Assignment:
Inferential and
Hypothesis Testing

Problem Statement:

Comprehension

The pharmaceutical company Sun Pharma is manufacturing a new batch of painkiller drugs, which are due for testing. Around 80,000 new products are created and need to be tested for their time of effect (which is measured as the time taken for the drug to completely cure the pain), as well as the quality assurance (which tells you whether the drug was able to do a satisfactory job or not).

Question 1

The quality assurance checks on the previous batches of drugs found that — it is 4 times more likely that a drug is able to produce a satisfactory result than not.

Given a small sample of 10 drugs, you are required to find the theoretical probability that at most, 3 drugs are not able to do a satisfactory job.

Let's Consider

P(S)=Probability of satisfactory result

P(NS)= Probability of not satisfactory result

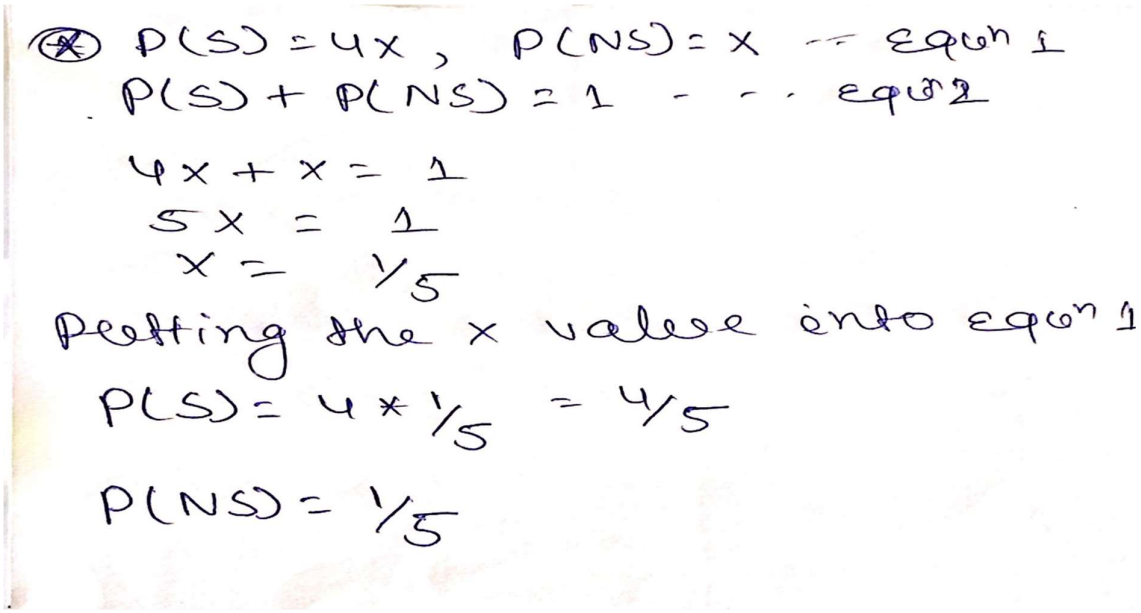
As per problem statement "It is 4 times more likely that a drug is able to produce a satisfactory result than not"

P(NS)= X

P(S)= 4X

As we know the sum of all possible probabilities should be 1

P(S)+P(NS)=1



Question 1.A

Propose the type of probability distribution that would accurately portray the above scenario, and list out the three conditions that this distribution follows.

Solution: -

From the details provided in the question, we can approach this as binomial distribution and below are the properties of binomial distribution.

- The **total number** of trials is **fixed** at **10**.
- Each trial is **binary**, i.e., Drugs can be either “Effective” or “Not effective”
- Probability **of success** is the **same** in all trials, **P=1/5**.

Question 1.B

Calculate the required probability.

Solution: -

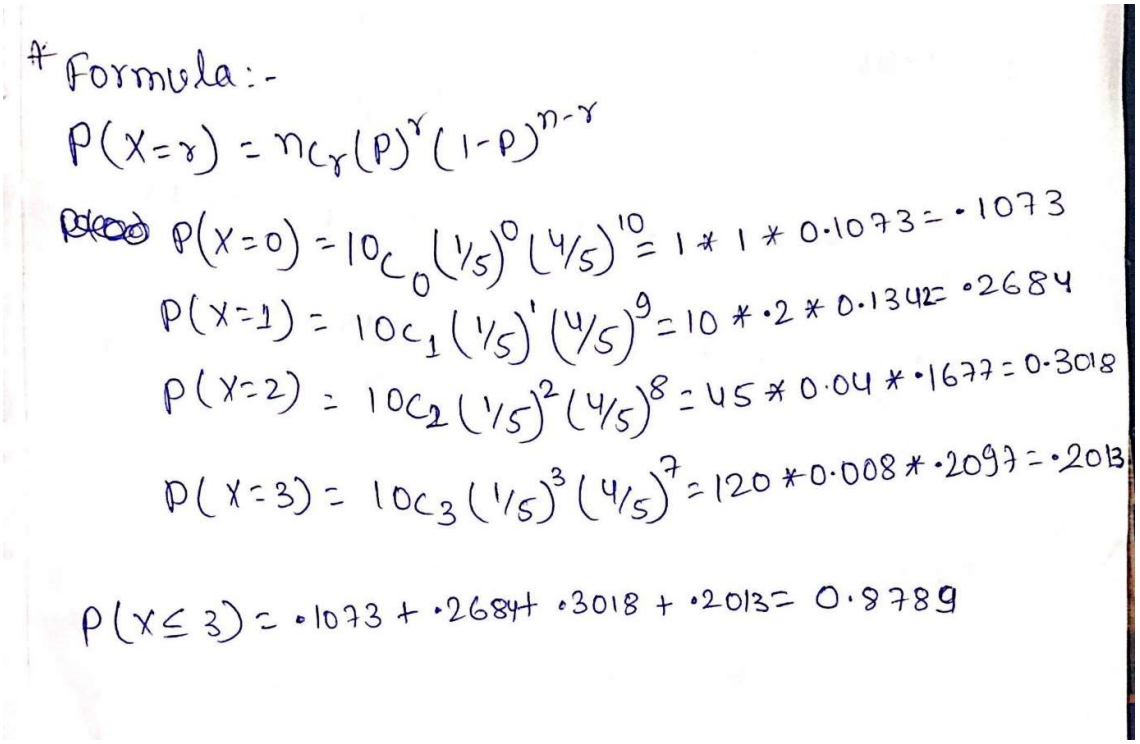
As per question “At most, 3 drugs are not able to do a satisfactory job”

Given:

n=10

Required Probability=P(X≤3)

P(X≤3) =P(X=0) + P(X=1) + P(X=2) + P(X=3)



X	P(X=r)
0	.1073
1	.2684
2	.3018
3	.2013

Answer: -Required probability=0.8789

Question 2

For the effectiveness test, a sample of 100 drugs was taken. The mean time of effect was 207 seconds, with the standard deviation coming to 65 seconds. Using this information, you are required to estimate the range in which the population mean might lie — with a 95% confidence level.

Question 2.A

Discuss the main methodology using which you will approach this problem. State all the properties of the required method. Limit your answer to 150 words.

Solution: -

From the details provided in the question, we can approach the problem using **Central Limit Theorem** as it meets all the 3 below conditions**.**

- Sampling distribution mean **($\mu_{\bar{X}}$)**= Population mean **(μ)**
- Sampling distribution standard deviation (Standard error) = $\frac{\sigma}{\sqrt{n}}$ where **σ** is the population standard deviation.
- For **n > 30**, the sampling distribution becomes a normal distribution

Question 2.B

Find the required range.

Solution: -

Considering population standard deviation same as Sample standard deviation because population standard deviation is not given in problem description.

Step 1: -Take a sample of size n.

Sample Size(**n**) =100

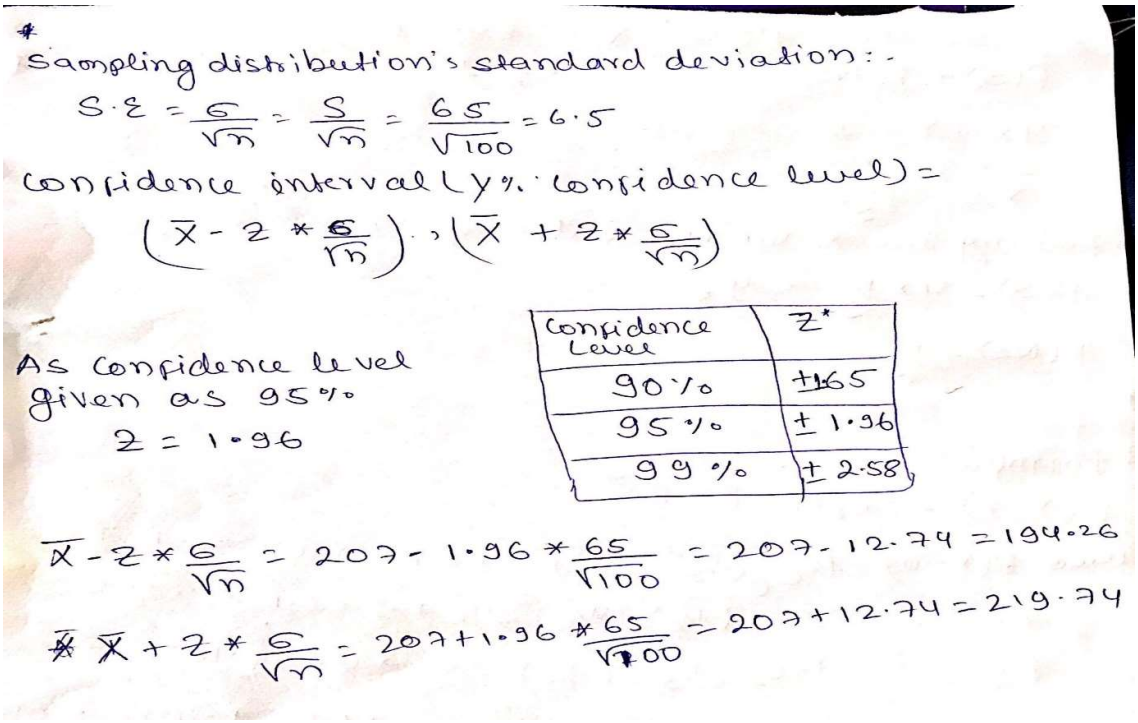
Step 2: - Then, find the mean $\bar{X}$ and standard deviation S of this sample

Step 3: - Now, you can say that for a y% confidence level, the confidence interval for the population mean μ is given by $(\bar{X} - \frac{Z^*S}{\sqrt{n}}, \bar{X} + \frac{Z^*S}{\sqrt{n}})$

Sampling distribution mean ($\mu_{\bar{X}}$)= Population mean **(μ)**(Unknown)=207 Secs

Sampling standard deviation(S)= 65 Secs

Confidence level(Y%) =95%



Answer: **Required Range will be between 194.26 and 219.74**

Question 3.A

The painkiller drug needs to have a time of effect of at most 200 seconds to be considered as having done a satisfactory job. Given the same sample data (size, mean, and standard deviation) of the previous question, test the claim that the newer batch produces a satisfactory result and passes the quality assurance test. Utilize 2 hypothesis testing methods to make your decision. Take the significance level at 5 %. Clearly specify the hypotheses, the calculated test statistics, and the final decision that should be made for each method.

Solution: -

Given

Sample Size(n)=100

Sample Mean ($\bar{X}$)=207 Secs

Standard Deviation = 65 Secs

Significance Level =0.05=5%

Null Hypothesis H0 ≤ 200 Secs

Alternate Hypothesis H1 >200 Secs

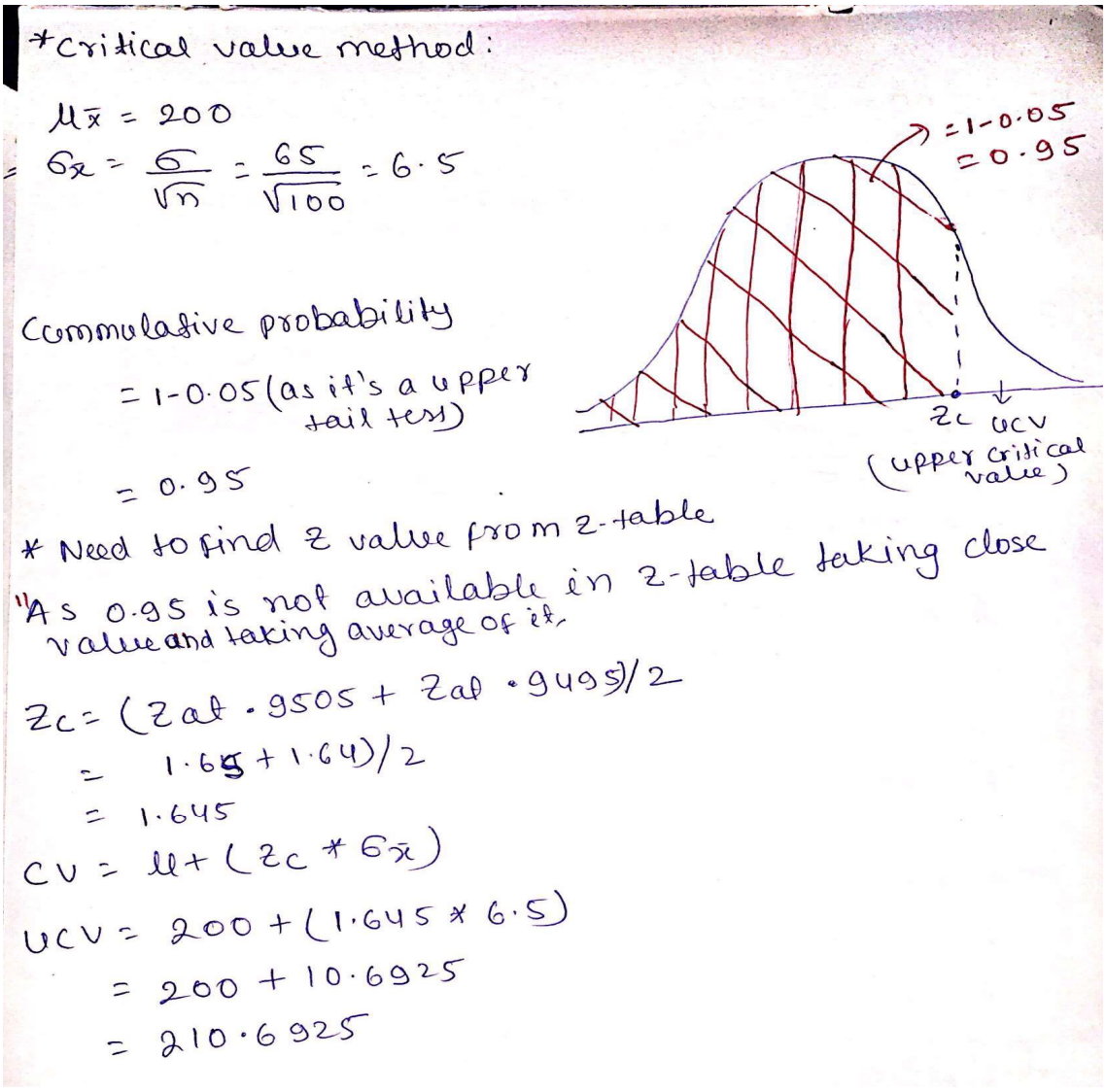
As H1 has > sign so it will be 1 tail (Upper) test

Critical Value Method:

There will be 3 steps to calculate Critical Value

Step 1: -Calculate the value of Zc from the given value of α (significance level).

Step 2: -Calculate the critical values (UCV and LCV) from the value of Zc.



Step 3: -Make the decision on the basis of the value of the sample mean x with respect to the critical values (UCV AND LCV).

Since the mean value **207 secs** is less than the UCV (**210.69**),

So, failed to reject Null Hypothesis

P Value Method:

There will be 3 steps to calculate Critical Value

Step 1: -Calculate the value of the z-score for the sample mean point on the distribution.

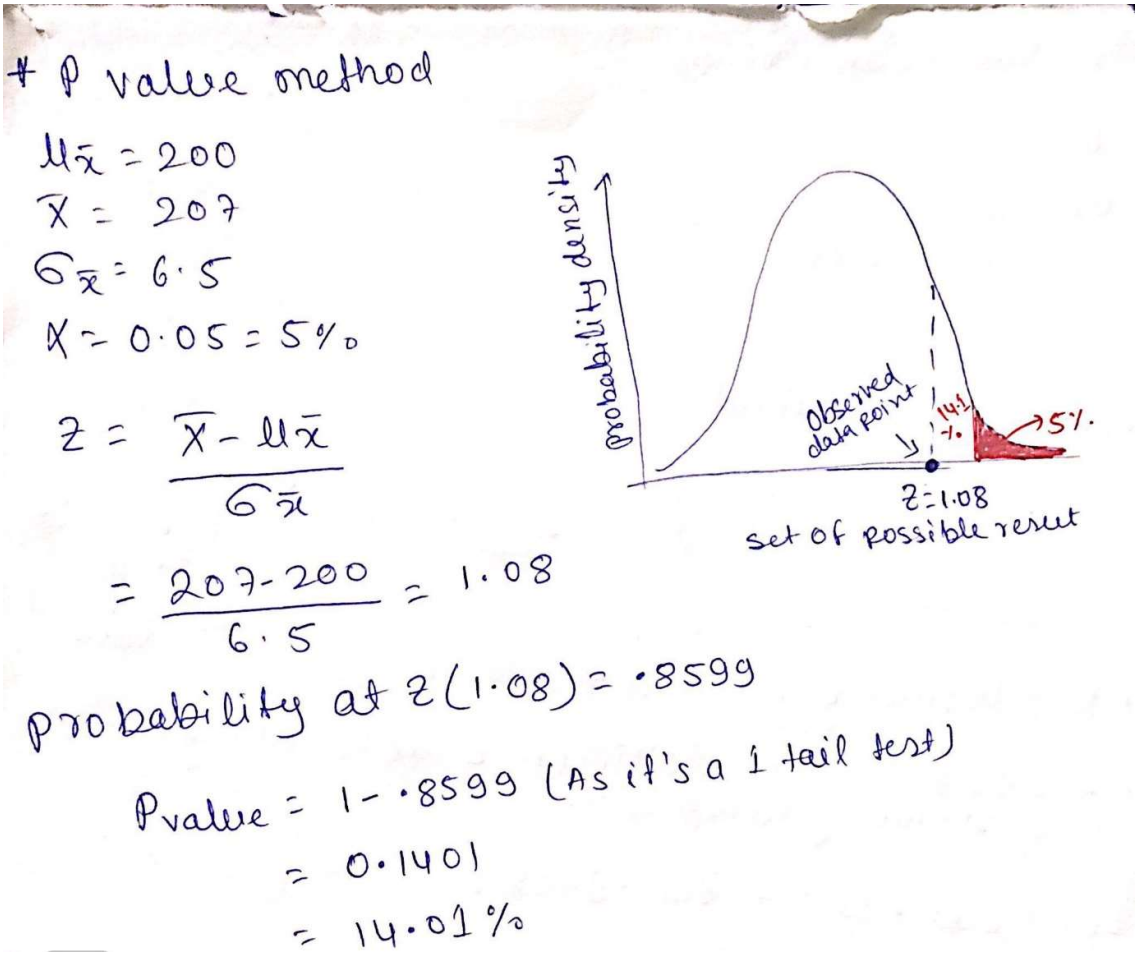
Step 2: -Calculate the p-value from the cumulative probability for the given z-score using the z-table.

$\mu_{\bar{X}}$=200

Sample Mean ($\bar{X}$)=207 Secs

Standard Deviation = 65 Secs

Significance Level =0.05=5%



Step 3: -Make a decision on the basis of the p-value with respect to the given value of α (significance value).

Since P value 14.01%> 5 %,

So, failed to reject Null Hypothesis

Question 3.B

You know that two types of errors can occur during hypothesis testing — namely Type-I and Type-II errors — whose probabilities are denoted by α and β respectively. For the current sample conditions (sample size, mean, and standard deviation), the value of α and β come out to be 0.05 and 0.45 respectively.

Now, a different sampling procedure (with different sample size, mean, and standard deviation) is proposed so that when the same hypothesis test is conducted, the values of α and β are controlled at 0.15 each. Explain under what conditions would either method be more preferred than the other, i.e. give an example of a situation where conducting a hypothesis test having α and β as 0.05 and 0.45 respectively would be preferred over having them both at 0.15. Similarly, give an example for the reverse scenario - a situation where conducting the hypothesis test with both α and β values fixed at 0.15 would be preferred over having them at 0.05 and 0.45 respectively. Also, provide suitable reasons for your choice (Assume that only the values of α and β as mentioned above are provided to you and no other information is available).

Solution:

Given

	Case 1	Case 2
α	0.05	0.15
β	0.45	0.15

H0=Drug is effective

H1≠ Drug is effective

- A **type I-error**, represented by α, occurs when you reject a true null hypothesis.
- A **type-II error**, represented by β, occurs when you fail to reject a false null hypothesis.

	Drug is effective	Drug is not effective
Reject Null Hypothesis(H0)	**Type 1 error α**	Correct
Failed to reject Null Hypothesis(H1)	Correct	**Type 2 error β**

Case 1 when **α=0.05** and **β=0.45**

Scenario 1: -Impact to company if Type 1 and Type 2 error happened during the test

Type 1 error

- Company will stop the manufacturing of drug even drug was **effective**.
- There is no harm to customer.
- It will increase the manufacturing cost for the company

Type 2 error

- Company will not stop the manufacturing of drug even drug was **not effective**.
- It causes harm to customer.
- It will decrease the brand value of the company as drug was not effective.

Let's consider drug is launched to the market with Type 1 and Type 2 error.

Scenario 2: -Impact to customer (Non sensitive drug like Paracetamol) if Type 1 and Type 2 error happened during the test

Type 1 error

- Customer will get cured from disease because drug was **effective**.
- There will be no harm to the company

Type 2 error

- Customer will not get cured from disease because drug was **not effective**.
- Customer will stop buying product as company drug was not effective.
- Company sale will decrease.

Scenario 3: -Impact to customer (Sensitive drug like Heart Medicine) if Type 1 and Type 2 error happened during the test

Type 1 error

- Customer will get cured from disease because drug was **effective**.
- There will be no harm to the company

Type 2 error

- It will harm the Customer and customer died drug was **not effective**.
- There will be serious issue with the company and company might get closed.

Case 2: -when **α=0.15** and **β=0.15**

As α value is increased from 0.05 to 0.15

- **There will be no harm to the customer as they will get the correct drugs**
- **But it will increase the manufacturing cost to the company**

As β value is decreased from 0.45 to 0.15

- **As β value is decreased it will reduce the chances of harm to the customer and this will be good for Company's brand value and sale.**

Question 4

Now, once the batch has passed all the quality tests and is ready to be launched in the market, the marketing team needs to plan an effective online ad campaign to attract new customers. Two taglines were proposed for the campaign, and the team is currently divided on which option to use.

Explain why and how A/B testing can be used to decide which option is more effective. Give a stepwise procedure for the test that needs to be conducted.

Solution: -

In case of two-sample proportion test if it's difficult to find the correct option to choose then **A/B** testing comes into picture. **A/B** testing provides a way to test two different versions of the same element and see which one performs better.

Procedure to do A/B testing:

Step 1. Define Control and Challenger

Tagline 1 (Control) and Tagline 2 (Challenger)

Step 2. Create a hypothesis and consider significance value(α)=0.05

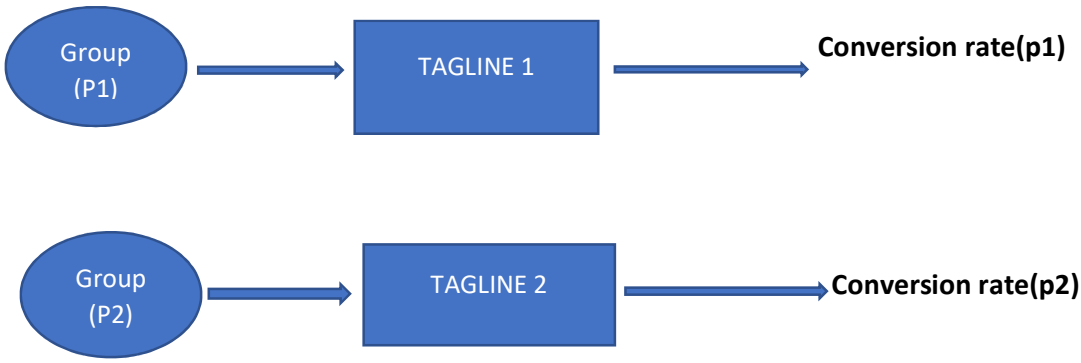
H0= Tagline 1 is better

Ha≠ Tagline 1 is better

Significance value(α)=0.05

Step 3. Collecting the data

- We will take some people (say 1000) to participate in A/B testing.
- Divide 1000 people into 2 groups of 500 each and will say 1st group as P1 and 2nd Group as P2 and give Tagline 1 to P1 and Tagline 2 to P2 group
- We will collect the response from both group and in terms of "True" or "False"
- Convert "True" as 1 and "False" as 0 to calculate the P value



Step 4: To find P value

Find the P value by using any tool like XLSAT.

- First find the frequency and sample size of both control and variable.
- Provide above frequency and sample size in tool like frequency 1, frequency 2 and sample size 1 and sample size 2.
- Select the significance level as 5%
- Select the range for the output
- Verify the frequency and sample size once result is populated

Step 5: Decision Making:

- **If P value is greater than the significance value then we cannot reject the null hypothesis and Tagline 1 will be better for campaign.**
- **If P value is less than the significance value then we can reject the null hypothesis and Tagline 2 will be better for campaign.**
- **If we don't have the proper significant value to reject or failed to reject null hypothesis then we should increase the sample size to conclude the test result.**