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.

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

Answer:

The type of probability distribution required for this problem is **Binomial Distribution.** All the three conditions listed below, can be seen in the given problem which helped in concluding the required distribution.

$$P(X=r)= {}_nC_r(p)_r(1-p)_{n-r}$$

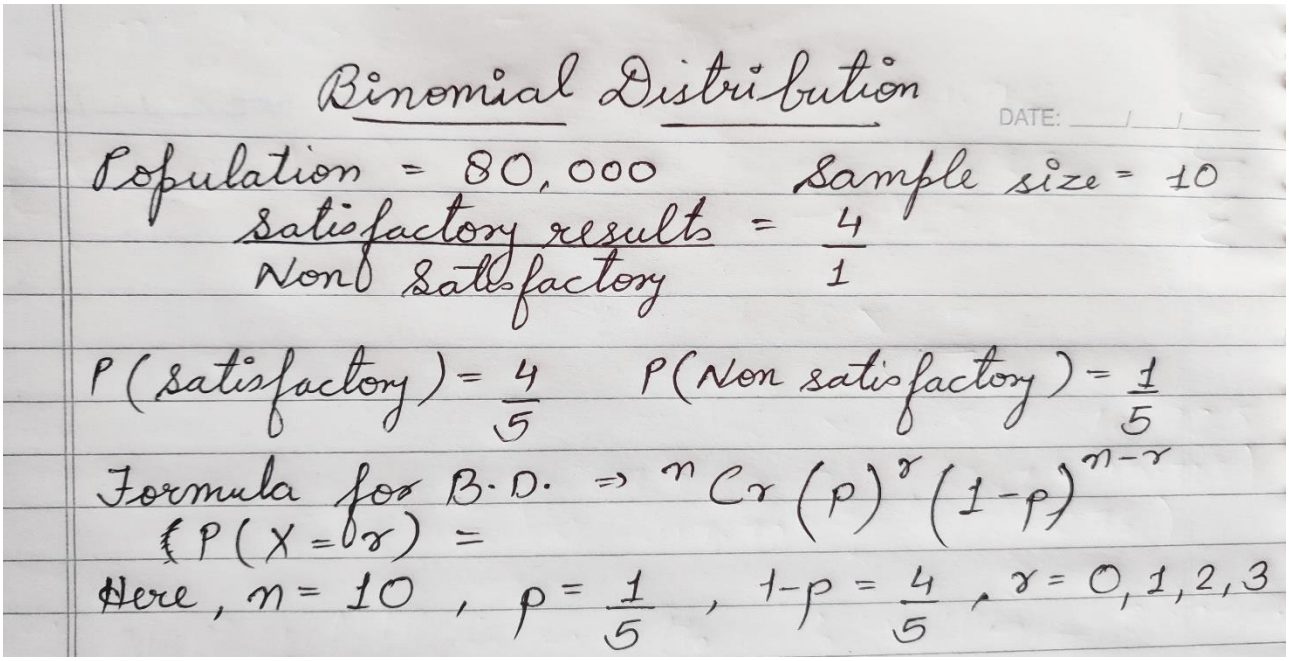
Three conditions of binomial distribution are-

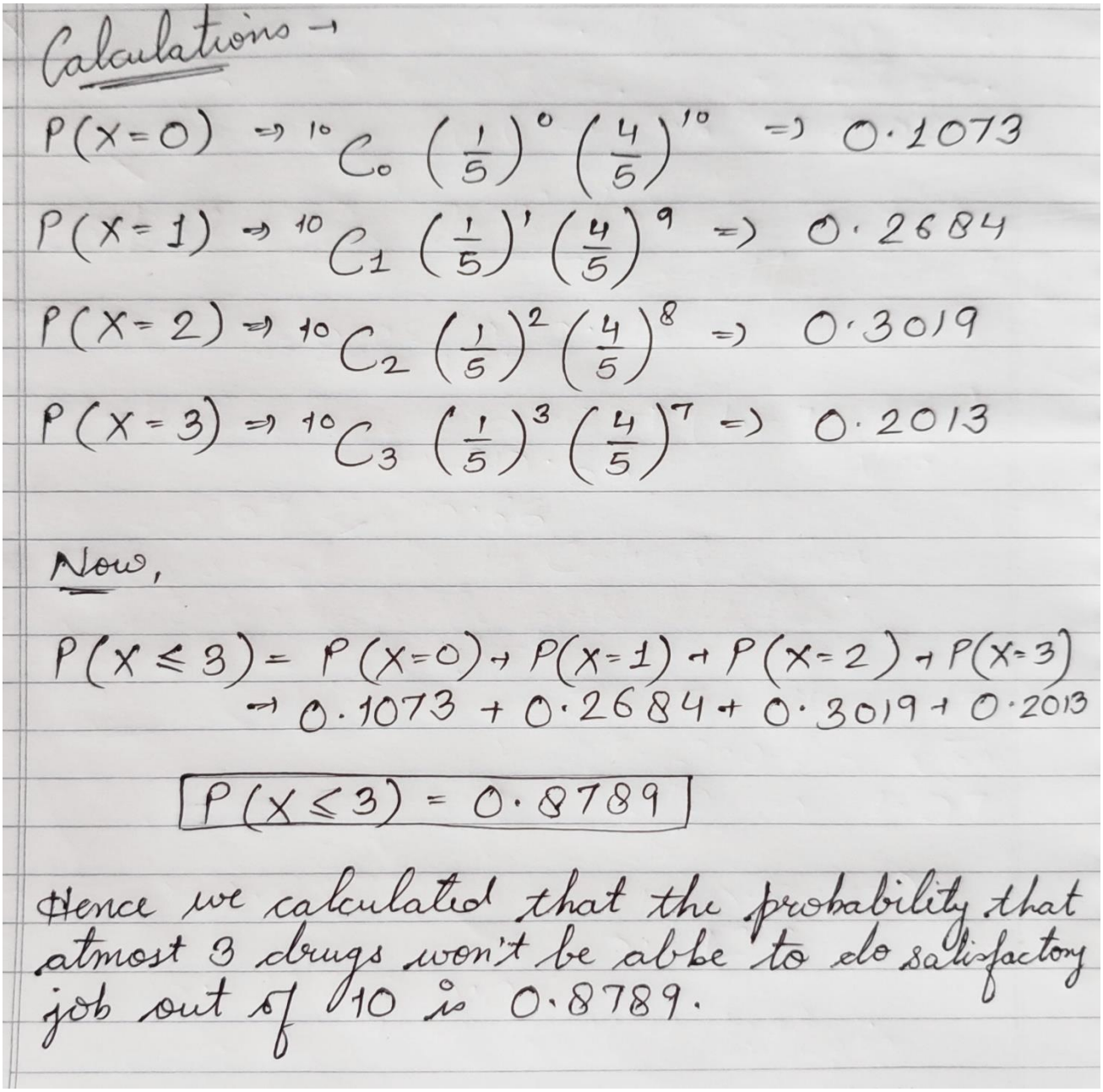
1. Total numbers of trial are fixed which is denoted by n.

2. Each trial is binary, i.e. has only two possible outcomes, either success or failure.

3. Probability of success is same in all the trials, which is denoted by p.

b.) Calculate the required probability.





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 interval in which the population mean might lie — with a 95% confidence level.

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.

Answer:

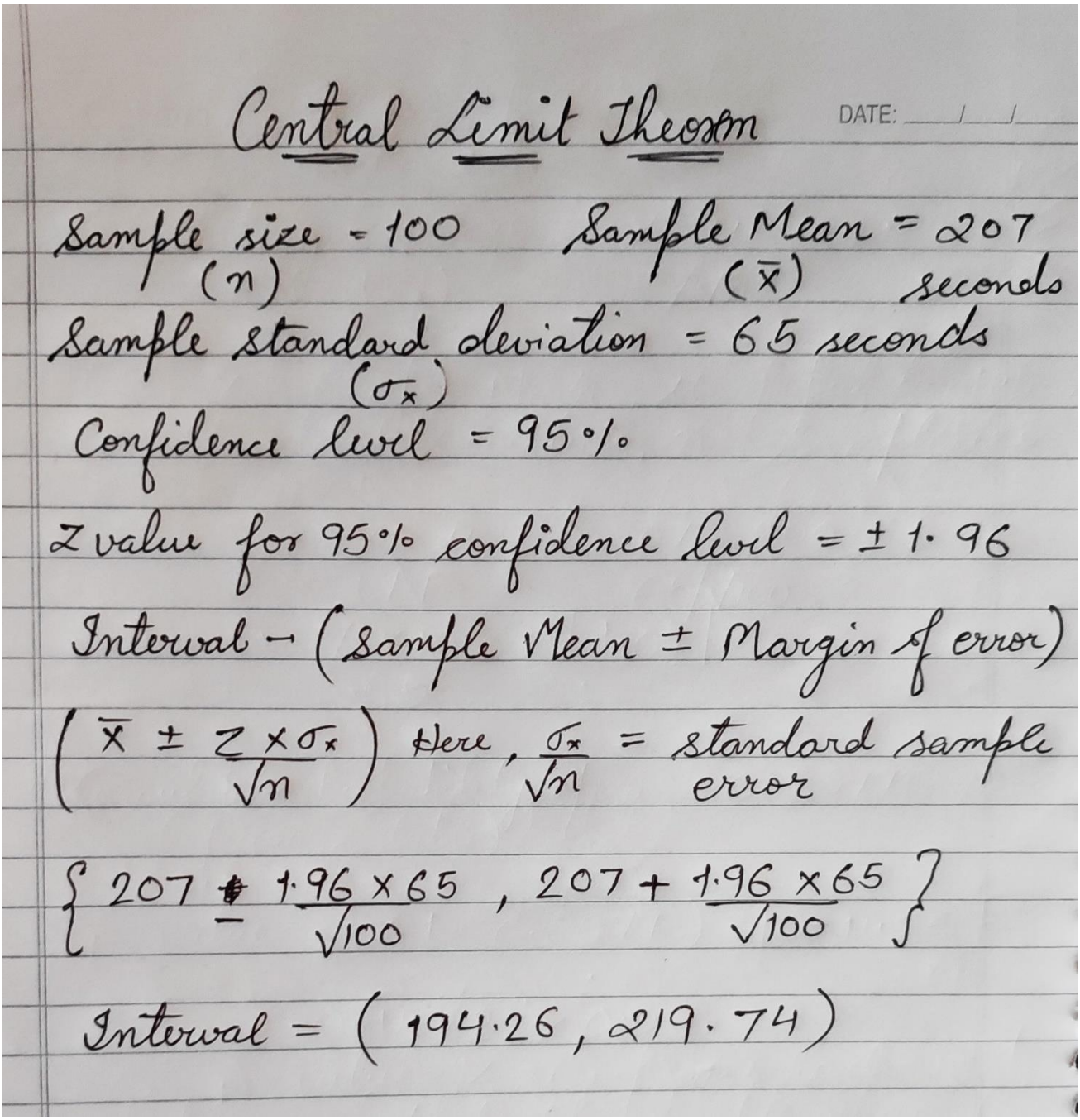
In this question, we were given with the sample mean but not the population mean or sample variables. The methodology which works on sample mean and can be used to calculate interval in which the population mean

might lie. **Central Limit Theorem** is the one which comes in mind when we need to work with only sample mean.

There are some properties of central limit theorem which need to be applied on the given data to find the interval.

1. **Sampling distribution's mean** (μ_x) = **Population mean** (μ)
2. Sampling distribution's standard deviation (**Standard error**) = $\sigma/n^{1/2}$
3. **For n > 30**, the sampling distribution becomes a **normal distribution**

b.) Find the required interval.



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.

Answer: **Critical Value Method of testing**

Null Hypothesis → H_0 → Time of effect $\leq$ 200 sec

Alternate Hypothesis (H_1) → Time of effect > 200 sec

Population Mean (μ) = 200 sec.

Sample Mean ($\bar{x}$) = 207 sec

Sample size (n) = 100 Standard Deviation (σ_x) = 6.5 sec.

Significance level → 5% = 0.05

Critical Value Method

Upper Tailed Test because alternate hypothesis has '>' sign

Standard Error ($\sigma_{\bar{x}}$) = $\frac{\sigma}{\sqrt{n}} \rightarrow \frac{65}{\sqrt{100}} \rightarrow 6.5$

Probability of failing to reject H_0 = (1 − 0.05) → 0.95

Z-critical value (Z_c) for 0.95 → [1.64]

Finding the critical value (C.V) → $\mu + (Z_c \times \sigma_{\bar{x}})$

→ 200 + (1.64 × 6.5)

→ [210.66]

Region beyon 210.66 is critical region.

207 < 210.66

Sample mean does not lie in the critical region.

Fail to reject the Null Hypothesis

This is a Critical Value Method in which we find the critical value or critical region and compare it with the sample mean. If the sample mean is in the critical region, we reject the null hypothesis. Otherwise, we fail to reject the null hypothesis.

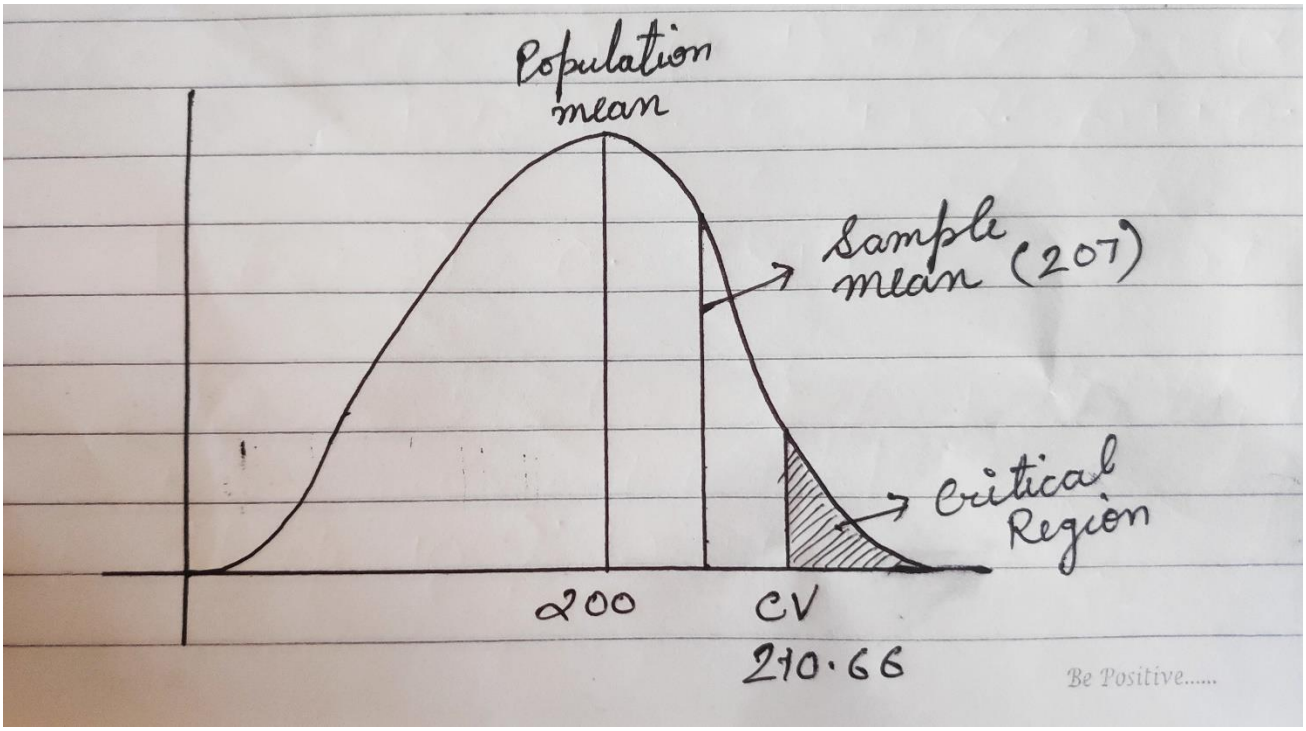


Figure 1: Graph for Critical Value Method

P-Value Approach of testing

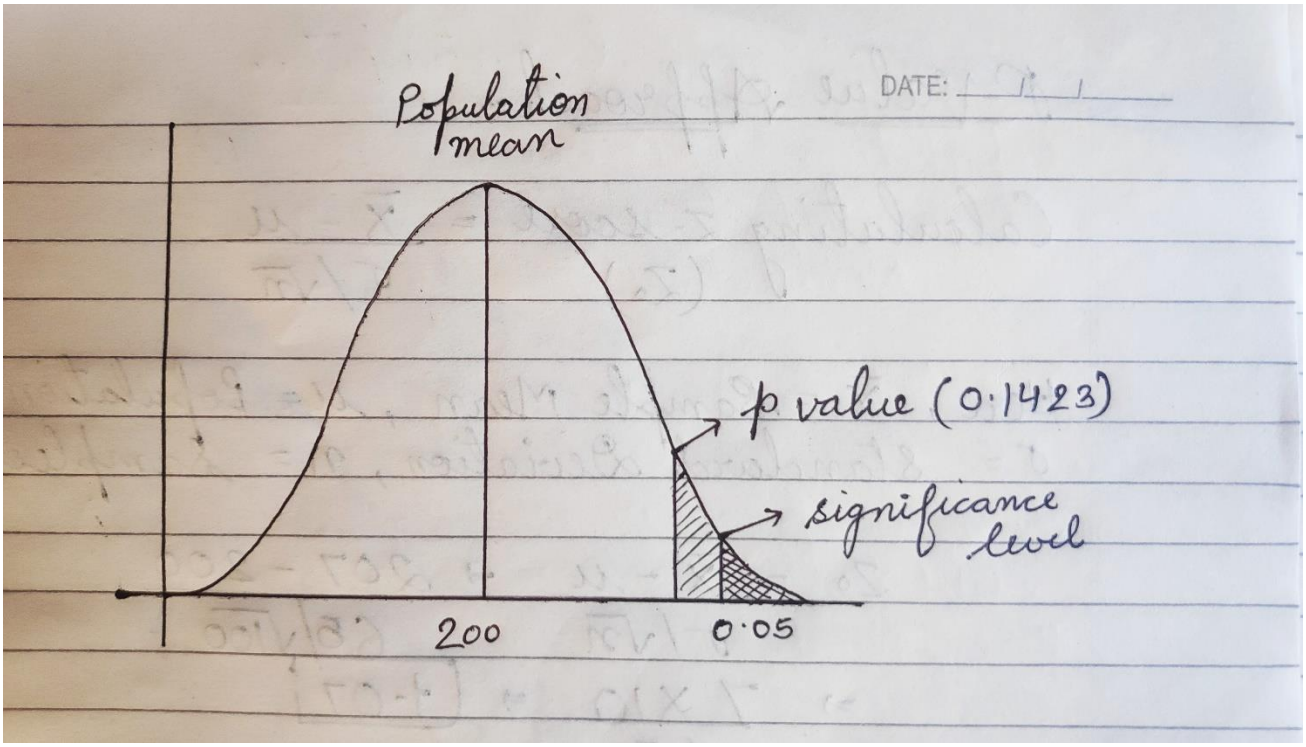
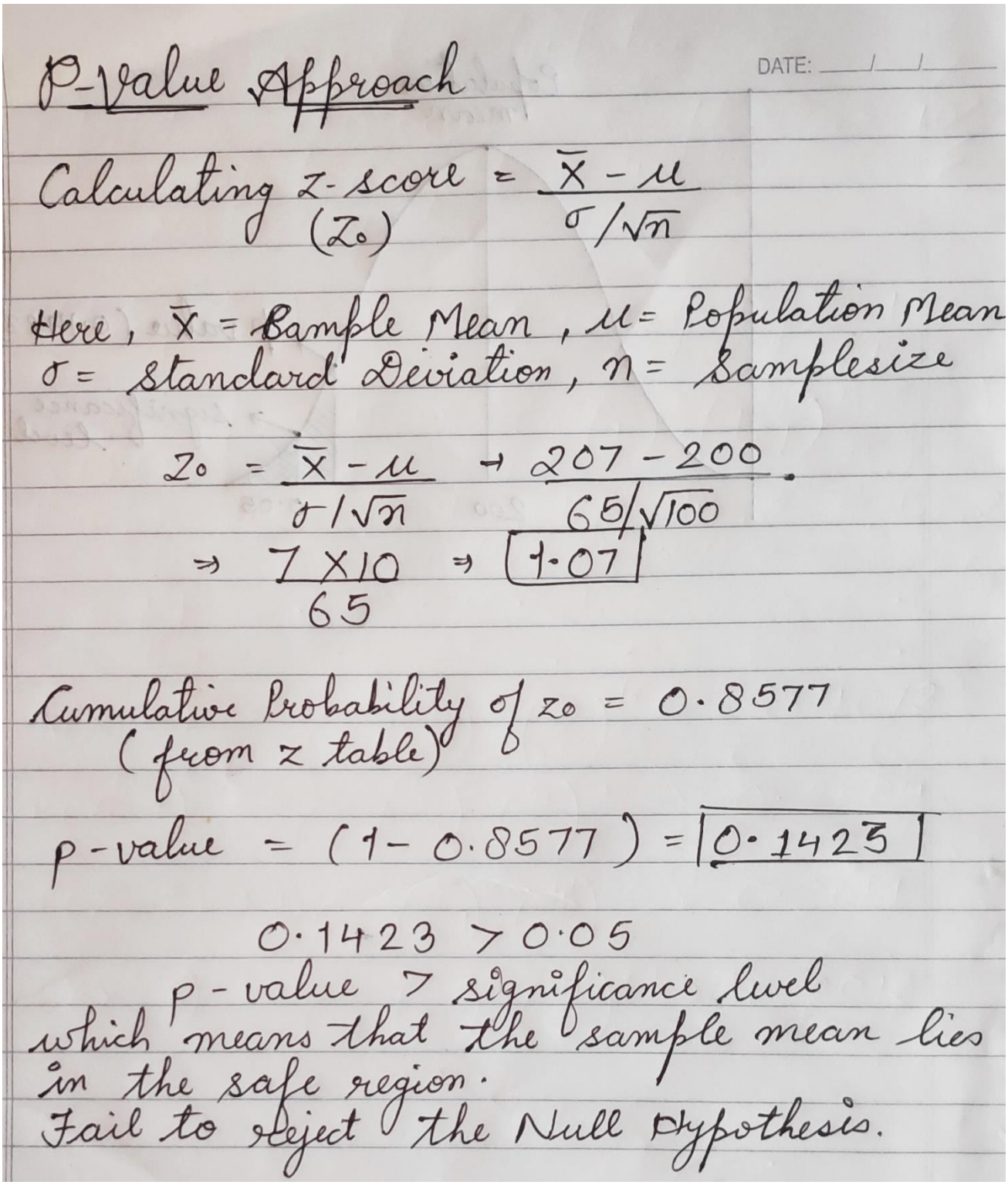


Figure 2: Graph for p-Value Test

P-Value test is a statistical sigifocance test which checks the largest probability of obtaining test results at least as extreme as the results actually observed, under the assumption that the null hypothesis is correct or true. If the p-value is greater than the significance level, we can say that we failed to reject the null hypothesis and if we have a smaller p-value, the null hypothesis gets rejected.



After both the testing, it can be said that we failed to reject the null hypothesis. The time of effect of the drugs produced is at most 200 seconds.

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 (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 the hypothesis test with α and β as 0.05 and 0.45 respectively would be preferred over conducting the same hypothesis test with α and β at 0.15 each. Similarly, give an example for the reverse scenario- where conducting the same hypothesis test with α and β at 0.15 each would be preferred over having them at 0.05 and 0.45 respectively.

For each example, give suitable reasons for your particular choice using the given values of α and β only. (Assume that no other information is available. Also, the hypothesis test that you are conducting is the same as mentioned in the previous question - you need to test the claim whether the newer batch produces a satisfactory result.)

Answer:

For first case, α = 0.05 and β = 0.45

This means that in 5% of cases, the painkiller drug has a time of effect of at most 200 seconds but we rejected that. And that in 45% of cases, the painkiller drug has a time of effect of at more than 200 seconds but we accepted that. This means roughly 45% of our drugs did not produce a satisfactory result and 5% of drugs that could produce satisfactory results were rejected.

The result of the painkiller drug will be unsatisfactory in roughly 45% cases, hence, it's a concern for the company. Only 5% of good drugs are being rejected, hence, it will reduce the wastage. It will reduce the wastage of good drugs but the accepted drugs will be very less effective. **If the company wants to focus only on profits and reduce operational cost and wastage, this sampling method is better.**

For second case, α = 0.15 and β = 0.15

This means that in 15% of cases, the painkiller drug has a time of effect of at most 200 seconds but we rejected that. And that in 15% of cases, the painkiller drug has a time of effect of at more than 200 seconds but we accepted that. This means roughly 15% of our drugs did not produce a satisfactory result and 15% of drugs that could produce satisfactory results were rejected.

The result of the painkiller drug will be unsatisfactory in roughly 15% cases which is less than the previous case. But the 15% of good drugs are being rejected which is 3 times the previous case. It will have more wastage of good drugs but the accepted drugs will be much more effective. **If the company wants to focus on the effectiveness of painkiller drug and is comfortable with higher operational cost and wastage, this sampling method is better.**

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 subscribers. 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.

Answer:

A/B testing is a method of comparing two webpages or web apps against one another and analyze the to analyze which one performs better. In this, two variations of a webpage are shown to the users and some analysis is performed to determine.

In this example, the marketing team wants to test the taglines for their online ad campaign. Two taglines were proposed and now they need to test which one's going to attract more new subscribers. The testing can be performed using the below listed steps.

1. Need two different ad templates for the testing, with one tagline on each of them.

2. Half of the users must be shown ad with one tagline, and half of them must be shown the other one.

3. The users will click on the ad if they get attracted by the tagline of the ad or for some other region.

4. The analysis should be done (for e.g. Using google analytics) for how many users clicked the ad with one tagline and how many clicked the other.

5. The one which shows the highest variation and attract most new subscribers should be chosen in the final decision.