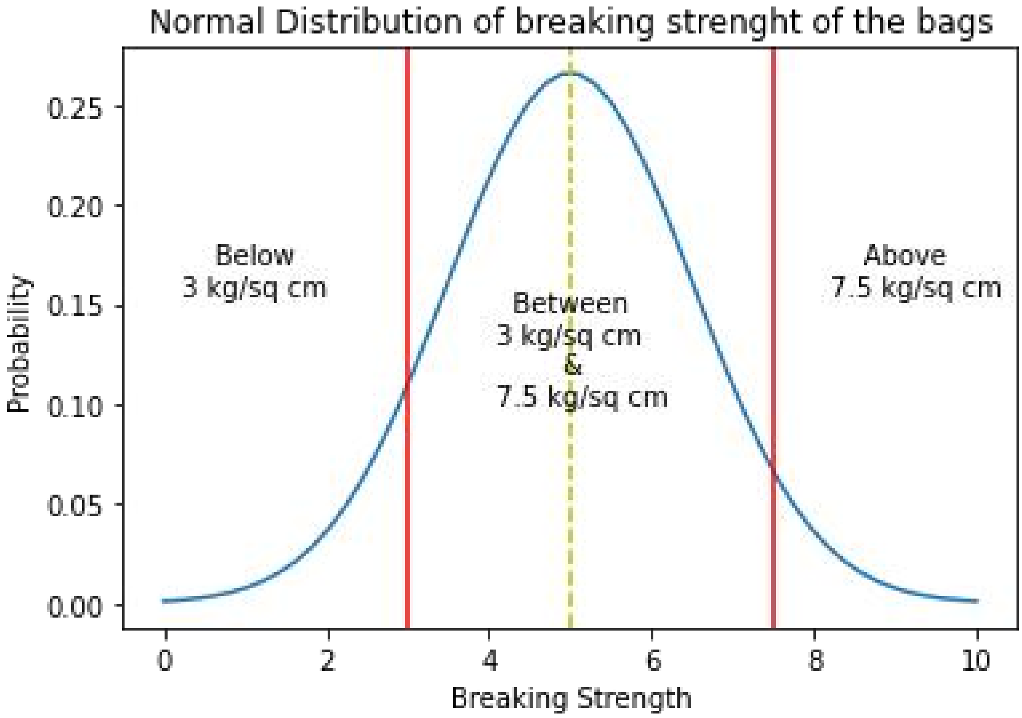

FIGURE 3.3 - BEAKING STRENGTH not between 3 Kg/ Cm^2 and 7.5 Kg/ Cm^2

The area on the left side and right side of ***the red line*** for 3 Kg/ Cm^2 and 7.5 Kg/ Cm^2, respectively, on both sides of the figure shows the amount of gunny bags that has breaking strength which does not lie between 3 Kg/ Cm^2 and 7.5 Kg/ Cm^2

The above figure shows that the we find that the proportion of the gunny bags have a breaking strength between 3 Kg/ Cm^2 and 7.5 Kg/ Cm^2 is only 13.9 %.

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.

Q4.1.
What is the probability that a randomly chosen student gets a grade below 85 on this exam?
Solution :
From the above data,

Population Mean, μ (mu) = 77

Standard deviation, σ (sigma) = 8.5

Let, X = 85

Using cumulative probability, applying cdf method we get that

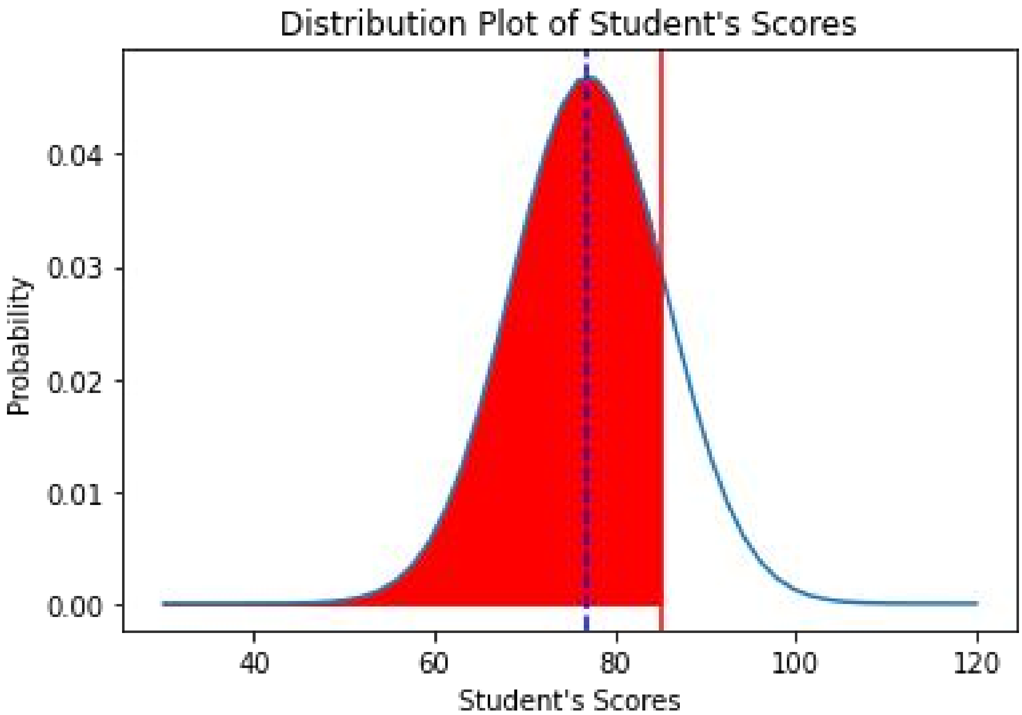

FIGURE 4.1 - Student's Score Below 85

The the probability that a randomly chosen student gets a grade below 85 on this exam is around 82.67 %.

Q4.2.

What is the probability that a randomly selected student scores between 65 and 87?

Solution :

From the above data,

Population Mean, μ (mu) = 77

Standard deviation, σ (sigma) = 8.5

Let,

X1 = 65

X2 = 87

Using the probability, applying cdf method we get that,

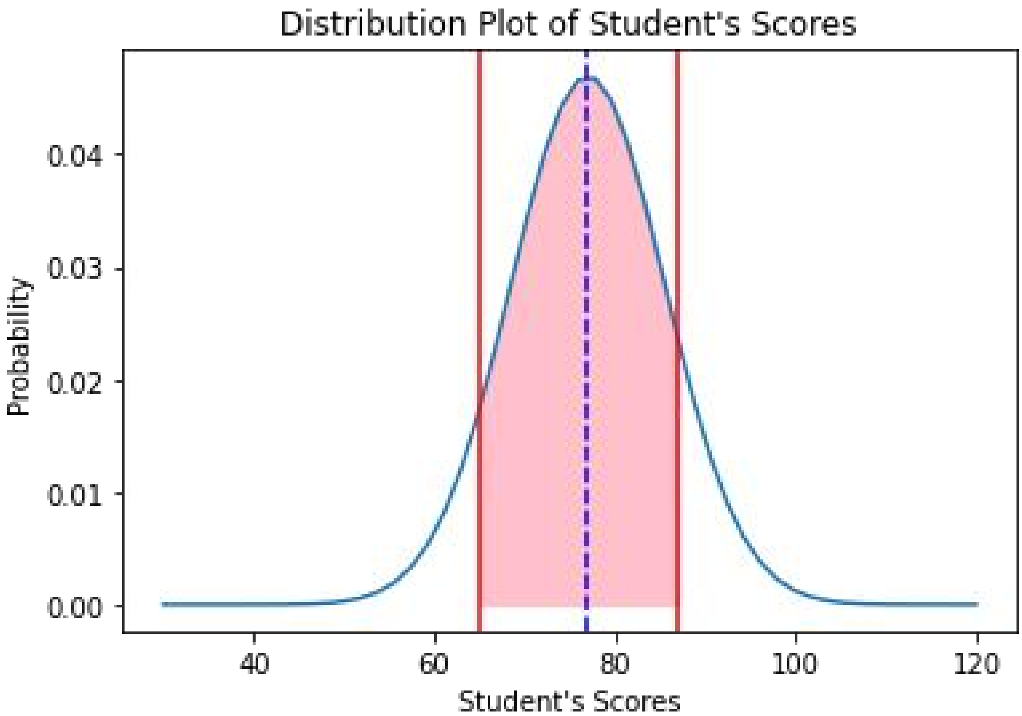

FIGURE 4.2 - Student's Score Between 65 and 87

The probability that a randomly selected student scores between 65 and 87 is around 4.07%.

Q4.3.

What should be the passing cut-off so that 75% of the students clear the exam?

Solution :

From the above data,

Population Mean, $\boldsymbol{\mu}$ (mu) = 77

Standard deviation, $\boldsymbol{\sigma}$ (sigma) = 8.5

Using the probability, applying PPF method we get that,

Passing cut-off = 71

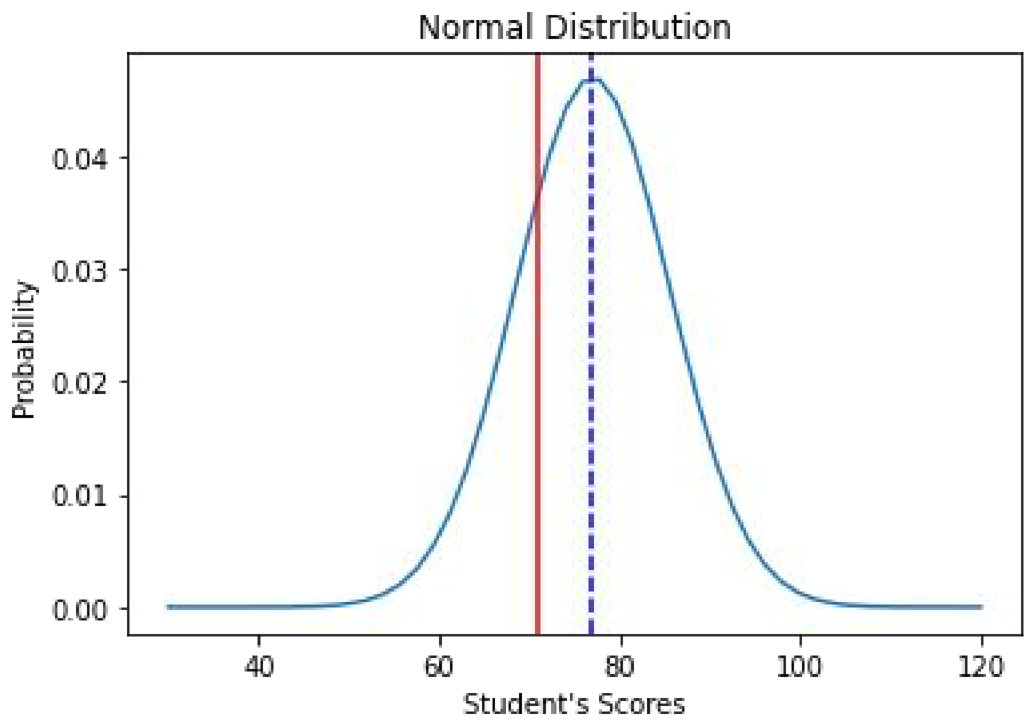

FIGURE 4.3 - Student's cut-off mark at 71

The passing cut-off should be 71 so that 75% of the students clear the exam.

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).

Checking the data for any irregularities :

	Unpolished	Treated and Polished
0	164.481713	133.209393
1	154.307045	138.482771
2	129.861048	159.665201
3	159.096184	145.663528
4	135.256748	136.789227

```
<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
```

- Data Check - Data Head() and Info ()function after loading the CSV file.
- Data in the columns is of float type (decimal values).
- There are no missing values.
- There are no duplicate values.

Q5.1

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 :

- The **level of significance** (Alpha), α = 0.05
- Checking the Unpolished stone is not suitable for printing or not, with two sample T-test

Hypothesis Assumption:

For T Test:

- Null hypothesis states that the Unpolished stone is suitable for printing.
- Alternative hypothesis states that the Unpolished stone is not suitable for printing.

We have to obtain the p-value using two sample T-test ,
We get,
t value = -3.5589113215869057
&
P-value = 0.0006545976110249849

T statistic is -3.5589113215869057
Corresponding P-value is 0.00032729880551249244

The two-sample t-test p-value= 0.0006545976110249849
We have to reject the null hypothesis in favour of alternative hypothesis.

Unpolished stone is suitable for printing:

- ***Since, the p-value is 0.00065 which is way less than the level of significance (Alpha), α = 0.05. Hence we can reject the null hypothesis that Unpolished stones is suitable for printing.***

- ***Therefore, statistically, the Unpolished stones are not suitable for printing.***

Q5.2

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

Solution:

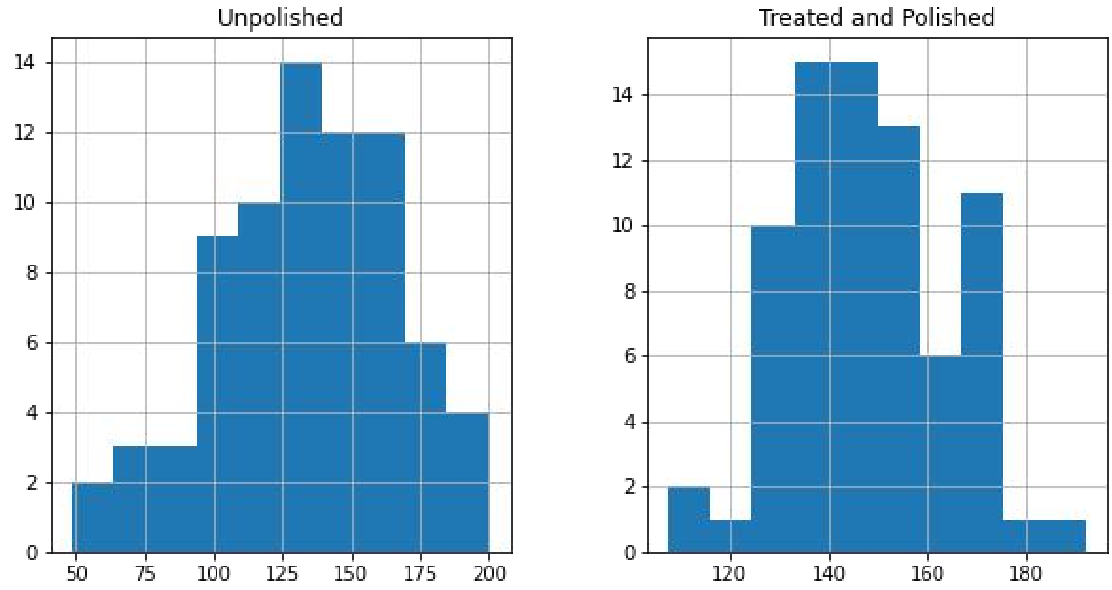

FIGURE 5.2.1 - HISTOGRAM PLOT FOR UNPOLISHED AND POLISHED STONES

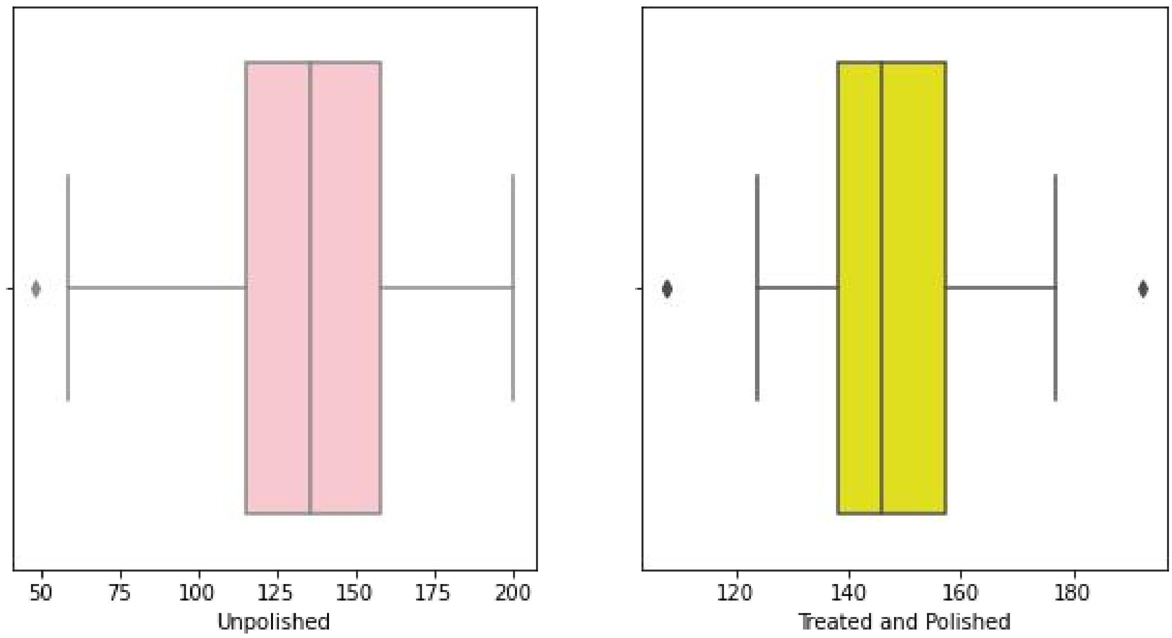

FIGURE 5.2.2- BOX PLOT FOR UNPOLISHED AND POLISHED STONES.

From the graph and statistically, the mean value of the hardness of both the stones are not equal.

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 :

Testing whether the new program in the gym is successful or not.

Hypothesis assumption:

- For null hypothesis, the candidate is able not to do more than 5 push-ups even after entering in the program, $\mu_{2aqua} \geq \mu_{1aqua}$.
- For alternative hypothesis, the candidate is able to do more than 5 push-ups after entering to the program, $\mu_{2aqua} < \mu_{1aqua}$.

$$H_0 : \mu_{1aqua} - \mu_{2aqua} \geq 0$$
$$H_A : \mu_{1aqua} - \mu_{2aqua} < 0$$

Here, μ_{1aqua} denotes the mean of push-ups before joining the program and μ_{2aqua} denotes the mean of push-ups after joining the program.

Here we have **Level of Significance**, ***α = 0.05*** as given in the problem statement.

By applying paired T-test method, we get the following:

T-statistics = -19.323

Paired two-sample t-test p-value = 1.1460209626255983e-35

Since, (p-value/2) < alpha.

We can conclude the following:

- We have enough evidence to reject the null hypothesis in favour of alternative hypothesis
- The candidate is able not to do more than 5 push-ups even after entering in the program