Statistics Assignment

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 that would accurately portray the above scenario is **Binomial distribution**. As it follows the following three conditions.

- Total number of trials is fixed at n
- Each trial is binary, i.e., has only two possible outcomes - success or failure
- Probability of success is same in all trials, denoted by p

---

b.) Calculate the required probability.

Answer:

Let p be the probability that the drug is satisfactory.
p = 4 (1-p)
p = 0.8

We have to to find the probability that at most, 3 drugs are not able to do a satisfactory job. Here probability for unsatisfactory job is (1-p).

$P(X=x_i) = {}_nC_r * (1-p)^r * p^{(n-r)}$

X	P(X)
0	0.1073
1	0.2684
2	0.3019
3	0.2013
Total	0.8791

Thus the probability that at most 3 drugs are not able to do a satisfactory job is 0.88

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.

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:

We can approach this problem using Central Limit Theorem.

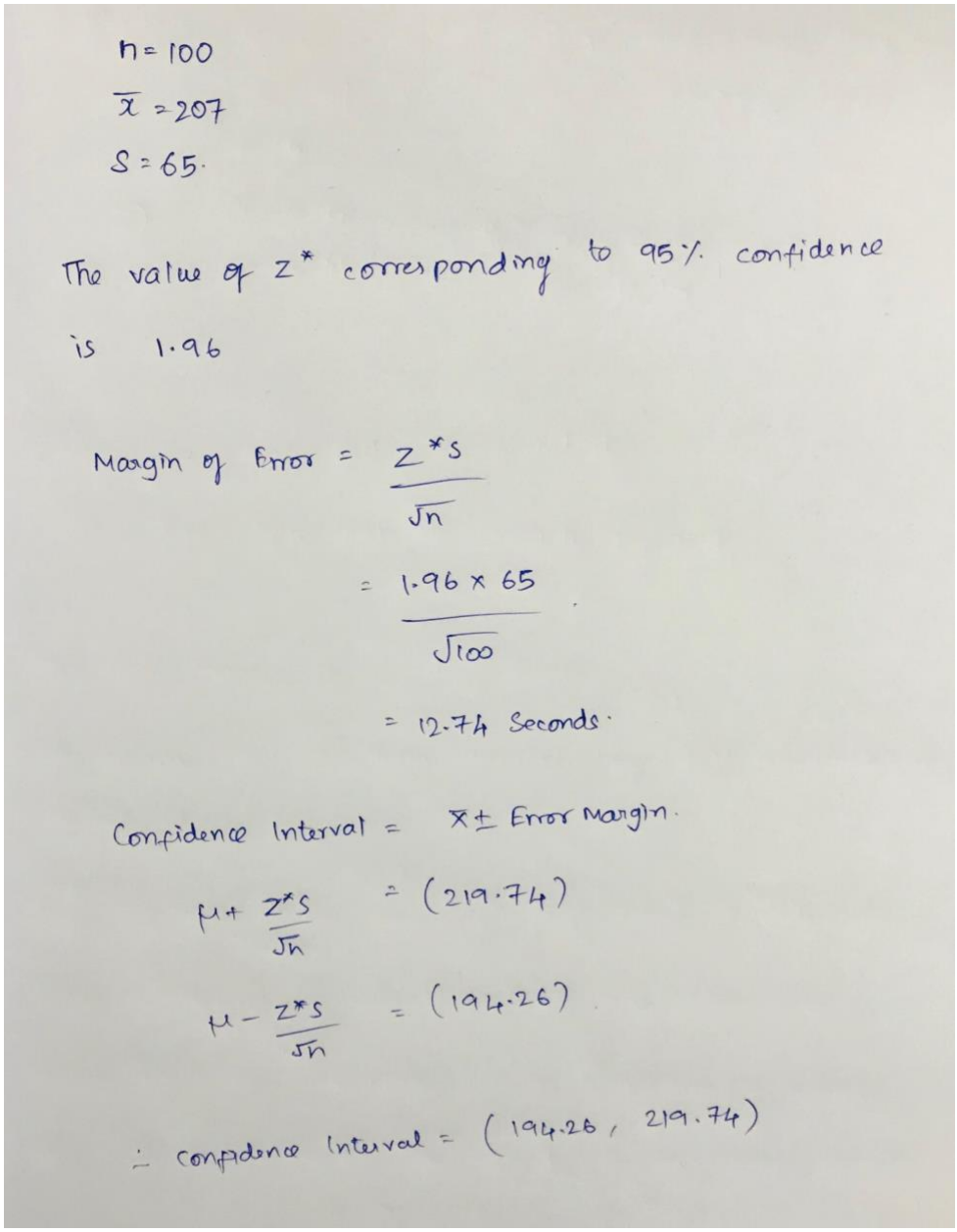
The central limit theorem says that, for any kind of data, provided a high number of samples has been taken, the following properties hold true:

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

---

b. Find the required range.

Answer:



Thus the confidence interval is (194.26, 219.74)

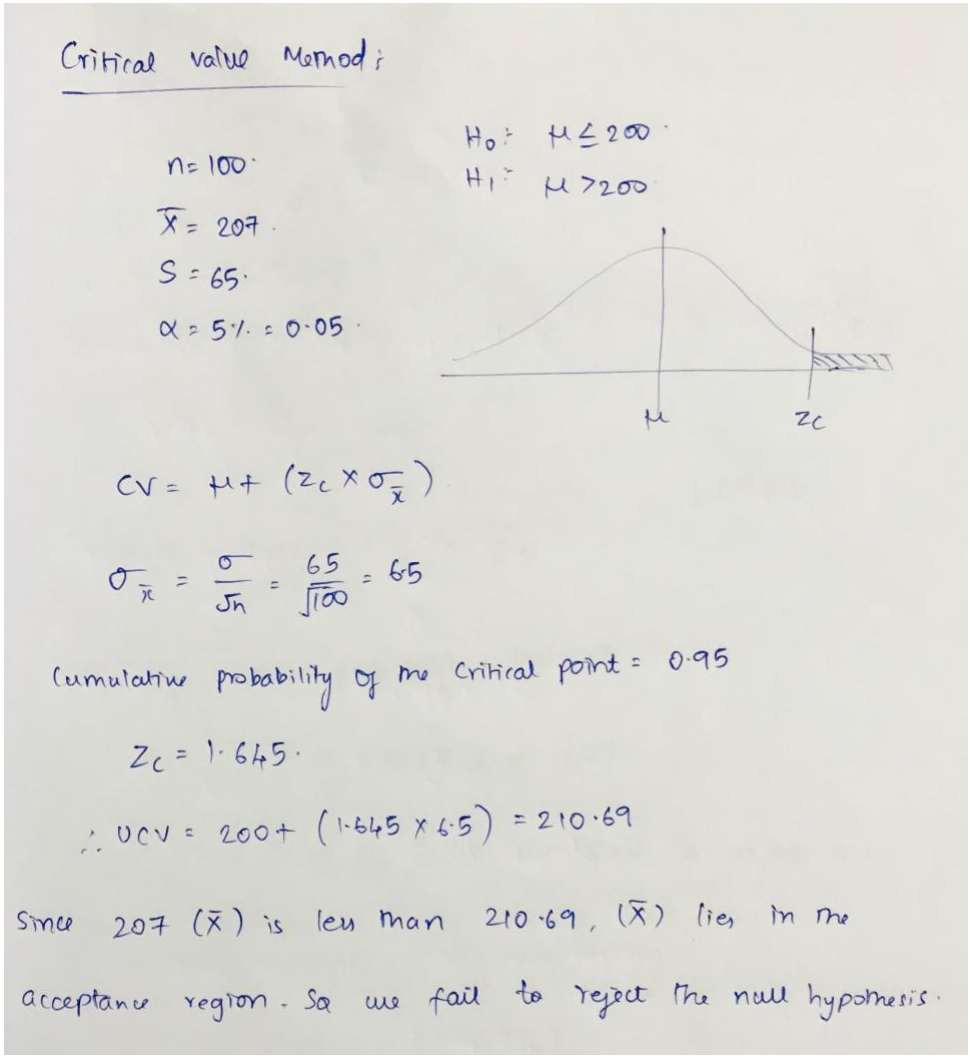
Question 3:

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

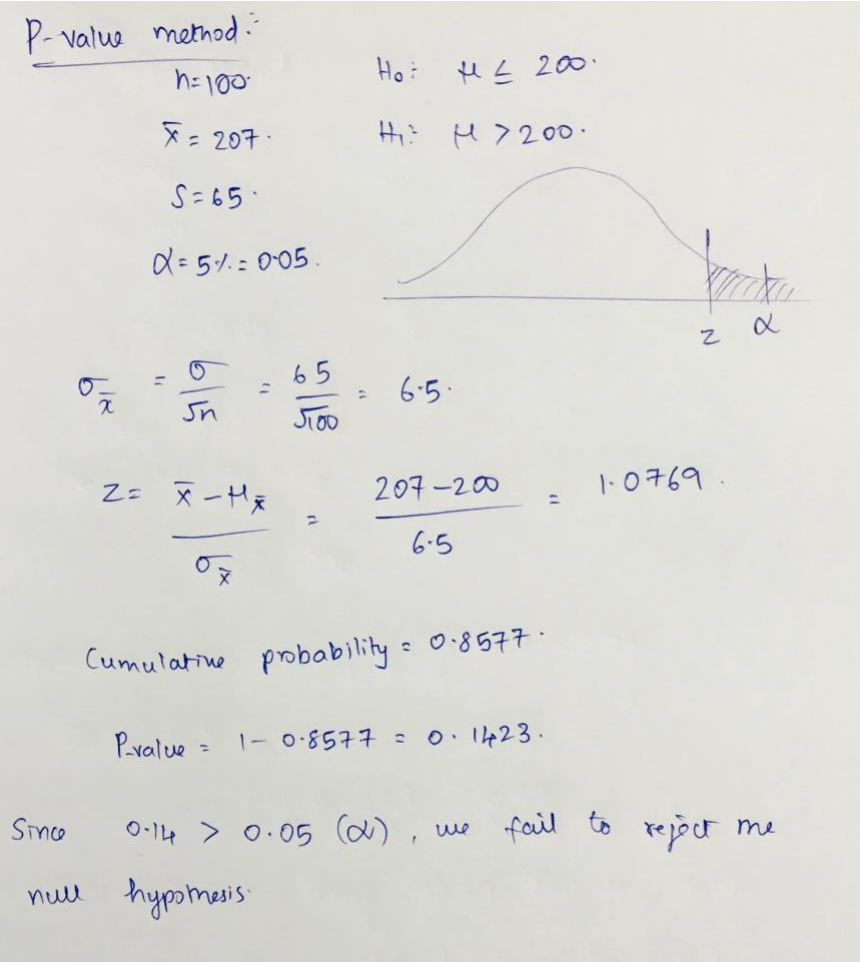
methods to make your decision. Take the significance level at 5 %. Clearly specify the hypotheses, the calculated test statistics, and the final decision that should be made for each method.

Answer:

Critical Value Method:



P-Value Method:



From the above two methods, we fail to reject the null hypothesis.

---

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 hypothesis test conditions (sample size, mean, and standard deviation), the value of α and β come out to 0.05 and 0.45 respectively.

Now, a different sampling procedure 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.

Answer:

μ : Mean time of Effect
H_0: μ<=200 (Job Satisfactory)
H_1: μ > 200 (Job Unsatisfactory)

α : H_0 is true. But we rejected it.
β : H_0 is false. But we failed to reject it.

Let the Time of Effect be the job.

α: 0.05 and β: 0.45
Type II error β means that the job was unsatisfactory, but we failed to reject this claim. So, in this case, setting the β value (0.45) will favour the alternate hypothesis H_1.

α: 0.15 and β: 0.15
In the Second case, both α and β are equal. In this case, it will favour the null hypothesis H_0.

You can decrease your risk of committing a type II error by ensuring your sample size is large enough to detect a practical difference.

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 for its existing 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 provides a way for you to test two different versions of the same element and see which one performs better.

Steps:

1. We can produce the two different taglines (i.e.) Tagline1 and Tagline2 and see how many made the purchase in each version by two populations P1 and P2.
2. We can assign values to purchase and Non-purchase as 1 and 0 correspondingly.
3. Then using two-sample proportion we have to compare the proportions of two different samples. We have to find the P-value and the significance level.
4. We can make a decision on a test based on the p-value and the significance level of the test.
5. If the P-value is greater than the significance level, we fail to reject the null hypothesis that the first tagline is better or as good as the second tagline.
6. We can wait for a longer period to collect more data, to be statically sure whether to accept or reject the null hypothesis.