Question-1:

The quality assurance checks on the previous batches of drugs found that — it is 4 times more likely that a drug can 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.

b.) Calculate the required probability.

Answer-1:

a.) I would propose **Binomial Probability Distribution** for the above scenario. Reason being, this model is used when there are only two possible outcomes, which is similar to this case.

Conditions for Binomial Probability Distribution

- Total no. of trials is fixed at N
- Each trial is binary, that is has only 2 possible outcomes (Success/Failure)
- Probability of success is same in all trails, denoted by 'p'

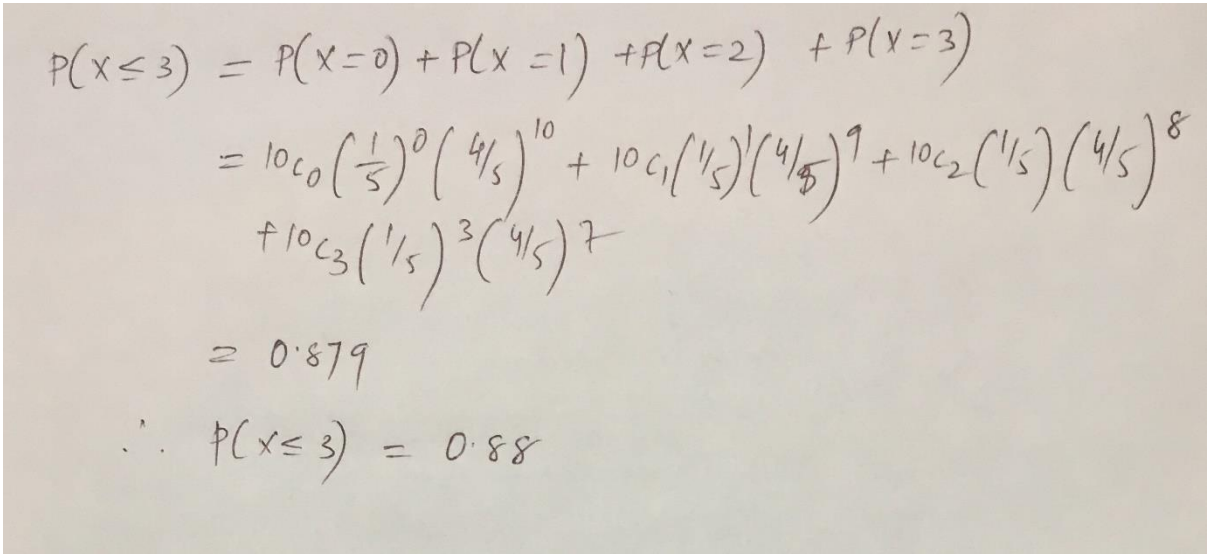
Binomial Distribution Formula

$$P(x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{(n-x)!\,x!} p^x q^{n-x}$$

where
n = the number of trials (or the number being sampled)
x = the number of successes desired
p = probability of getting a success in one trial
$q = 1 - p$ = the probability of getting a failure in one trial

b.) Probability that the drug is NOT satisfactory: p = 1/5
Probability that the drug is satisfactory: q = 1-p = 4/5

Sample size (n) = 10



P(x<=3) = 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.
b.) Find the required range.

Answer-2:

a.) **Sampling distribution** is basically distribution of sample means of population instead of original population.

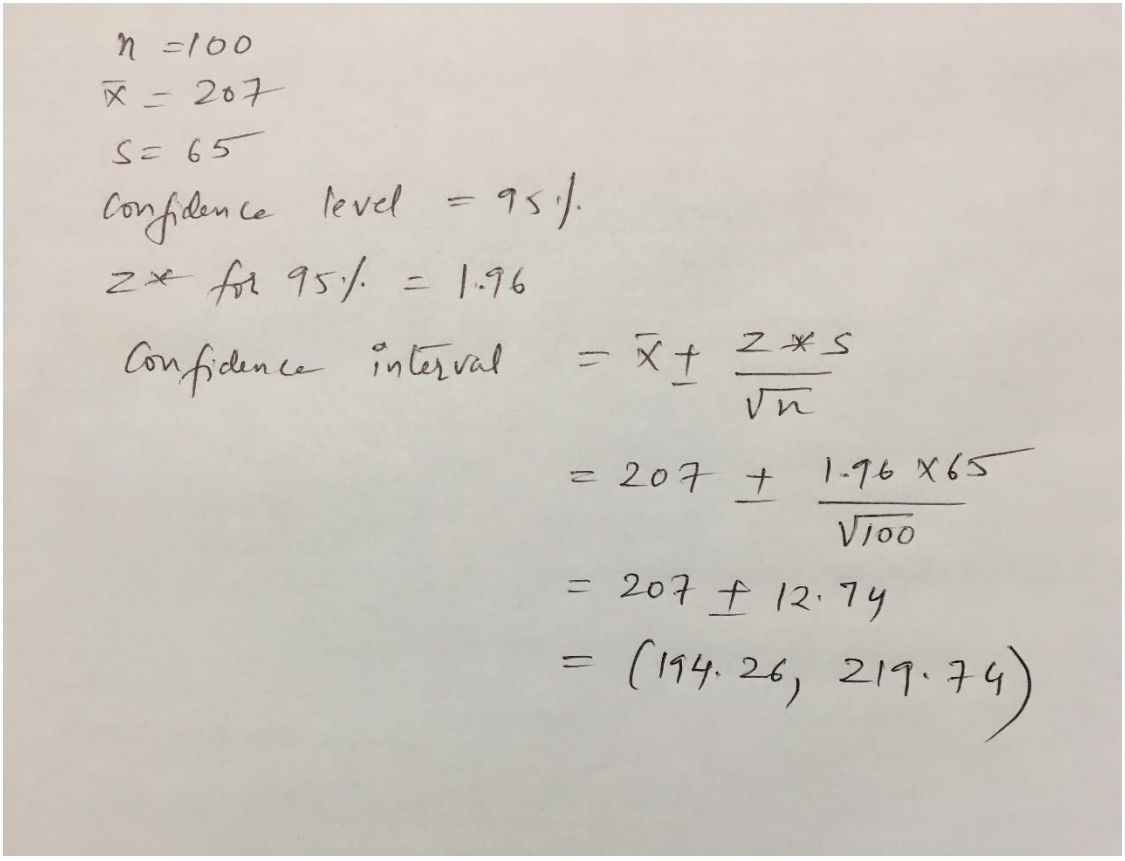
Properties of Sampling distributions

- Sampling distribution's mean ($\mu_{[\bar{x}]}$) = Population mean (μ)
- Sampling distributions standard deviation

Standard error = $\frac{\sigma}{\sqrt{n}}$, where σ (population size) and n (sample size)

- For N>30, the sampling distribution becomes a **Normal Distribution**

b.)



Hence, the range in which the population mean might lie — with a **95%** confidence level 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.

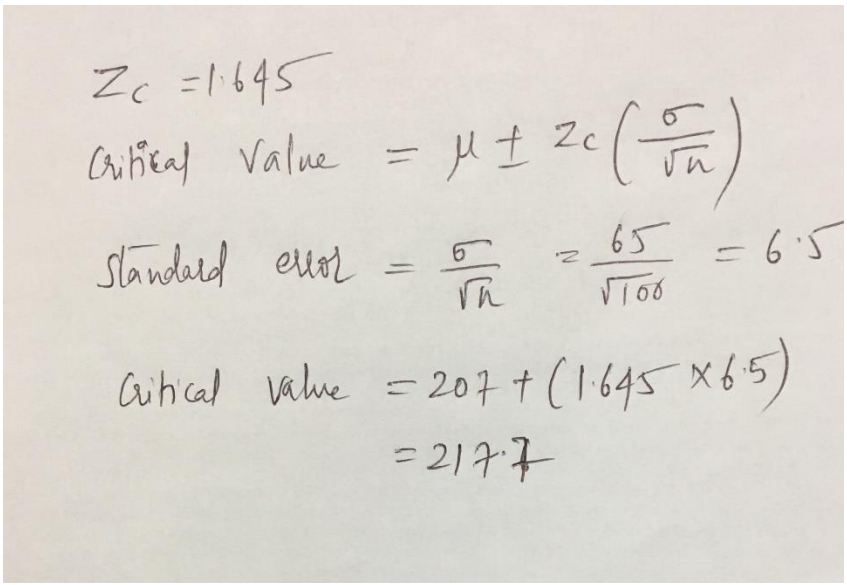
b) You know that two types of errors can occur during hypothesis testing — namely Type-I and Type-II errors — whose probabilities are denoted by α and β respectively. For the current hypothesis test conditions (sample size, mean, and standard deviation), the value of α and β come out to 0.05 and 0.45 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-3:

a.) Null Hypothesis:
H0 = Painkiller drug have a time effect of ≤ 200
Alternate Hypothesis:
H1 = Painkiller drug have a time effect of > 200

sample size = n = 100
sample mean = μ = 207
standard deviation of sample = S = 65
significance level = 5%
cumulative probability of the critical point = 0.95



Since, (217.7 > 207) **We fail to reject the Null hypothesis**

b.) If the consequences of a type I error are serious, then small significance level (α) is ok. Hence, if we consider the first method where we are getting low probability of type I error. It is very ok if we are not treating for severe cases where the pain is huge.

Likewise, If the results of a Type I error are not very serious, then a larger significance level is ok. Whereas for second case we are maintaining α and β at 0.15, it is suitable where type I or type II error is not going to make a significance difference for some less severe pain.

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-4:

A/B testing allows you to test the impact your changes might have on your portal.

A/B Testing comprises of a set of processes that one must follow sequentially to arrive at a **realistic conclusion.**

The following is an A/B testing framework you can use to start running tests:

Data gathering:

- This can be done by looking at the high traffic areas of the site
- Look for pages with low conversion rates or high drop-off rates that can be improved on

Find Objectives:

- Metrics to determine whether the variation is more successful than the original thing
- Number of product purchases and e-mail signups

Generate Hypothesis:

- Once you've found the goals to be accomplished you can begin generating A/B testing ideas and hypotheses for why you think they will be better than the current version.
- Once you have a list of ideas, prioritize them in terms of expected impact and difficulty of implementation.

Run Experiments:

- Start the experiment and wait for visitors to participate
- Then, start measuring how each performs

Analyse Results:

- A/B testing software will present the data from the experiment and show the difference between how the two versions of the page performed, and whether there is a statistically significant difference.