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.

Ans
As the effectiveness of the drug is an event which has 2 possible outcome, either the drug is effective or the drug is not effective. It is a **Binomial distribution**.

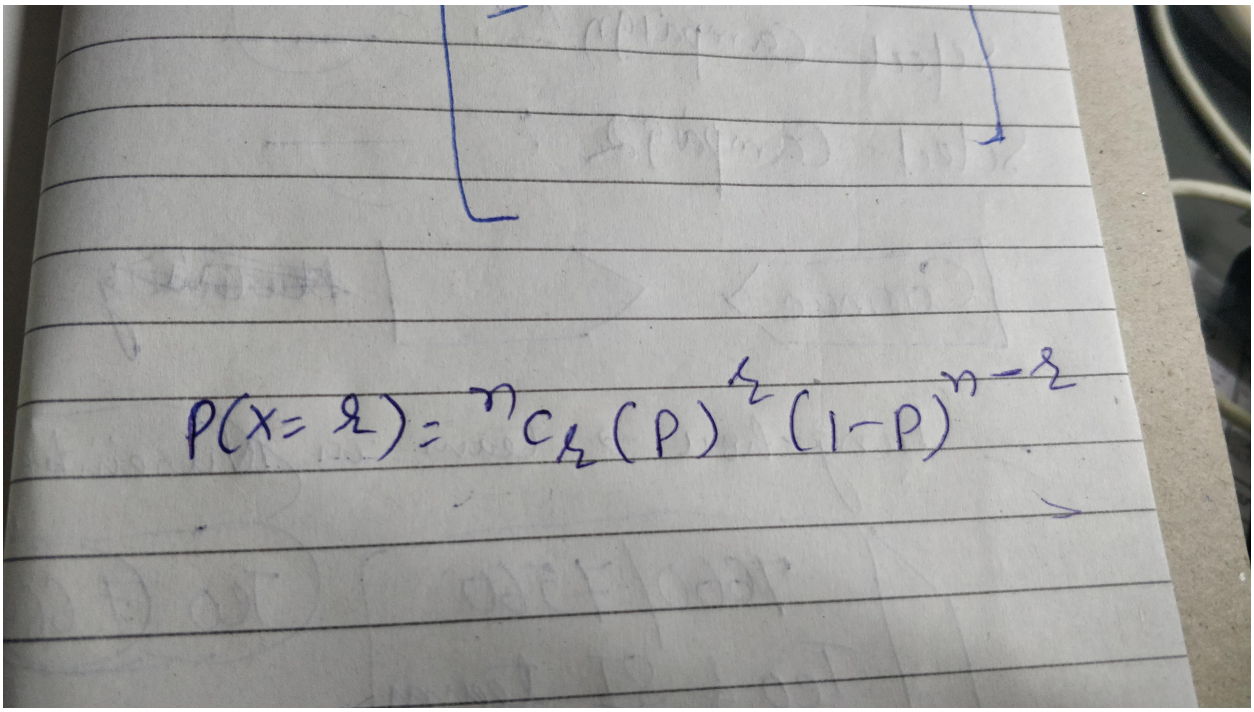
Three Condition of binomial distribution are as follows.

1. The **total** number of trials is **fixed**

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

3. The probability of success is the same for all the trials.

As at the most 3 drugs out of 10 drugs are not able to do a satisfactory job. We need to find the probability. For such a situation, the probability of r successes, is given by –



Where,
n is the total number of trials/questions **p** is the probability of success in 1 trial
r is the number of successes after n trials

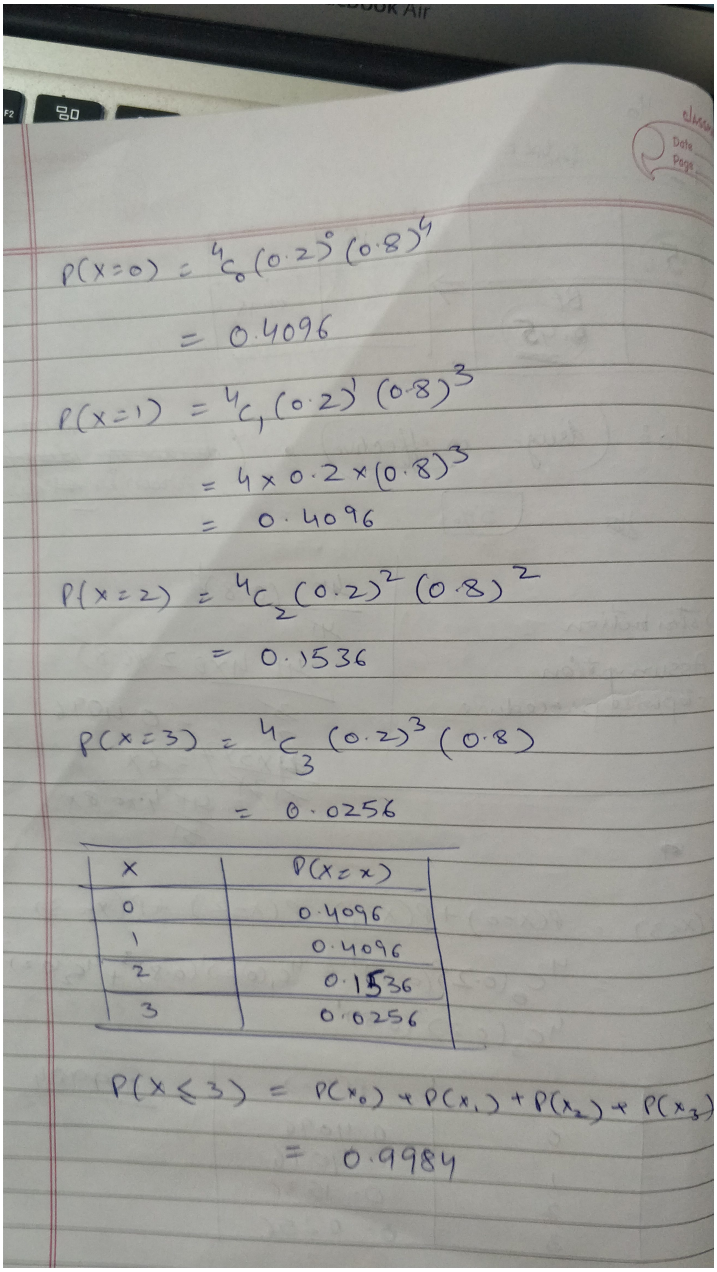
b.) Calculate the required probability.
Ans

As at the most 3 drugs out of 10 drugs are not able to do a satisfactory job We need to find the probability of X<=3.

Sample size (n) = 10
p = probability of a drug to produce the satisfactory result. = 4/5 = 0.2
(1-p) = 1-0.2 = 0.8

P(X<=3) = P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3)

F(X) = (P X<=x) = 0.9984



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.

Ans

A sample of size 100 is taken with sample mean as 207 seconds.
As the population mean is sample mean (+/-) SE(sampling Error).
We can follow the sampling distribution rules OR Central limit theorem rules which are as follows.

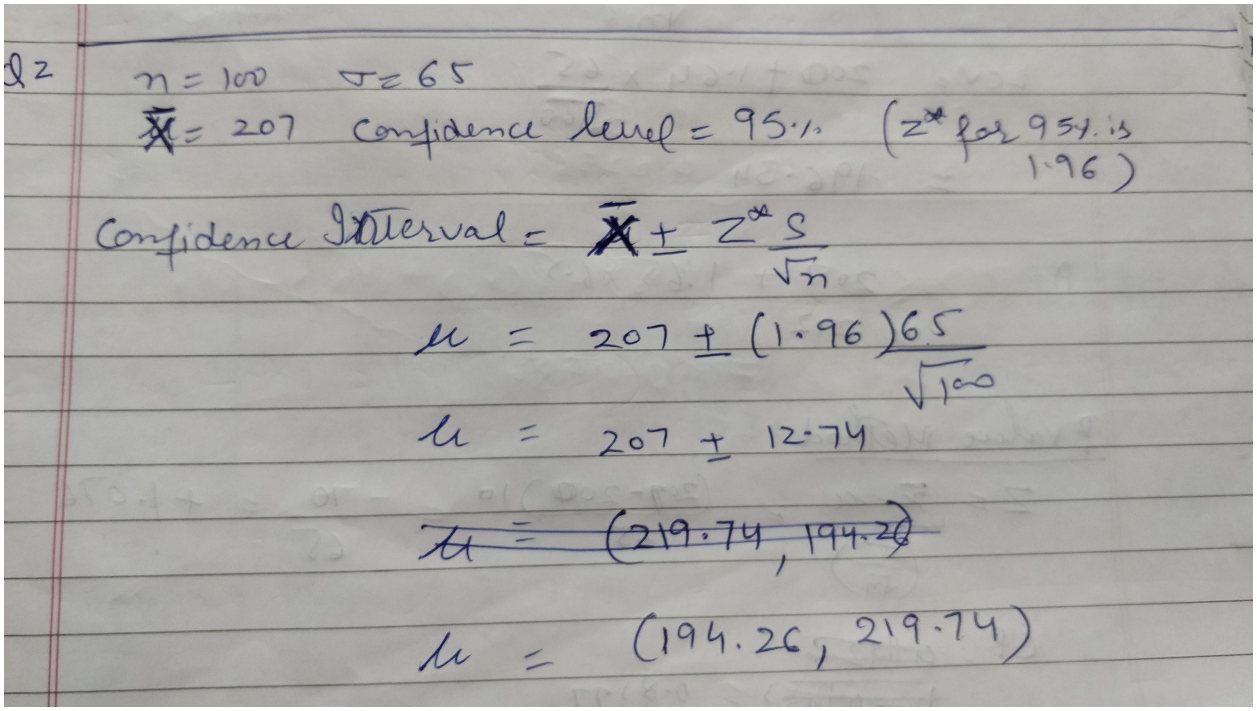
1. Sampling Distribution's Mean (μ^-) = Population Mean (μ) X .

2. Sampling Distribution's Standard Deviation (Standard Error) = $\frac{\sigma}{\sqrt{n}}$, where σ is the population's . s standard deviation and n is the sample size
3. For n > 30, the sampling distribution becomes a normal distribution

Using CLT, you can estimate the population mean from the sample mean and standard deviation.

b.) Find the required range.

Ans range is (194.26, 219.74)

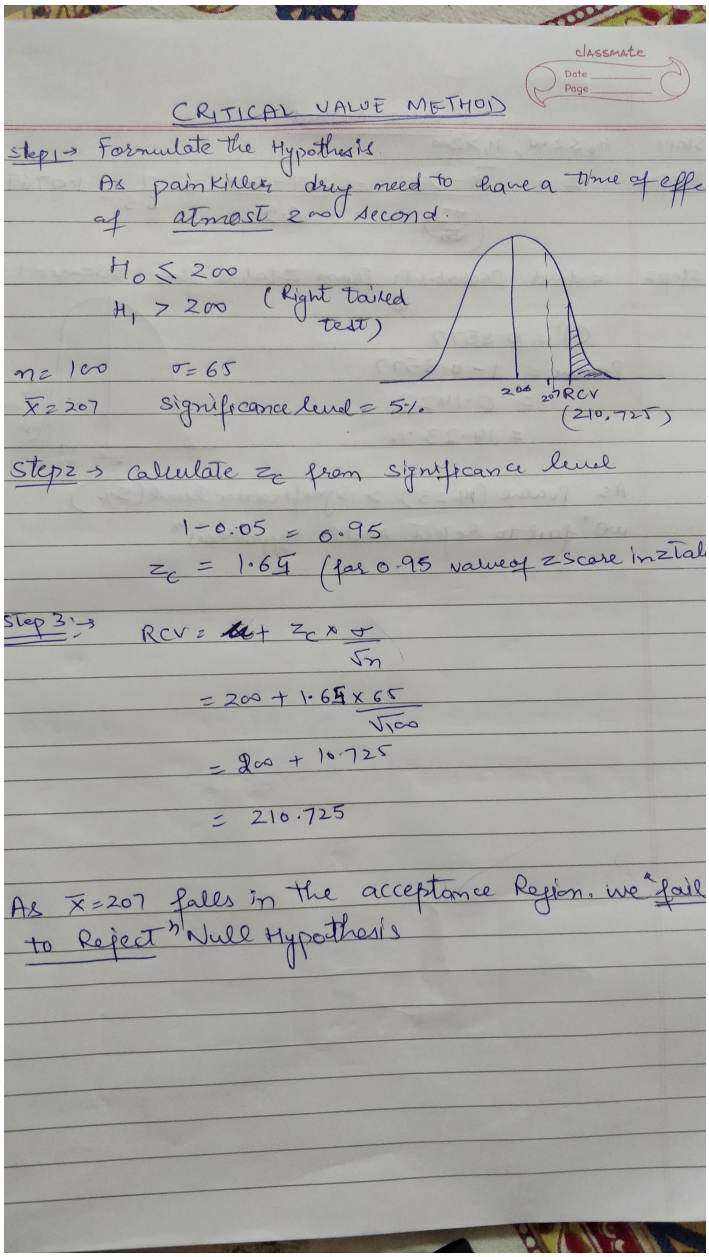


Question 3:

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.

Ans

Critical Value Method : **Fail to reject Null Hypothesis** As per below calculation



CRITICAL VALUE METHOD

Step 1 → Formulate the Hypothesis

As painkiller drug need to have a time of effect of atmost 200 second.

$H_0 \leq 200$

$H_1 > 200$ (Right tailed test)

$n = 100$ $\sigma = 65$

$\bar{x} = 207$ significance level = 5%.

200 207 RCV (210.725)

Step 2 → Calculate z_c from significance level

$1 - 0.05 = 0.95$

$z_c = 1.65$ (for 0.95 value of z score in z table)

Step 3 → $RCV = \mu + z_c \times \frac{\sigma}{\sqrt{n}}$

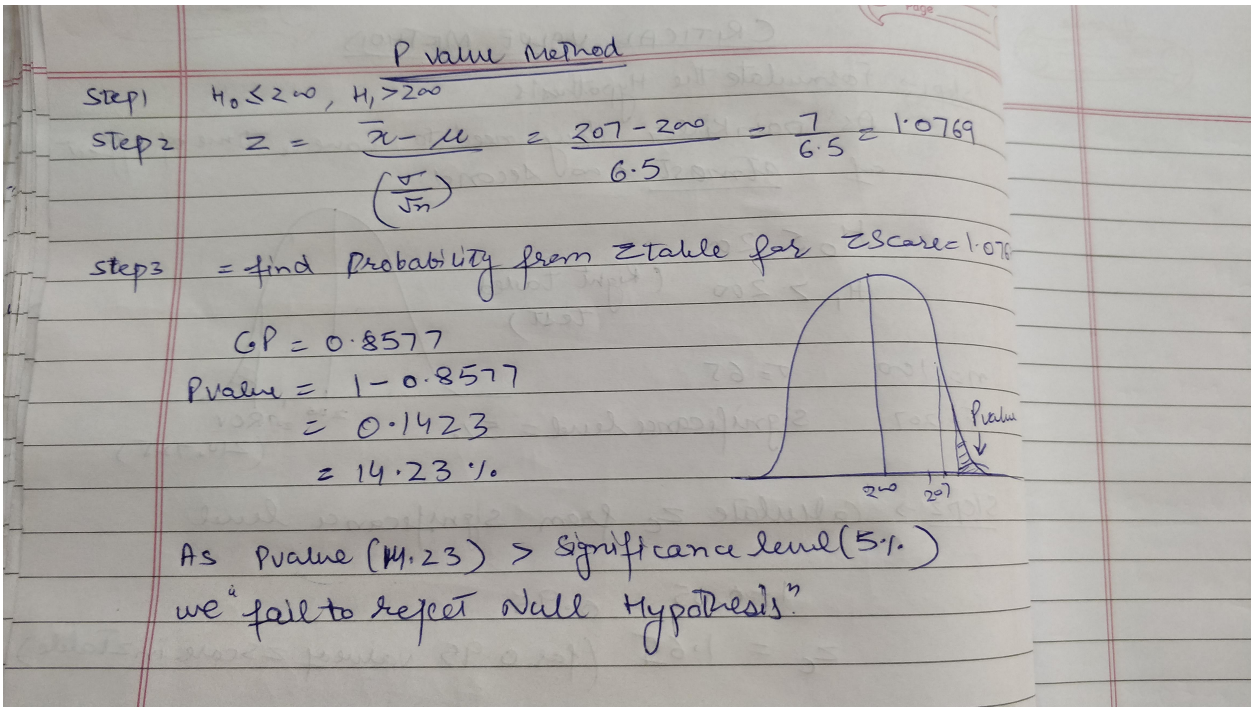
$= 200 + 1.65 \times \frac{65}{\sqrt{100}}$

$= 200 + 10.725$

$= 210.725$

As $\bar{x} = 207$ falls in the acceptance Region, we fail to Reject Null Hypothesis

P Value Method: Fail to reject Null Hypothesis As per below calculation



Conclusion:
painkiller drug have a time of effect of at most 200 seconds is considered to be doing a satisfactory job.

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).

Ans

1) where α and β as 0.05 and 0.45 respectively would be preferred over having them both at 0.15.

Example 1:

Lets say we want to know if the patient has the deadly disease or not **Null Hypothesis** is "Patient is having deadly disease(may be Cancer)".

Type 1 error : Lets say null hypothesis is correct but we reject it that is we make **type 1 error α**, **Means patient has disease but we said he/she do not have disease**. In that case we will not give treatment to the patient even if he/she is effected with Deadly disease, As a result patient may died and we are putting the life of patient at risk. So chances of making such error should be very low. Hence low value of **α** will be preferred here.

Type 2 error : Lets say null hypothesis is incorrect but we fail to reject it that is we make type 2 error β , Means patient does not have disease but we said he/she do have disease. In that case we will give treatment to the patient even if he/she is **NOT** effected with Deadly disease, As a result patient may have taken some extra medicine but life of patient in not at risk. So chances of making such error can be still tolerated. Hence even a high value of β will not have much adverse effect.

Conclusion: Type 1 Error is more critical in this example, so value of **α should be preferred low.**

2) 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.

Example 2:

Lets say we want to know if the Drug is effective or not

Null Hypothesis is "Drug is effective".

Type 1 error : Lets say null hypothesis is correct but we reject it that is we make **type 1 error α**, **Means drug is effective but we say that it is not effective**. In that case we will not make use of the drug, As it is a low risk scenario Hence value of **α** need not be critically low.

Type 2 Error: Lets say null hypothesis is incorrect but we fail to reject it that is we make type 2 error β , **Means drug is not effective but we say that it is effective.** In that case we will be consuming the drug but it has no effect, As it is a low risk scenario Hence value of β need not be critically low. But having too high value of β as 0.45 will not be advisable as drug is not effective it will not cure the patient and patient will be still suffering.

Conclusion: As effectiveness of a drug is low risk scenario and the type 2 error is more significant in this case, Hence high value of type 2 error will not be preferred.

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.

Ans:

As we have 2 tagline and we need to identify which ad is doing better than other. A/B testing is helpful in this case as it is is entirely based on the two-sample proportion test.

Stepwise procedure:
1) Create 2 samples with population size n1 and n2 and lets say x1 and x2 are the people who bought the product after seeing the ad campaign respectively.

Sample 1(Tagline 1) => Population (n1) ==> x1
Sample 2(Tagline2) => Population(n2) ==> x2

2) formulate the null hypothesis. Lets say Tagline one is better then or as good as tagline 2
T1>= T2

3) lets say significance level is 5%, find the P values by using the stats tool in Excel and we can verify if our assumed Null hypothesis is correct or not. Hence, results will indicate which ad campaign will perform better.