Statistics and Hypothesis Testing 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.*

Here the probability of the event has only two possible outcome i.e.

i. Able to do satisfactory result
ii. Not able to do satisfactory result.

So these type of scenarios are captured via **binomial distribution**. The conditions of binomial distribution are as follows:

a. The total number of trials is fixed
b. Each trial is binary, i.e. has only two possible outcomes, success and failure
c. The probability of success is the same for all the trials.

b) Calculate the required probability. (We have to find probability that at most, 3 drugs are not able to do a satisfactory job)

This can be calculated by finding cumulative probability of x, denoted by F(x).

Q1.b

Here we have to find $F(3)$

$$F(3) = P(X \leq 3)$$
$$= P(X=0) + P(X=1) + P(X=2) + P(X=3)$$

And it is given that $4P(NS) = P(S)$

where $P(S)$ = probability of satisfactory result.
$P(NS)$ = probability of not satisfactory result.

we know that, $P(S) + P(NS) = 1$

$$\therefore 4P(NS) + P(NS) = 1$$
$$5P(NS) = 1$$
$$P(NS) = \frac{1}{5} = 0.2$$
$$\therefore P(S) = \frac{4}{5} = 0.8$$

And,

$$P(X=r) = {}^nC_r (p)^r (1-p)^{n-r} \quad \& \quad {}^nC_r = \frac{n!}{(r!(n-r)!}$$

Here n = no. of trials = 10
p = probability of success $\Rightarrow$ in our case probability of not ~~doing~~ able to do satisfactory job.
r = no. of successes after n trials.

$$P(X=0) = {}^{10}C_0 (0.2)^0 (1-0.2)^{10-0}$$

$\hookrightarrow$ 0.2 because we are finding probability of not able to do satisfactory job which is nothing but our $P(NS)$.

$$P(X=0) = {}^{10}C_0 (0.2)^0 (1-0.2)^{10-0}$$

$$= 1 \times 1 \times 0.10737374$$

$$= 0.10737374$$

$$P(X=1) = {}^{10}C_1 (0.2)^1 \times (1-0.2)^{10-1}$$

$$= 10 \times 0.2 \times 0.1342217$$

$$= 0.268435$$

$$P(X=2) = {}^{10}C_2 (0.2)^2 \times (1-0.2)^{10-2}$$

$$= 45 \times 0.04 \times 0.167777$$

$$= 0.301989$$

$$P(X=3) = {}^{10}C_3 (0.2)^3 \times (1-0.2)^{10-3}$$

$$= 120 \times 0.008 \times 0.2097151$$

$$= 0.201326$$

$$\therefore F(3) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$$

$$= 0.10737374 + 0.268435 + 0.301989 + 0.201326$$

$$= 0.8791$$

$$\therefore F(3) = 0.8791$$

i.e probability that almost 3 drugs in a sample of 10 drugs are not able to do a satisfactory job is 0.8791 or 87.91%.

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.*

Here our aim is to find the range in which the population mean lie.

As we know **population mean = sample mean ± some margin of error**.

In the above statement we know about sample mean (207 seconds) but we don't know about margin of error.

With the help of **central limit theorem**, we will be able to find the margin of error. Because central limit theorem states that no matter how the original population is distributed, the sampling distribution will follow these three properties:

i. Sampling distribution's mean ($\mu_{\bar{X}}$) = Population mean (μ), which is 200 seconds in our case.
ii. Sampling distribution's standard deviation (Standard error) = $\frac{\sigma}{\sqrt{n}}$
where σ is the population's standard deviation and n is the sample size. n = 100 in our case.
iii. For n > 30, the sampling distribution becomes a normal distribution

Given these above three condition, in our case the sample size n = 100, which is greater than 30. This mean it lets us assume that the sample mean would be normally distributed. So we assume standard deviation $\boldsymbol{\sigma \approx S}$. Where 'S' is standard deviation of the sample which is 65 seconds in our case.

So, margin of error is given as $\pm\frac{Z^{*}S}{\sqrt{n}}$

Where Z* is the Z-score associated with a y% confidence level.

Therefore, the confidence interval for the population mean μ or the range of the population mean is given by **population mean = sample mean ± some margin of error**.

i.e confidence interval = $\left(\bar{X}-\frac{Z^{*}S}{\sqrt{n}}, \bar{X}+\frac{Z^{*}S}{\sqrt{n}}\right)$

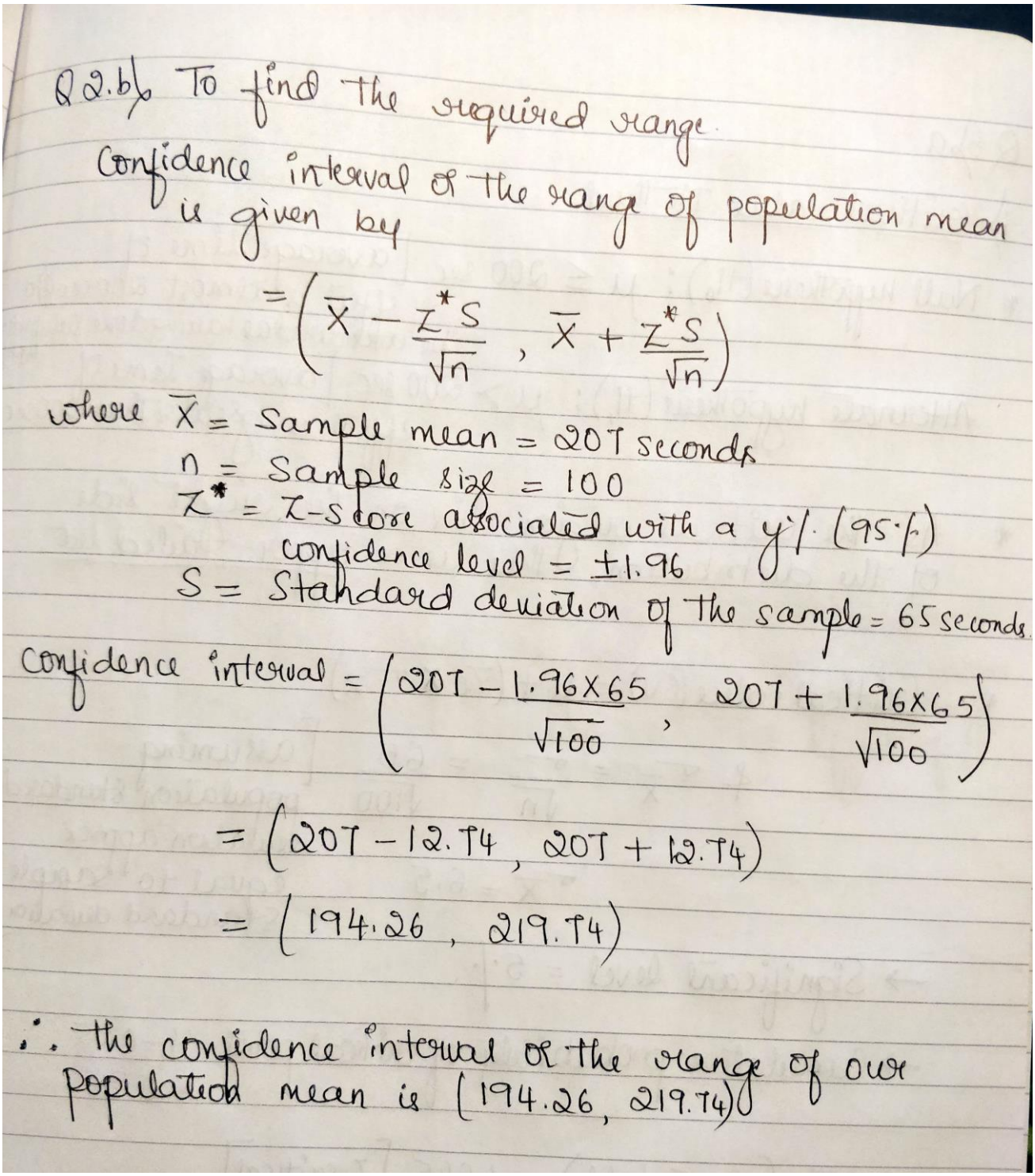
Where $\bar{X}$ is sample mean = 207 seconds.

S is standard deviation of the sample= 65 seconds

n is sample size = 100

Z* is the Z-score associated with a y% confidence level. y % = 95%

b) *Find the 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.*

1. Critical value method:

Q 3 a,b

i) Critical value method

* Null hypothesis (H_0): $\mu \leq 200$ sec [average time of effect of drug is less than or equal to 200 sec for it to be considered as effective doing satisfactory job]

* Alternate hypothesis (H_1): $\mu > 200$ sec [average time of effect of drug is greater than 200 sec for it to be considered as doing satisfactory job].

Alternate hypothesis (H_1):

* As the critical region lies on the right side of the distribution, this is a upper-tailed test.

* Critical value (CV) $= \mu + (Z_c * \sigma_{\bar{x}})$

& $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{65}{\sqrt{100}}$ [assuming population standard deviation approx. equal to sampling standard deviation]

$\sigma_{\bar{x}} = 6.5$

→ Significant level = 5%

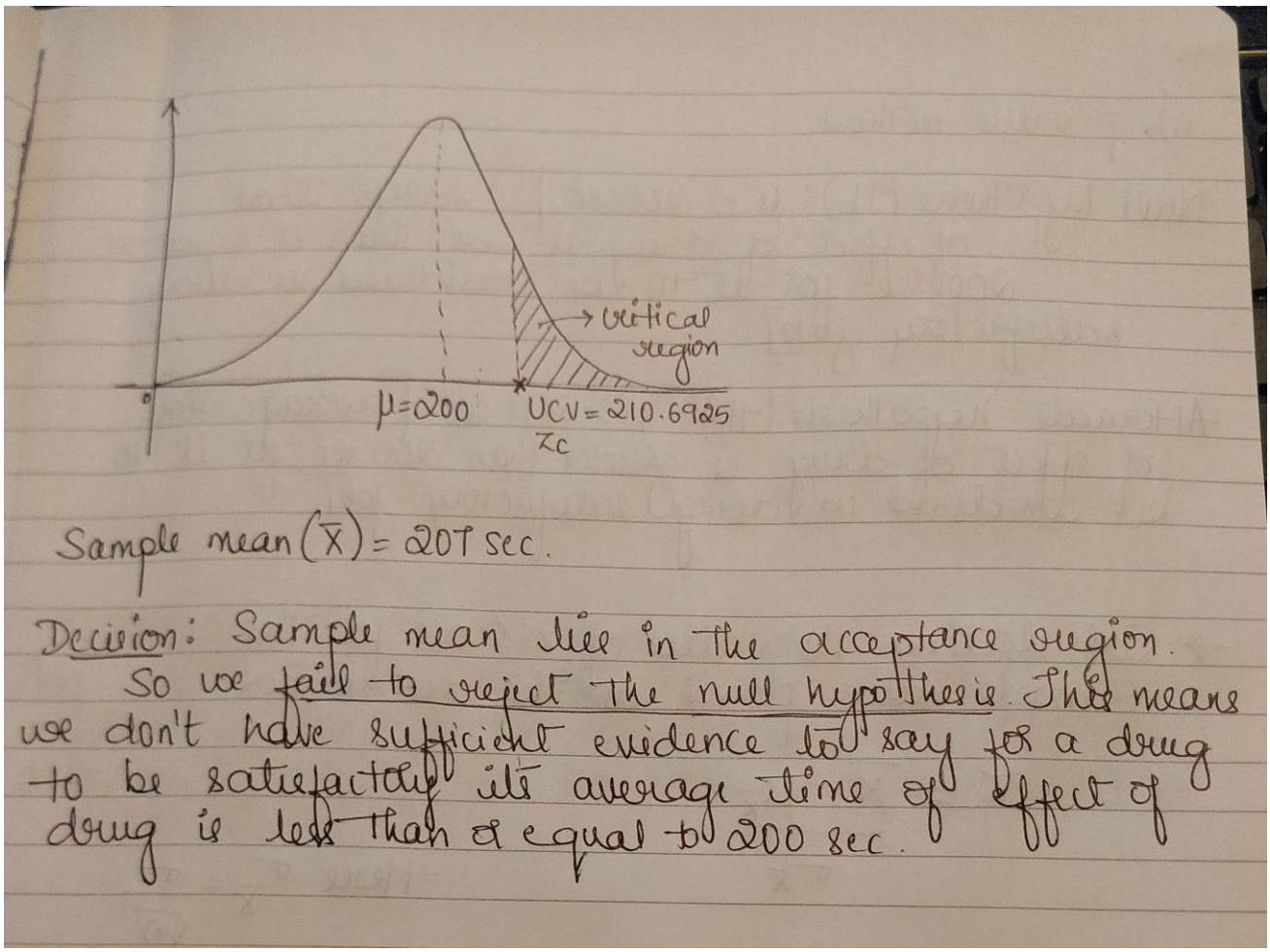
→ Cumulative probability of critical point = 1 − 0.05
= 0.95

→ Z_c (via Z-table) = 1.645 [Z-critical]

CV $= \mu + (Z_c * \sigma_{\bar{x}})$

$= 200 + (1.645 \times 6.5)$

$= 210.6925$ sec



Decision: By critical value method, we found that we fail to reject the null hypothesis. That is we **don't have** sufficient statistical evidence to say that time of effect of the drug (new batch) is less than or equal to 200 seconds for it consider as doing satisfactory job.

2. p- value method:

ii) p-value method

Null hypothesis (H_0): $\mu \leq 200$ sec [average time of effect of drug is less than or equal to 200 sec for it to be considered as doing satisfactory job]

Alternate hypothesis (H_1): $\mu > 200$ sec. [average time of effect of drug is greater than 200 sec for it to be considered as doing satisfactory job].

→ Calculating the value of z-score for the sample mean point on the distribution.

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

Here $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

$= \frac{65}{\sqrt{100}}$

$= 6.5$

$= \frac{207 - 200}{6.5}$

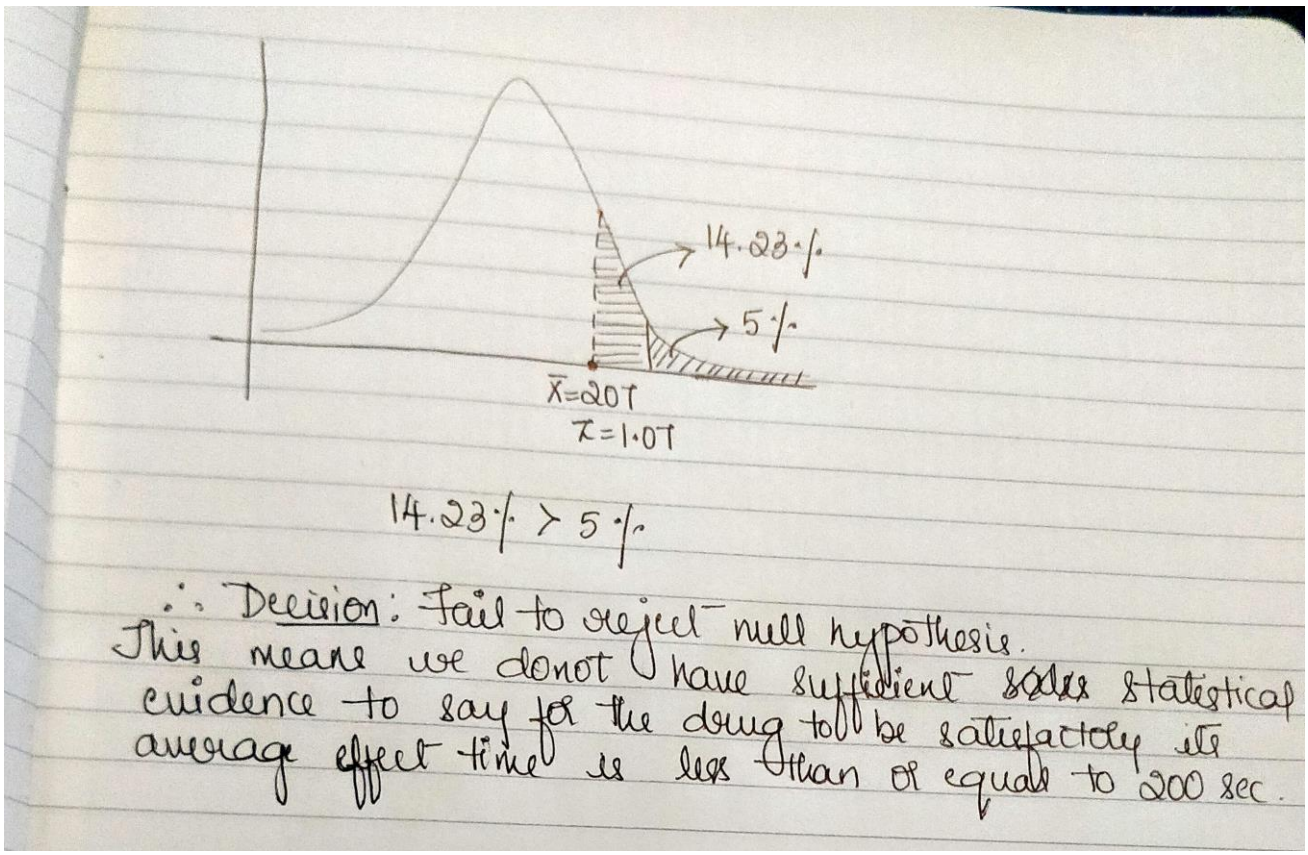
$= 1.0769$

→ Cumulative probability of sample point (via z-table) is 0.85777

∴ $P = 1 - 0.85777$ [one-tailed test]

$P = 0.1423$ or 14.23%.

$14.23\% > 5\%$



Decision: By p value method we got the same result as critical value method. We found that we fail to reject the null hypothesis. That is we **don't have** sufficient statistical evidence to say that time of effect of the drug (new batch) is less than or equal to 200 seconds for it consider as doing 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).

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

Scenario: Consider a scenario where the production cost of new batch is significantly lower than the old batch for the same type of drug .And we are testing our quality assurance based on time of effect of the drug. Lesser the time of effect of the drug higher is the chance of passing the quality assurance test as it is doing satisfactory job.

Hypothesis: Sticking to our same hypothesis but for new batch. Which is:

Null hypothesis: $\boldsymbol{Ho}$: $\mu \leq \mathbf{200}\ \boldsymbol{sec}$ – time of effect of the drug is less than or equal to 200 sec
Alternate hypothesis: $\boldsymbol{H1}$: $\mu > \mathbf{200}\ \boldsymbol{sec}$ - time of effect of the drug is greater than 200 sec.

Reason for the choice: In this condition **α is less than β**. That means we are trying to decrease the % of type 1 error.

Imagine, if we had large % of type 1(high value of α) for the given scenario (for new batch) then it would be - Rejecting the true null hypothesis that is -though time of effect of the drug was less than or equal to 200 sec, it has been rejected. This means the new batch failed the assurance test. This in turn means though the production cost of new batch is lower than the old batch and though time of effect of new batch is less than or equal to 200(doing satisfactory job), it is being rejected. The consequence will be- we are going with old batch of drug where the production cost is high.

In such a scenario, the producer is at risk as it is costing them more money in spite of having same effectiveness.

Therefore, in these type of scenario, we want to reduce the producer risk by keeping α less than β.

α and β values fixed at 0.15 would be preferred over having them at 0.05 and 0.45 respectively

Scenario: Consider the scenario where that painkiller drug which has been tested for the quality assurance in this hypothesis is to be used for **immediate pain relief** i.e. in case of excruciating pains/severe conditions. Lesser the time of effect of the drug higher is the chance of passing the quality assurance test as it is doing satisfactory job.

Ex: during a severe accident, if patient is suffering from kidney stones etc. If the painkiller fails to give immediate relief, it might lead the patient into trauma.

Hypothesis: Sticking to our same hypothesis, which is:

Null hypothesis: $\boldsymbol{Ho}$: $\mu \leq \mathbf{200}\ \boldsymbol{sec}$ – time of effect of the drug is less than or equal to 200 sec
Alternate hypothesis: $\boldsymbol{H1}$: $\mu > \mathbf{200}\ \boldsymbol{sec}$ - time of effect of the drug is greater than 200 sec.

Reason for the choice: In this condition **β is lesser than previous β (0.45 to 0.15)**. That means we are trying to decrease the % of type 2 error.

Imagine we had high % of type 2 error (high value of β), then it would lead to – failing to reject false null hypothesis. That is though the time of effect of the drug is not less than or equal to 200 sec, we are not rejecting. In other words, the drug was supposed to not pass the quality assurance test as it is not doing satisfactory job but still it passed! The consequence will be, the consumer (patient) is at risk- that is though the patient has taken the drug, the drug is not effective. This in turn might lead the patient into trauma.

Other way of seeing the scenario is: By reducing the β value, we are potentially increasing the α value (0.05 to 0.15) as both are inversely proportional. Increase in α value is nothing but increase in % of type 1 error i.e rejecting the true hypothesis. In this scenario, even if we

commit type1 error -though time of effect of the drug was less than or equal to 200 sec, it is been rejected, this will **not** negatively impact the consumer.

Therefore, in these type of scenario, we want to reduce the consumer risk by reducing the β value.

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.

Why A/B testing is effective:

This testing method will compare between two versions of the same element. This will help us determine which version is performing better. In our case the two versions are the two taglines for the online ad campaign. By conducting A/B testing we will know which tagline will be effective to attract new consumers. This will help us - to carefully make changes for the better user experiences.

Few of the benefits of A/B testing are:

- Improved engagement.
- Lot of money is invested in marketing, we definitely don't want negative surprises after the actually launch or we can say-we want to reduce bounce rates.
- Increased sales or in our case increase in the customers.
- Major revisions to the site can cost huge amount of money. A/B testing helps us foresee it - before committing to major decisions and help us increase chances of success.

How A/B testing works:

In an A/B test, we take a webpage (where the ad is published) and modify it to create a second version of the same page with different tagline i.e two version of the same element is created. Then, half of our traffic (visitors) is shown the one version of the page (say first tagline) and other half are shown the different version of the page (say second tagline).

The visitor's engagement is measured and collected in an analytics dashboard and analysed through a statistical engine. We can then determine which tagline had a positive, negative, or no effect on visitor behaviour.

Steps of A/B test:

- **Determining our goals:** Before the test is conducted we should know what we are hoping to accomplish. Sometimes goals can be clear by **collecting the data** to check for low conversion rates or high drop-off rates or things to be improved.
- **Generating Hypothesis:** Once we've identified a goal we will begin to generating A/B testing ideas and hypothesis.
- Making the changes on the page as per our testing scenario.

- Start the experiment and wait for visitors to participate. Their interaction is measured and counted.
- **Analysing the result:** Once the experiment is over, we can use tools to determine which of the either version performed better or was there no impact.