Advanced Statistic Project

Submitted by: Taniya Dubey

INDEX

No	Problem
1	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?
	1.5 What is the probability that a randomly chosen injured player is either a forward or an attacking midfielder?
2	2.1 What are the probabilities of a fire, a mechanical failure, and a human error respectively?
	2.2 What is the probability of a radiation leak?
	2.3 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. • A Mechanical Failure. • A Human Error.
3	3.1 What proportion of the gunny bags have a breaking strength less than 3.17 kg per sq cm?
	3.2 What proportion of the gunny bags have a breaking strength at least 3.6 kg per sq cm.?
	3.3 What proportion of the gunny bags have a breaking strength between 5 and 5.5 kg per sq cm.?
	3.4 What proportion of the gunny bags have a breaking strength NOT between 3 and 7.5 kg per sq cm.?
4	4.1 What is the probability that a randomly chosen student gets a grade below 85 on this exam?
	4.2 What is the probability that a randomly selected student score between 65 and 87?
	4.3 What should be the passing cut-off so that 75% of the students clear the exam?
5	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
	5.2 Is the mean hardness of the polished and unpolished stones the same?
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?

List of plots

List of tables

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

Table 1: table of football team

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

As we know the formula of probability of any event is:

Probability (Event) = Favorable Outcomes/Total Outcomes

Event: In probability, the set of outcomes from an experiment.

Total Outcomes: the total number of possible outcomes for the probability experiment.

Favorable Outcomes: those outcomes that satisfy the condition of an event.

According to the question:

event will be the chosen player who suffer injury.

total outcomes will be the total number of players.

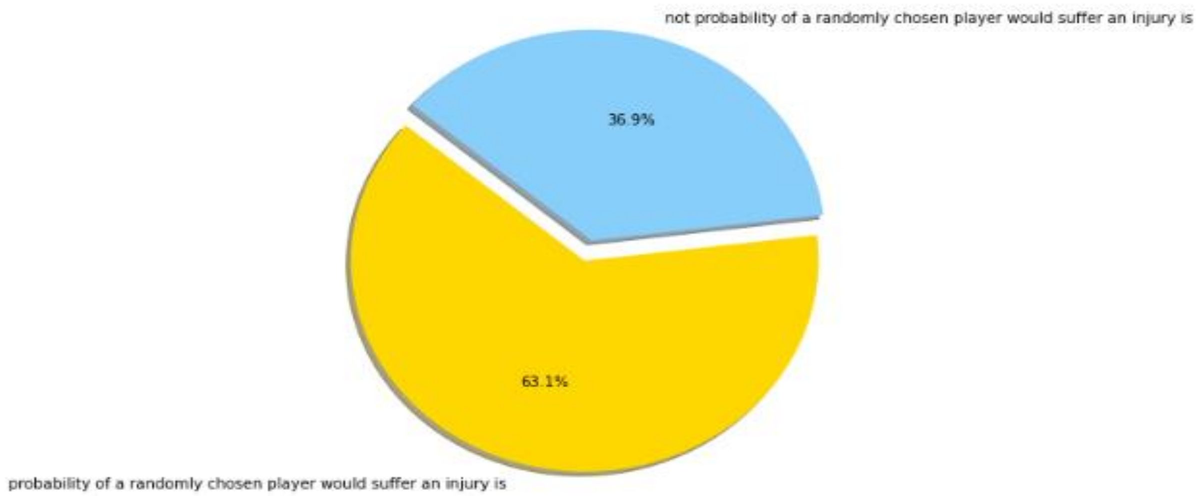
Favorable outcomes will be the numbers of injured players.

The probability that a randomly chosen player would suffer an injury = $\frac{\text{Total number of injured players}}{\text{Total number of players}}$

$$= \frac{145}{235}$$

$$= 0.6170$$

It shows that there is 61% chance that a chosen player would suffer an injury.



Plot 1: probability of player suffer injury

```
probability that a randomly chosen player would suffer an injury is 0.6170
```

Image 2

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

Same probability formula will be used here to solve this question.

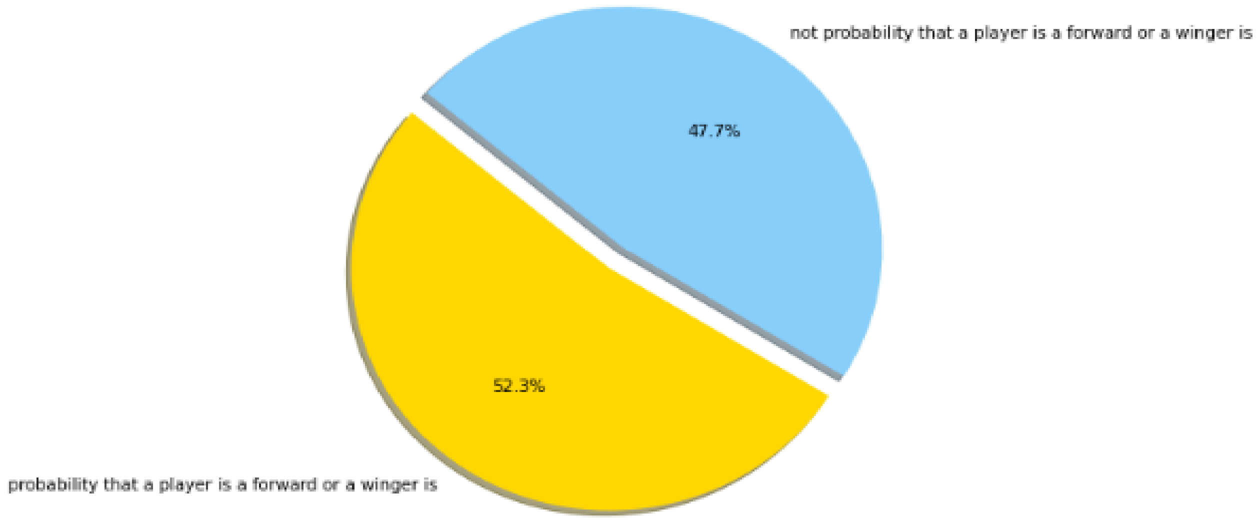
The probability that a player is a forward or a winger

$= \frac{\text{Total players of forward \& winger}}{\text{Total number of players}}$

$= \frac{123}{235}$

$= 0.5234$

It shows that there is a 52% chance that a chosen player is a forward or a winger.



Plot 2: probability of player is a forward or a winger

```
The probability that a player is a forward or a winger is 0.5234
```

Image 3

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

Same probability formula will be used here to solve this question.

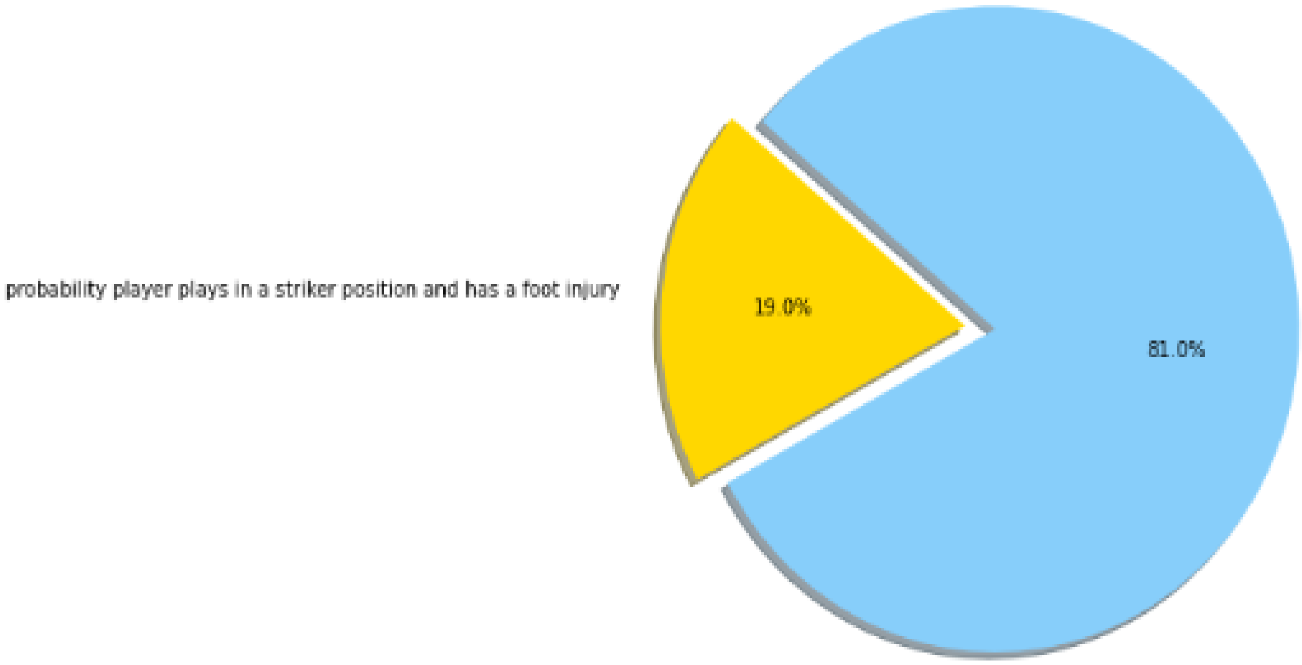
The probability that a randomly chosen player plays in a striker position and has a foot injury

$= \frac{\text{Total Injured striker}}{\text{Total numbers of player}}$

$= \frac{45}{235}$

$= 0.191$

It shows that there is a 19% chance that a randomly chosen player is a striker who has foot injury.



Plot3: probability of player is a striker has an injury

```
probability that a randomly chosen player plays in a striker position and has a foot injury is 0.1914
```

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

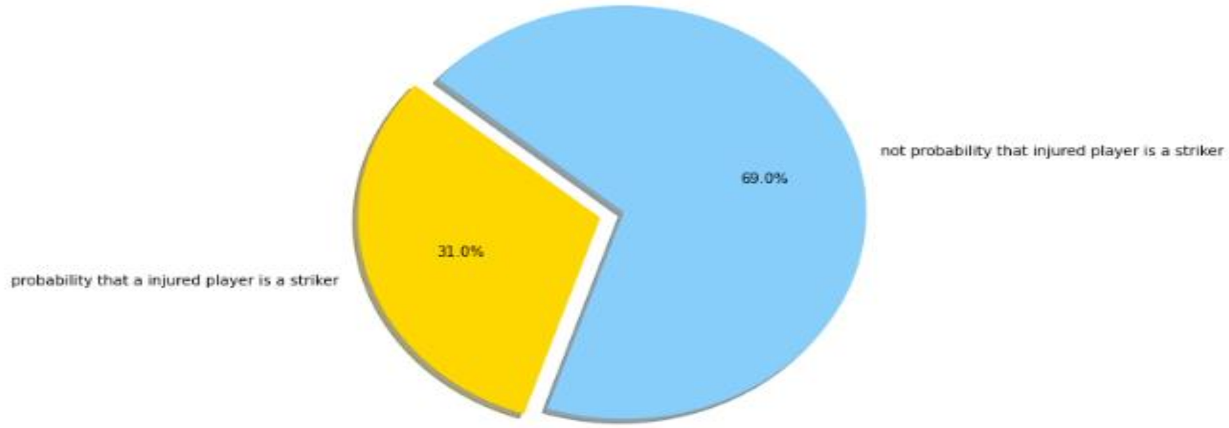
The probability that a randomly chosen injured player is a striker=

$$\frac{\text{Total injured players who are striker}}{\text{Total numbers of players injured}}$$

$$= \frac{45}{145}$$

$$= 0.31$$

It shows that there is 31% chance that the randomly chosen player is an injured striker.



Plot4: probability of an injured player is a striker

```
probability that a randomly chosen injured player is a striker is 0.310
```

Image 5

1.5 What is the probability that a randomly chosen injured player is either a forward or an attacking midfielder?

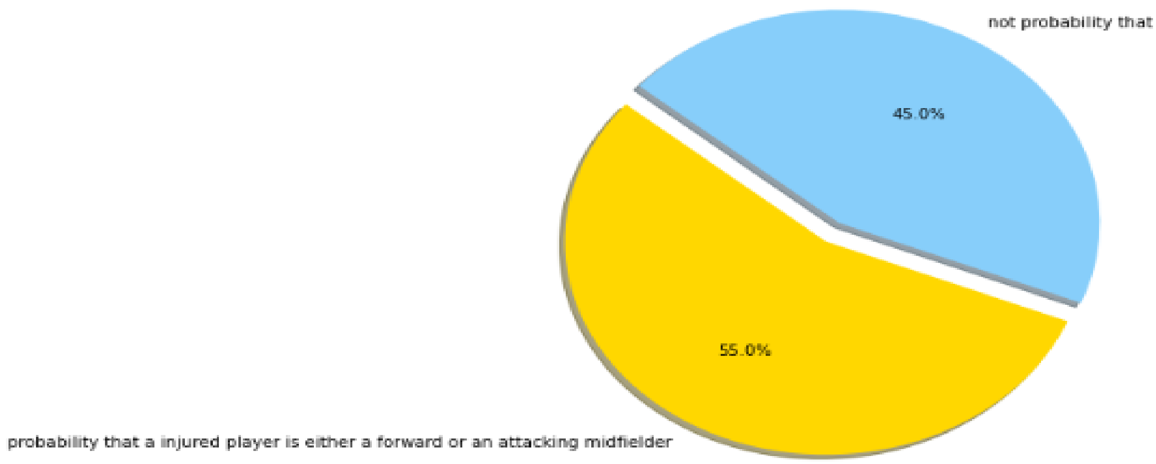
The probability that a randomly chosen injured player is either a forward or an attacking midfielder

$$= \frac{\text{Total number of injured forward or attacking}}{\text{Total number of players injured}}$$

$$= \frac{80}{145}$$

= 0.55%

It shows that there is a 55% chance that the randomly chosen injured player is either a forward or an attacking midfielder.



Plot5: probability of an injured player is a forward or a midfielder

```
The probability that a randomly chosen injured player is either a forward or an
attacking midfielder 0.5517241379310345
```

Image 6

Problem 2

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%.
Based on the given information answer the questions below.

2.1 What are the probabilities of a fire, a mechanical failure, and a human error respectively?

P (Fire) = P (Radiation Leak AND Fire)/ P (Radiation Leak | Fire) = 0.001/0.2 = 0.005 According to the information available, probability of a mechanical failure is; P (Mechanical Failure) = P (Radiation Leak AND Mechanical Failure)/ P (Radiation Leak |Mechanical Failure) = 0.0015/0.5 = 0.003 According to the information available the probability of a human error is; P (Human Error) = P (Radiation Leak

AND Human Error)/ P (Radiation Leak | Human Error) = 0.0012/0.1 = 0.012

2.2 What is the probability of a radiation leak?

P (Radiation Leak) = P (Radiation Leak AND Fire) + P (Radiation Leak AND Mechanical failure) + P (Radiation Leak AND Human error) = 0.001 + 0.0015 + 0.0012 = 0.0037

2.3 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.
A Mechanical Failure.
A Human Error.

This question will be solved with the help of bayes theorem

Bayes' theorem, named after Thomas Bayes, describes the probability of an event, based on prior knowledge of conditions that might be related to the event.

Formula

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

A, B = events
$P(A|B)$ = probability of A given B is true
$P(B|A)$ = probability of B given A is true
$P(A), P(B)$ = the independent probabilities of A and B

Image 7

P (Fire|Radiation Leak) = P (Fire and Radiation Leak)/P (Radiation Leak) = 0.001/0.0037 = 0.2703.

P (Mechanical Failure | Radiation Leak) = P (Mechanical Failure and Radiation Leak)/P (Radiation Leak) = 0.0015/0.0037 = 0.4054.

P (Human Error | Radiation Leak) = P (Human Error and Radiation Leak)/P (Radiation Leak = 0.0012/0.0037 = 0.3243.

Problem 3

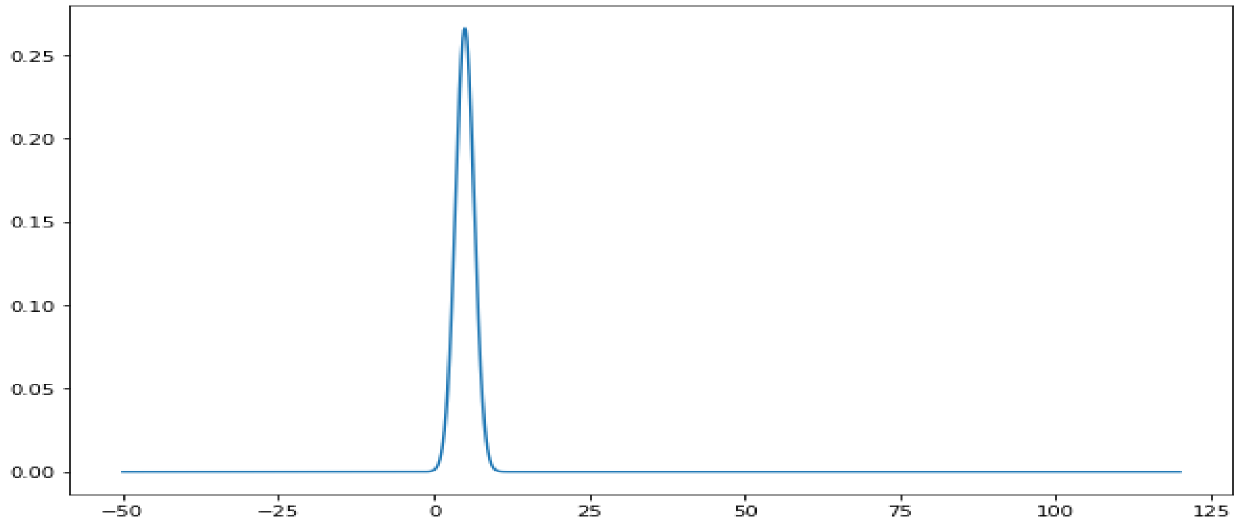
The breaking strength of gunny bags used for packaging cement is normally distributed with a mean of 5 kg per sq. centimeter and a standard deviation of 1.5 kg per sq. centimeter. The quality team of the cement company wants to know the following about the packaging material to better understand wastage or pilferage within the supply chain; Answer the questions below based on the given information

Mean (mu)= 5 kg per sq
Standard Deviation(sigma)= 1.5 kg per sq

To solve the following question the concept of Normal Cumulative Density Function (CDF)is used. A cumulative distribution function (CDF) tells us the probability that a random variable takes on a value less than or equal to some value. It can be represented mathematically as -

$F_X(x)=P(X=X)$

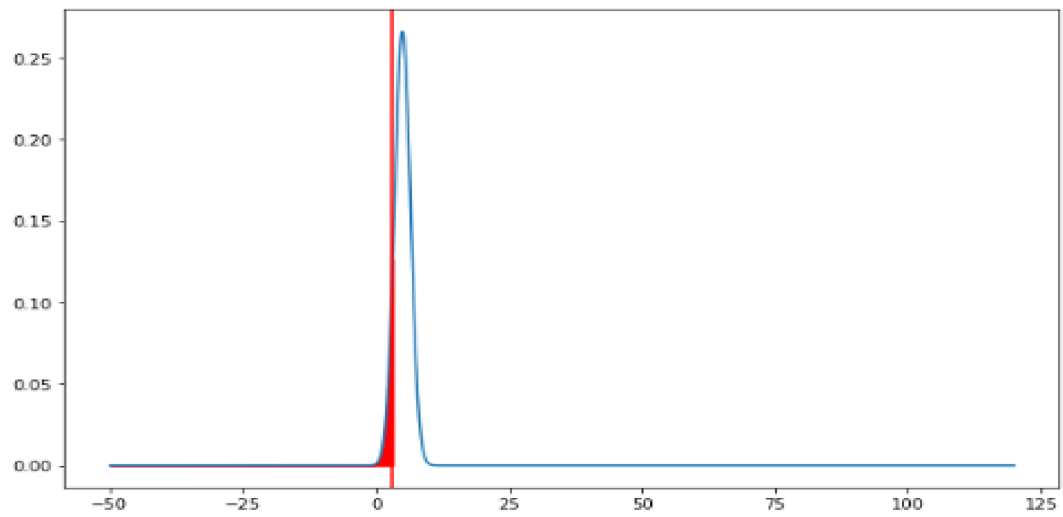
**hear we assume that the distribution is normal



Plot 6: normal distribution of gunny bags

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

With the help of formula, we will get 0.11123 probability which means that there is a 11% chance of a gunny bags have a breaking strength less that 3.17kg per sq cm.

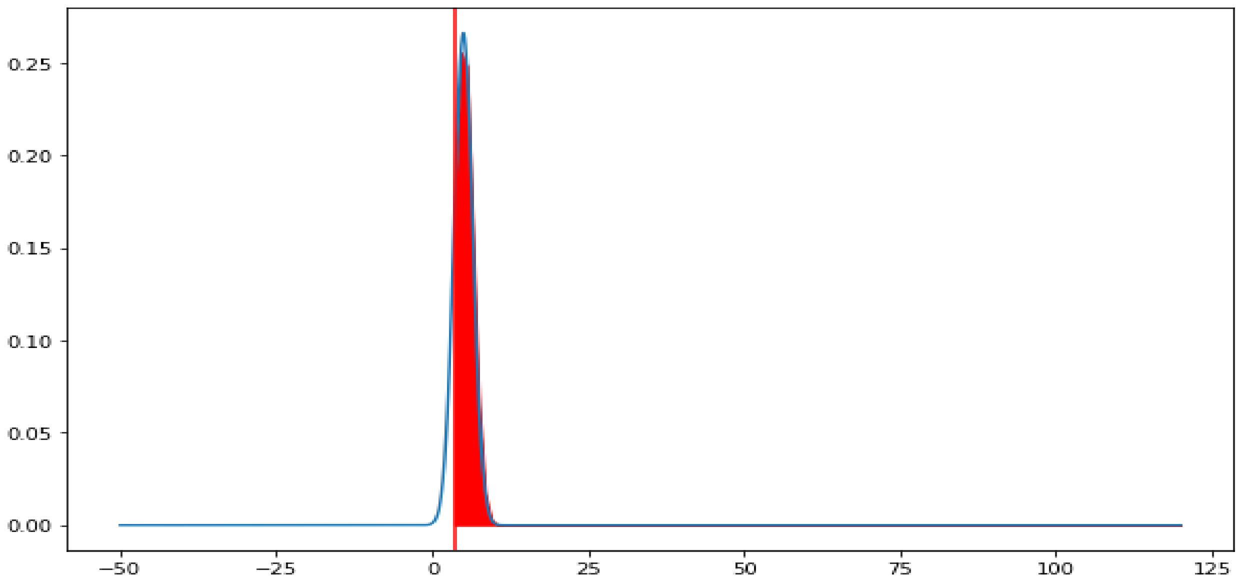


Plot:7 proportion of gunny bags breaking strength less than 3.17kg per sq cm

```
probability that the gunny bags have a breaking strength less than 3.17 kg per sq cm 0.1112
```

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

With the help of formula, we get 0.173 probability which means that there is a 82% chance of the gunny bags that have a breaking strength at least 3.6 kg per sq cm.

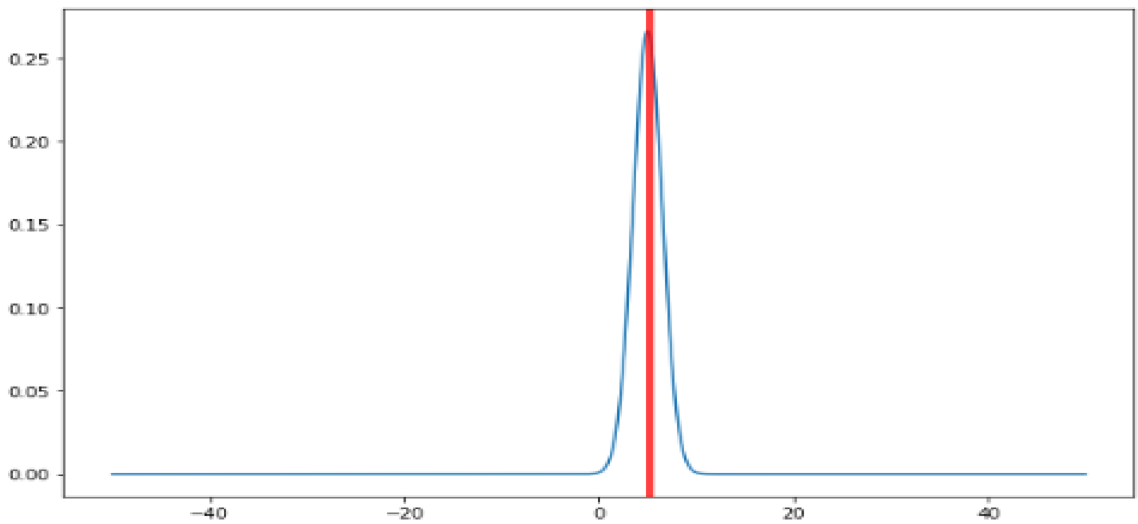


Plot 8: proportion of the gunny bags having a breaking strength at least 3.6kg per sq cm.

```
The probability that the gunny bags have a breaking strength at least 3.6 kg per sq cm
0.82
```

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

The proportion of the gunny bags that have a breaking strength between 5 and 5.5 kg per sq cm is 0.1305. which means that there is a 13% chance.

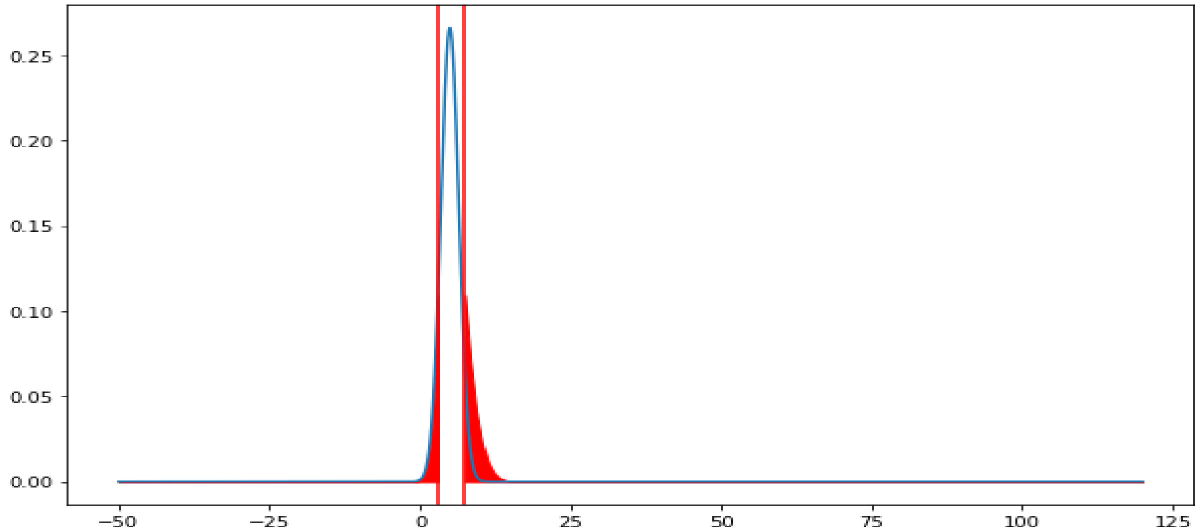


Plot 9: proportion of the gunny bags have a breaking strength between 5 and 5.5 kg per sq cm

```
probability that the gunny bags have a breaking strength between 5 and 5.5 kg per sq cm. 0.1305
```

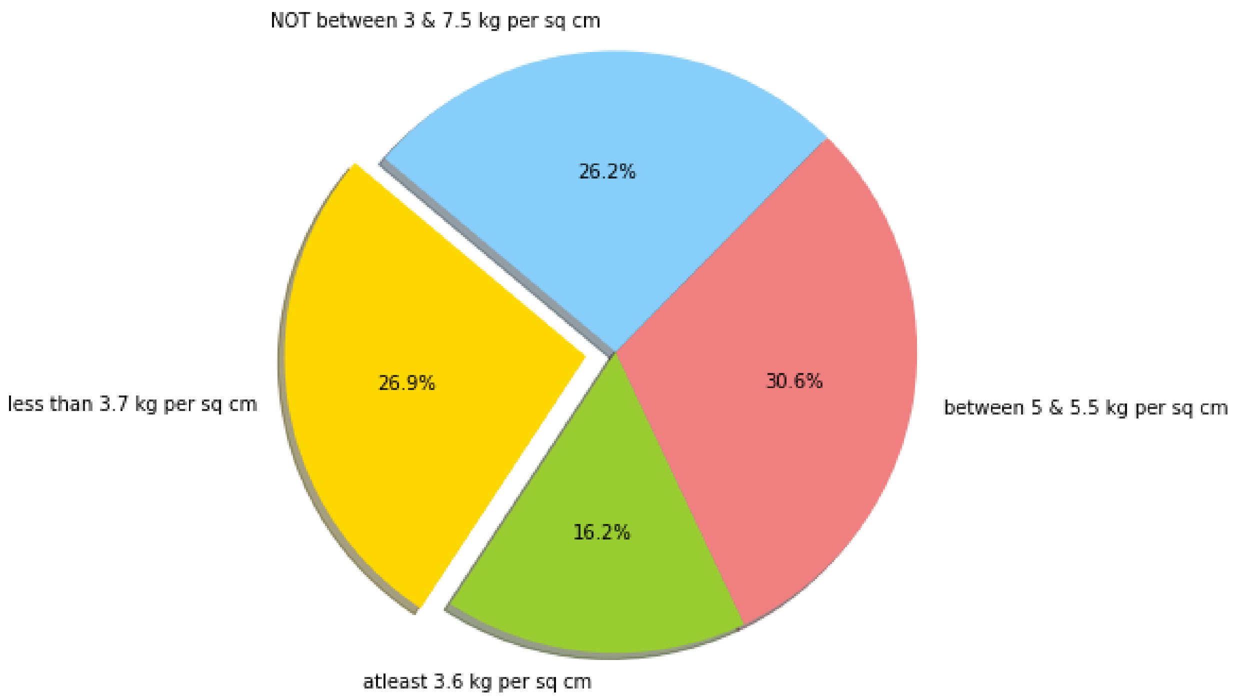
3.4 What proportion of the gunny bags have a breaking strength NOT between 3 and 7.5 kg per sq cm?

The chance of the gunny bags has a breaking strength NOT between 3 and 7.5 kg per sq cm is 13%.



Plot 10: proportion of gunny bags have a breaking strength not between 3 and 7.5 kg per sq cm.

```
probability that the gunny bags have a breaking strength NOT between 3 and 7.5 kg per sq cm. 0.139
```



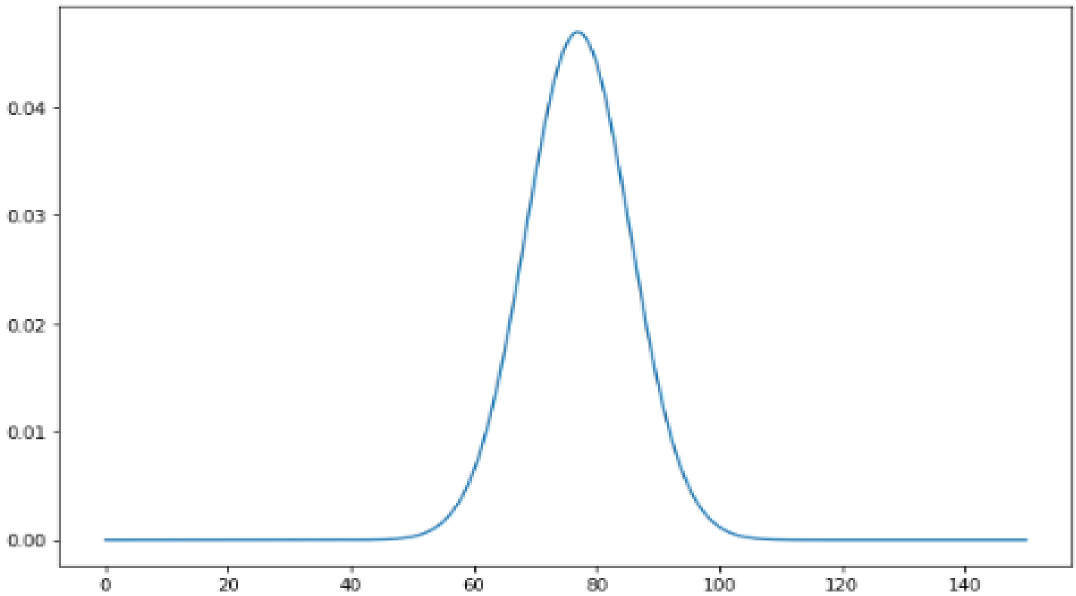
plot 11: final plot

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.

mean(mu) = 77

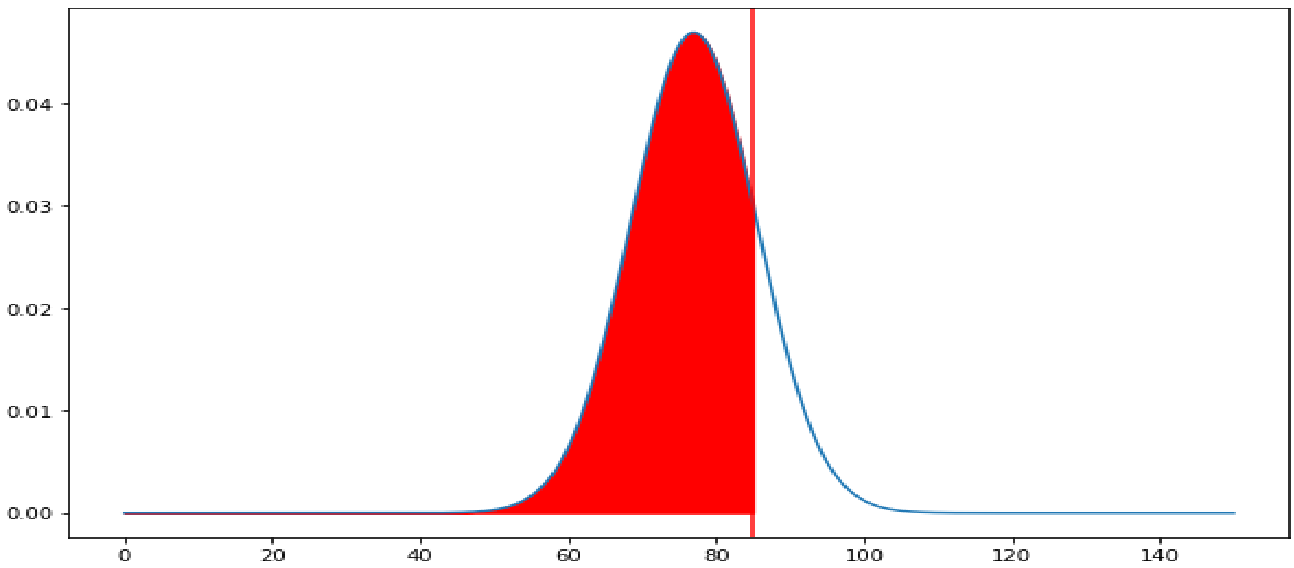
standard deviation(sigma)= 8.5



Plot 12: normal distribution of Grades of the final examination

4.1 What is the probability that a randomly chosen student gets a grade below 85 on this exam?

There is 82% chance that a randomly chosen student gets a grade below 85 on this exam.

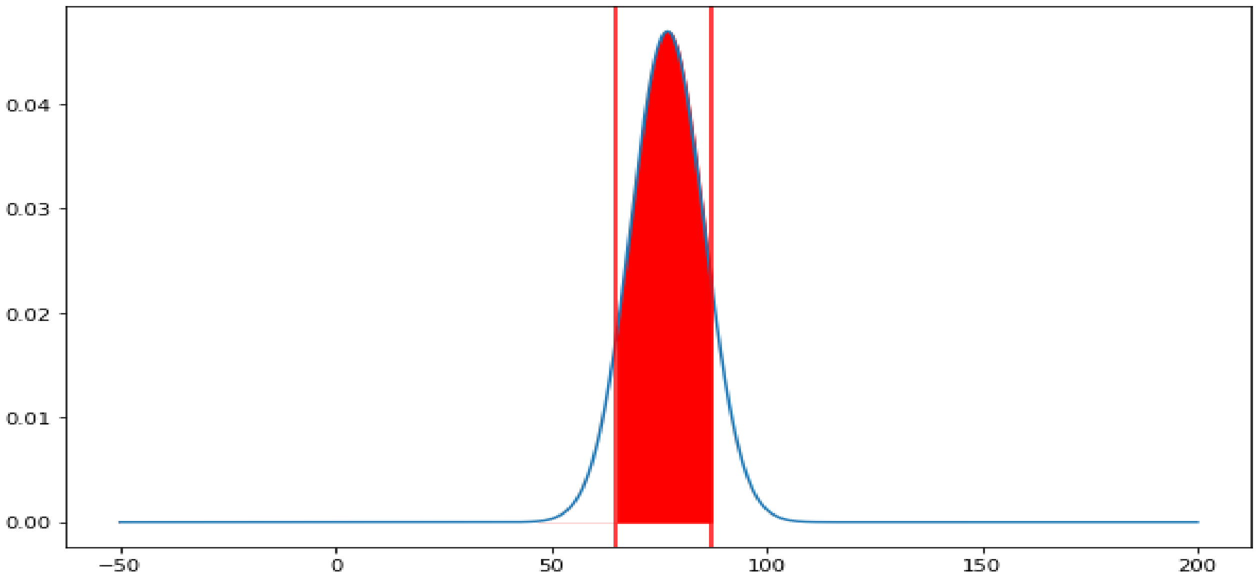


Plot 13: probability of randomly chosen students gets a grade below 85 on exam

```
probability that a randomly chosen student gets a grade below 85 on this exam is 0.826
```

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

There is an 80% chance that a randomly selected student score between 65 and 87.

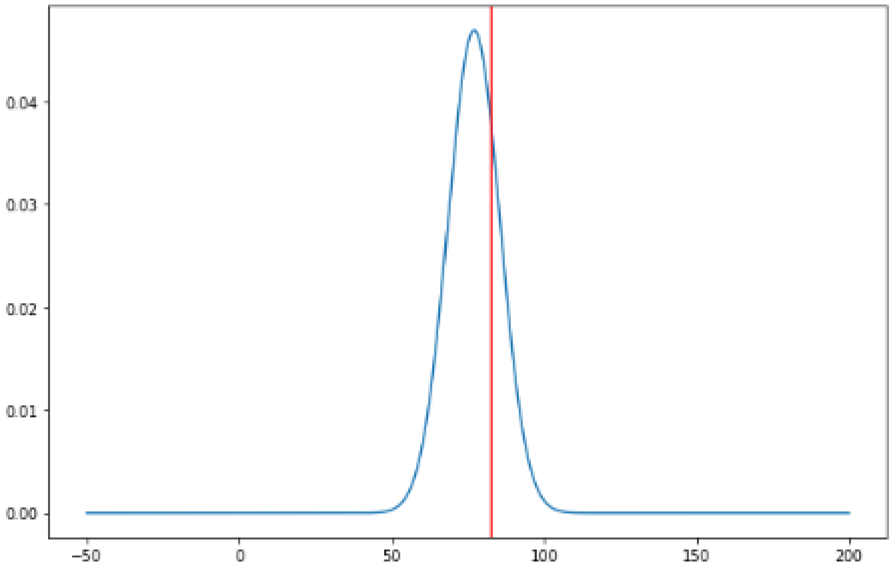


Plot 14: probability of randomly selected students' scores between 65 and 87

```
probability that a randomly selected student scores between 65 and 87 is 0.8012
```

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

The passing cut off is 82.7 so that 75% of the student clear the exam.



Plot 15: the passing cur off so that 75% of the students clear the exam.

```
The passing cut-off so that 75% of the students clear the exam is 82.73
```

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);

Few information about the data

- *Few rows of Data*

	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

Table 2: problem 5 few rows of data

- Statistic Description

	Unpolished	Treated and Polished
count	75.000000	75.000000
mean	134.110527	147.788117
std	33.041804	15.587355
min	48.406838	107.524167
25%	115.329753	138.268300
50%	135.597121	145.721322
75%	158.215098	157.373318
max	200.161313	192.272856

Table 3: table of problem 5 statistic description

- **There is 75 rows and 2 columns.**
- **There is no null value and the data is clean.**

mean of unpolished(μ1)=134.11

mean of polished(μ2)=147.70

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?

Solution:

To solve the question there are 5 steps involved which are as follows:

Step 1: Define null and alternate hypotheses

null hypotheses: $\mu 1 \geq 150$
alternative hypotheses: $\mu 1 < 150$

Where μ1 stands for mean of unpolished

null hypothesis: The null hypothesis is a typical statistical theory which suggests that no statistical relationship and significance exists in a set of given single observed variable, between two sets of observed data and measured phenomena.

alternate hypothesis: the alternative hypothesis is one of the proposed proposition in the hypothesis test.

Step 2: decide the level of significance

a =0.05.

Sample size = 75

Step 3: Identify the test statistic

as standard deviation is unknown and n=75. so, we use the t distribution and t statistic.

Step 4: Calculate p value and the t statistic

P value we get here is 8.342

T statistic is -4.164

Step 5: Decide to reject or accept null hypothesis

Level of significance: 0.05

p-value: 8.342

As we can see that p value is higher than 1 which means we cannot reject the null hypothesis.

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

Step 1: Define null and alternate hypotheses

null hypotheses: μ1 = μ2
alternative hypotheses: μ1 ≠ μ2

Where μ1 stands for mean of unpolished

Where μ2 stands for mean of polished

null hypothesis: The null hypothesis is a typical statistical theory which suggests that no statistical relationship and significance exists in a set of given single observed variable, between two sets of observed data and measured phenomena.

alternate hypothesis: the alternative hypothesis is one of the proposed proposition in the hypothesis test.

Step 2: decide the level of significance

α =0.05.

Sample size = 75

Step 3: Identify the test statistic

as standard deviation is unknown and n=75. so, we use the t distribution and t statistic.

Step 4: Calculate p value and the t statistic

P value we get here is 0.00060

T statistic is -3.58

Level of significance = 0.05

Step 5: Decide to reject or accept null hypothesis

As we can see that is more that the p value which means that we fail to accept the null hypothesis which was that the mean hardness of both the stones were same.

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.

Few information about the data

- *Few rows of data*

	Sr no.	Before	After
0	1	39	44
1	2	25	25
2	3	39	39
3	4	6	13
4	5	40	44

Table 4: problem 6 few rows of data

- Statistic description

	Sr no.	Before	After
count	100.000000	100.000000	100.000000
mean	50.500000	26.940000	32.490000
std	29.011492	8.806357	8.779562
min	1.000000	3.000000	10.000000
25%	25.750000	21.750000	26.000000
50%	50.500000	28.000000	34.000000
75%	75.250000	32.250000	39.000000
max	100.000000	47.000000	51.000000

Table 5: problem6 statistic description

- There are 100 rows and 3 columns which are s.no, Before, After.
- The data is clean as the data is free from abnormalities and missing values.

Solution:

Note: We will add a new column named 'diff' to know the difference between the after and before column. This column will help is assessing the result.

Now the data is like:

	Sr no.	Before	After	diff
0	1	39	44	5
1	2	25	25	0
2	3	39	39	0
3	4	6	13	7
4	5	40	44	4

Table 6: problem 6 few rows of data with new column

To solve the question there are 5 steps involved which are as follows:

Step 1: Define null and alternate hypotheses

Null hypotheses: mean of diff (μ) ≥ 5

Alternate hypotheses: mean of diff (μ) < 5

null hypothesis: The null hypothesis is a typical statistical theory which suggests that no statistical relationship and significance exists in a set of given single observed variable, between two sets of observed data and measured phenomena.
alternate hypothesis: the alternative hypothesis is one of the proposed proposition in the hypothesis test.

Step 2: decide the level of significance

a =0.05.

Sample size = 100

Step 3: Identify the test statistic

as standard deviation is unknown and n=100.

so, we use the t distribution and t statistic.

Step 4: Calculate p value and the t statistic

P value we get here is 0.0583

T statistic is 1.9148

To calculate p value this is the formula:

Z=^p−p0√p0(1−p0) n

Where,

^p=p^=Sample Proportion

P0=P0= assumed population proportion in the null hypothesis

P-value defines the probability of getting a result that is either the same or more extreme than the other actual observations. The P-value represents the probability of occurrence of the given event.

To calculate the T statistic, take the difference between the population mean and the sample mean, then divide the result by the result of the standard deviation divided by the square root of the sample size

In statistics, the t-statistic is the ratio of the departure of the estimated value of a parameter from its hypothesized value to its standard error.

Step 5: decide to reject or accept the null hypothesis

Level of significance is 0.05

P value is 0.0583

And we know that if p value is low null will fly it means if the value of p is less than the level of significance then the null is rejected and vice versa.

So null is not rejected as p value is more than the level of significance.

Conclusion

So, from the hypothesis testing we conclude that the rigorous program for body conditioning is actually successful. The program is successful as the candidates are able to do more than 5 push-ups, as compared to when he/she enrolled in the program.