Statistics Assignment

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.

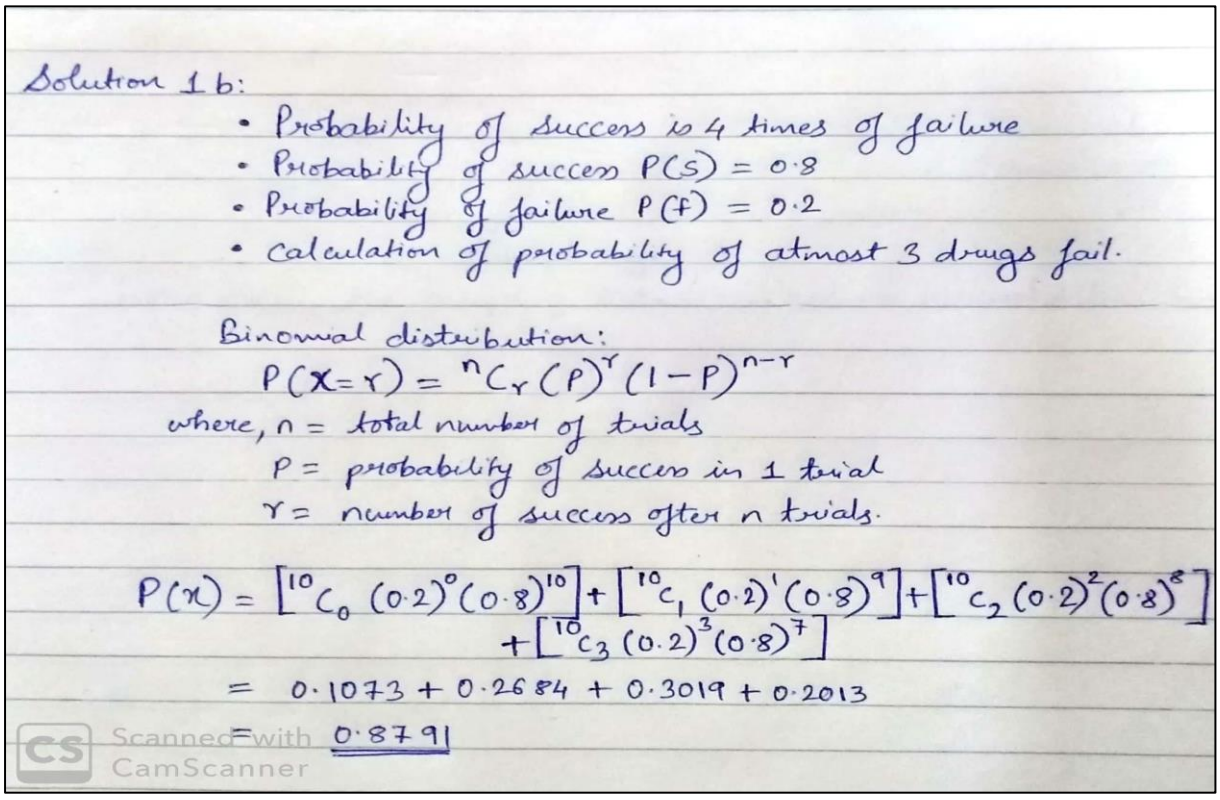
b.) Calculate the required probability.

Solution:

a.) The type of probability distribution used would be **Binominal distribution**. Below are the 3 that this distribution follows:

- Total number of trials is fixed.
- Each trial only has either success or failure as an outcome.
- Probability of success remains same for each trial.

b.) Calculation of the required probability:



Solution 1 b:

- Probability of success is 4 times of failure
- Probability of success $P(S) = 0.8$
- Probability of failure $P(f) = 0.2$
- Calculation of probability of atmost 3 drugs fail.

Binomial distribution:

$$P(X=r) = {}^{n}C_{r}(P)^{r}(1-P)^{n-r}$$

where, n = total number of trials

P = probability of success in 1 trial

r = number of success after n trials.

$$P(x) = \left[{}^{10}C_0\,(0.2)^0(0.8)^{10}\right] + \left[{}^{10}C_1\,(0.2)^1(0.8)^9\right] + \left[{}^{10}C_2\,(0.2)^2(0.8)^8\right] + \left[{}^{10}C_3\,(0.2)^3(0.8)^7\right]$$

$$= 0.1073 + 0.2684 + 0.3019 + 0.2013$$

$$= 0.8791$$

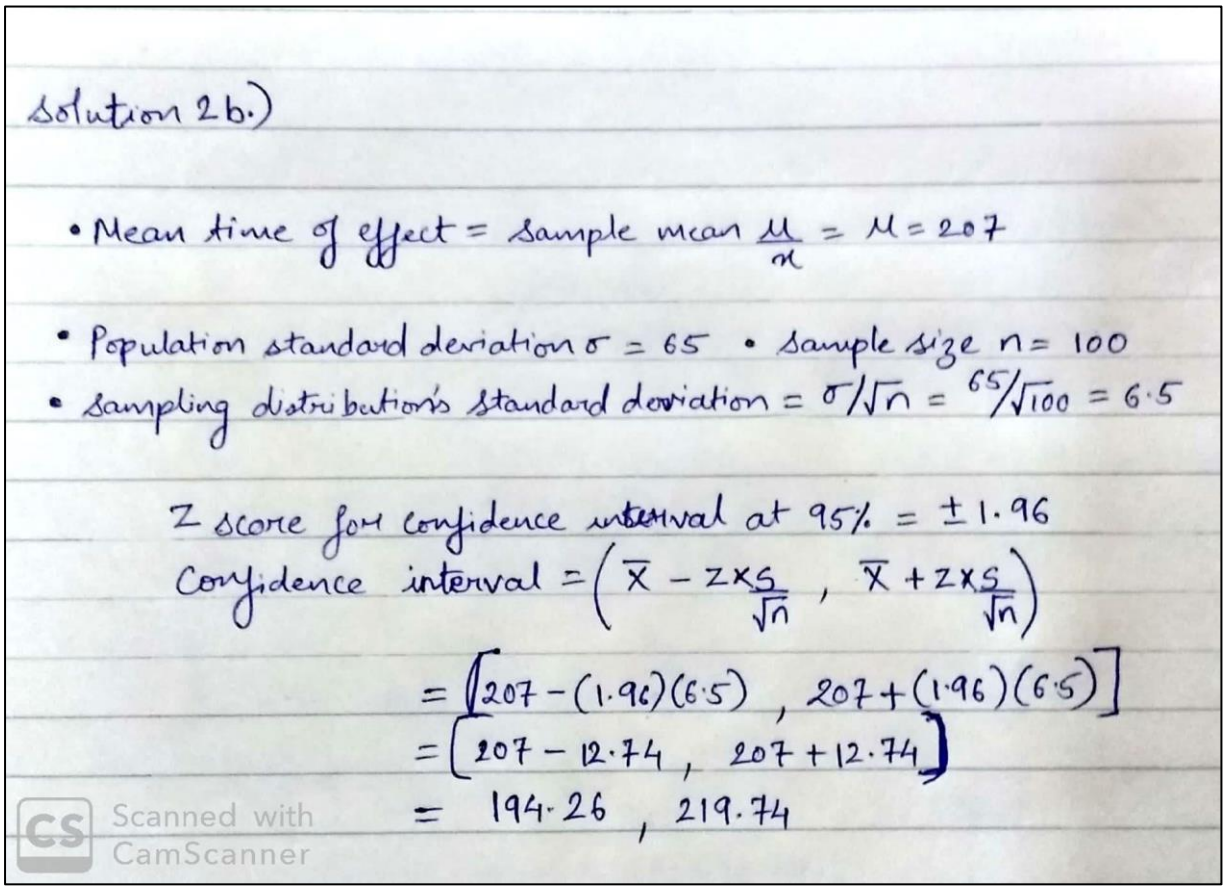
Question 2:
For the effectiveness test, a sample of 100 drugs was taken. The mean time of effect was 207 seconds, with the standard deviation coming to 65 seconds. Using this information, you are required to estimate the range in which the population mean might lie — with a 95% confidence level.

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.

Solution:
a.) The main methodology to approach this problem is **Sampling Distribution,** which is basically the distribution of sample means of a population and below are the properties which are collectively known as **central limit theorem**.

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

b.) Calculation of required range:



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 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).

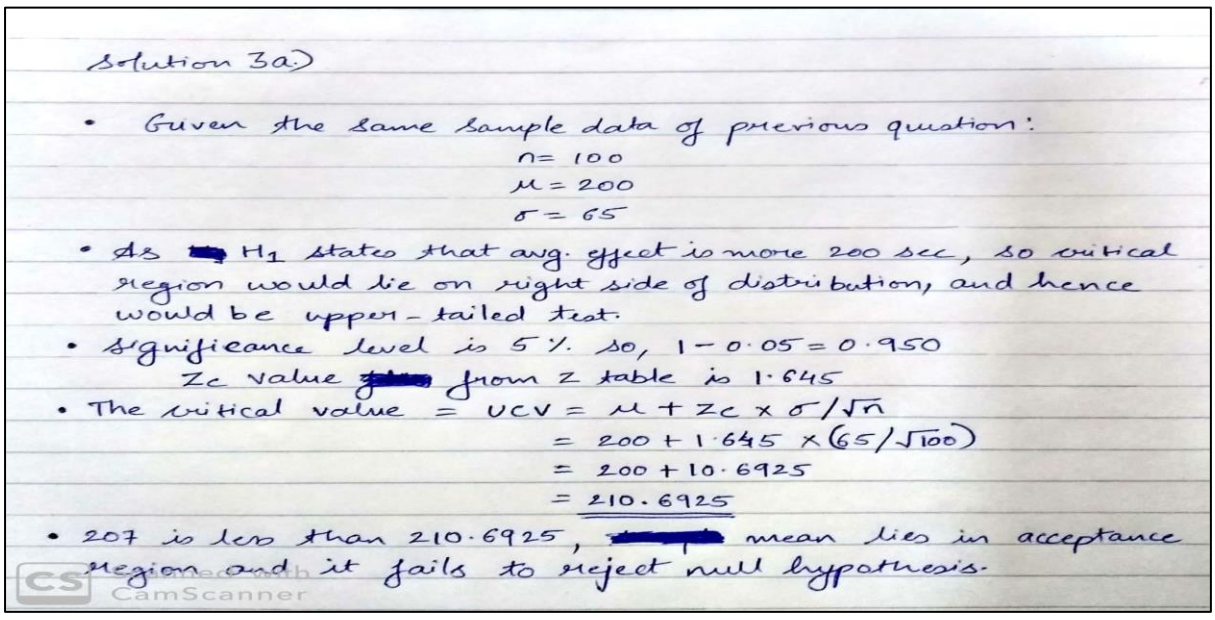
Solution:
a.) Defining null hypothesis(H_0) and Alternate hypothesis(H_1)

- Null Hypothesis: The average time of effect is less than or equal to 200 seconds.

$$\mathbf{H_0: \mu \leq 200}$$

- Alternate Hypothesis: The average time of effect is more than 200 seconds.

$$\mathbf{H_1: \mu > 200}$$


Solution 3a)

- Given the same sample data of previous question:
  $n = 100$
  $\mu = 200$
  $\sigma = 65$
- As [redacted] H_1 states that avg. effect is more 200 sec, so critical region would lie on right side of distribution, and hence would be upper-tailed test.
- Significance level is 5%. so, $1 - 0.05 = 0.950$
  Zc value [redacted] from z table is 1.645
- The critical value = UCV = $\mu + Zc \times \sigma/\sqrt{n}$
  $= 200 + 1.645 \times (65/\sqrt{100})$
  $= 200 + 10.6925$
  $= 210.6925$
- 207 is less than 210.6925, [redacted] mean lies in acceptance region and it fails to reject null hypothesis.

- **Final Decision:** Fail to reject the null hypothesis.

b.) Type I and Type II Error

- The type I error occurs when the **null hypothesis is true but we reject it.**
- The type II error occurs when the **null hypothesis is false but we fail to reject it**.

Scenario 1:

Consider that a weight of the car should not be above 1200kg, which would affect the performance of car.

- **Null Hypothesis:** The weight of the car is less than equal to 1200kg, $\mathbf{H_0: \mu \leq 1200}$
- **Alternative Hypothesis:** The weight of the car is greater than 1200kg, $\mathbf{H_0: \mu > 1200}$
- Here values of α (Rejecting a true null hypothesis) and β (Failing to reject a false null hypothesis) will be very crucial.
- If value of α increases that means the weight of car is less than 1200kg but we reject the fact and consider it to be of much higher weight (we reject $\mathbf{H_0: \mu \leq 1200)}$.
- So, we must prefer α value as 0.05 over 0.15.

Scenario 2:

- Average session duration by user on social media platforms is 15 minutes, a digital marketer needs to plan a budget to spend on social media ads accordingly.
- H_0 = Average session duration by user is 15 minutes
- H_1 = Average session duration by user is not 15 minutes
- In this case failing to reject a false H_0 will cost, a digital marketer by spending for ads in wrong channels.
- So here it's best to prefer β value as 0.15 rather than 0.45.

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:

A/B testing is a marketing experiment wherein we "split" the audience to test several variations of a campaign and decide which performs better. A/B testing can help in taking firm marketing decisions rather than just an intuition. A/B testing serves valuable to a business as it is low investment with higher ROI. It is helpful to increase website traffic, higher conversion rate, reduction on bounce rate and lower cart abandonment and many more.

Stepwise procedure for the A/B test:

- **Research:** It is very important to study data before conducting A/B test. Use tools like Google Analytics, Adobe Analytics, etc. to understand the behaviour of user, identify the pain areas and then conduct A/B test.
- **Data Collection:** Gather the data from various analytics tool, CRM, etc. identify the less traffic or conversion driving channels, major drop off in a journey of purchasing a product or ultimate conversion.
- **Goal Setting:** Based on above two step setup a firm business goal. A goal can help identify the metrics to decide whether a new version is more successful than its original version or not.
- **Create Hypothesis:** After setting goals we can start creating ideas and hypothesis about which version is better. Set an order of those ideas to be implemented based on business goals.
- **Create Variations:** Create multiple variations using A/B testing tools available in market. A variation can be like changing form fields, adding images or colours, etc.
- **Experimentation:** Publish all the variation for users and monitor their actions for multiple variations and compare the results for same.
- **Result Analysis:** Understand and analyse the performance for the two versions for which test was conducted from A/B testing software and identify if statistical significant difference exist or not.