Sun Pharma Case Study Assignment

Submitted By: Kunal Kakara

Problem Statement

Comprehension

The pharmaceutical company Sun Pharma is manufacturing a new batch of painkiller medications, 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 medicine to show effects) as well as quality assurance (which tells you whether the medicine was able to do a satisfactory job).

Question 1:

Quality assurance checks on the previous batches of medications found that it is four times more likely that a medicine is able to produce a satisfactory result than not.

Given a small sample of 10 medicines, you are required to find the theoretical probability that, at most, 3 medicines are unable to do a satisfactory job:

1. **Propose the type of probability distribution that would accurately portray the above-mentioned scenario and list out the three conditions that this distribution follows.**
2. **Calculate the required probability.**

Solution 1:

In the given scenario, we need to calculate the discrete theoretical probabilities of at most 3 medicines being unable to do a satisfactory job. We are presented with a fixed sample size and there are only two possible outcomes, with each outcome having the same probability throughout. Therefore, a **Binomial Distribution** would accurately portray the given scenario.

Three conditions of a Binomial Distribution are as follows:

- The number of trails, denoted by **n**, should be fixed. Here **n=10.**
- Each of the trials must have only two outcomes: success or failure. Here, success is when medicines are unable to do a satisfactory job and failure is when medicines are able to do a satisfactory job.

- The probability of success (or failure) must be same in all the cases.

The formula for finding the Binomial Probability 'P(X=r)' is given by:

$P(X=r)= {}^{n}C_r\ p^r\ q^{n-r}$

where 'n' is the total number of outcomes, 'p' is the probability of success, 'q' is the probability of failure and 'r' is the number of successes after n trials.

Note: Here, success refers to those medicines that are unable to do a satisfactory job and vice-versa.

Now, according to the data given in the question,

$p = 1/5 = 0.20$

$q = 4/5 = 0.80$

$n = 10$

$r = 0,1,2,3$ (because at most 3 means all values less than or equal to 3)

Using the Formula for finding the Binomial Probability, we have:

$\sum_{r=0}^{r=3} P(X=r)= {}^{n}C_r\ p^r\ q^{n-r}$

i.e. $P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$

On putting the values in the formula, we have :

$P(X\leq 3) = 0.879$

Therefore, the theoretical probability that, at most, 3 medicines are unable to do a satisfactory job is **0.879**

Question 2:

For the effectiveness test, a sample of 100 medicines 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 interval in which the population mean might lie – with a 95% confidence level:

1. **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.**
2. **Find the required interval.**

Solution 2:

We will use the **"Central Limit Theorem"** to solve this problem.

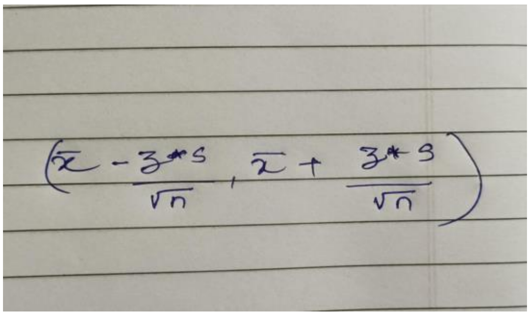
The properties of the Central Limit Theorem are as follows:

- Sampling Distribution's Mean (μ_X) = Population mean (μ),
- Sampling distribution's standard deviation (Standard Error) = $\sigma/\sqrt{n}$, and
- For n > 30, the sampling distribution becomes a Normal Distribution.

We will use the following formula to find the confidence interval for y% confidence level:

Confidence interval

(for y% confidence level) =



where Z* is the Z-score associated with y% confidence level, S is the Sample Standard Deviation, $\bar{X}$ is the Sample Mean and n is the Sample Size.

For 95% confidence level, Z* is ± 1.96.

S = 65 seconds

$\bar{X}$ = 207 seconds

n = 100

Therefore, using the formula for the given set of values, we have

a confidence Interval = 207 ± 12.74

= (194.26, 219.74)

Hence, the population mean might lie in the interval of **(194.26, 219.74)** seconds on a 95% confidence level.

Question 3:

1. **The painkiller 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) as that in the previous question, test the claim that the newer batch produces a satisfactory result and passes the quality assurance test. Utilise two hypothesis testing methods to take a 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.**
2. **You know that two types of errors can occur during hypothesis testing – 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 (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.

Under what conditions would either method be more preferred than the other? Give an example of a situation where conducting the hypothesis test with α and β as 0.05 and 0.45, respectively, would be preferred over conducting the same hypothesis test with α and β at 0.15 each. Similarly, give an example for the reverse scenario, where conducting the same hypothesis test with α and β at 0.15 each would be preferred over having them at 0.05 and 0.45, respectively.

For each example, give suitable reasons for your particular choice using the given values of α and β only. (Assume that no other information is available. Additionally, the hypothesis test that you are conducting is the same as mentioned in the previous question;

you need to test whether the newer batch produces satisfactory results.)

Solution 3:

The two sets of hypotheses formed are as follows:

Null Hypothesis: The painkiller has a time of effect of 200 seconds or less i.e., $\mu \leq 200$ seconds.

Alternate Hypothesis: The painkiller has a time of effect of more than 200 seconds i.e., $\mu > 200$ seconds.

First hypothesis testing method:

We will use the critical value method to make a decision.

We will use the given value of sample data as follows:

S = 65 seconds

μ = 200 seconds

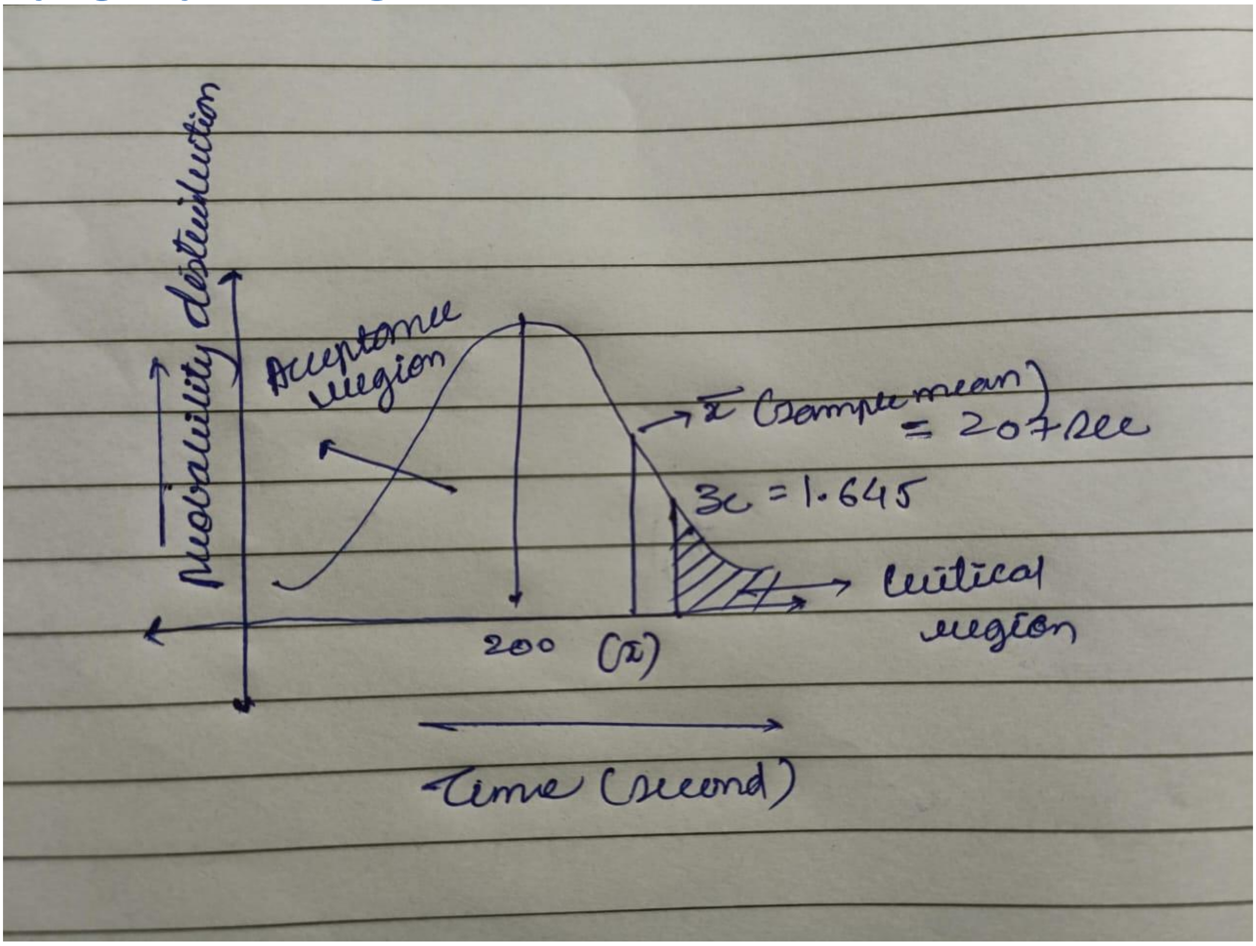
$\bar{X}$ = 207 seconds (Sample Mean)

n = 100

For the given significance value of 5%, we have the cumulative probability of the critical point as 95% or 0.95

Next, we will find the **Z-score** corresponding to 0.95 from the Z-table which will be equal to **1.645**

Since the alternate hypothesis involves a '>' (greater than) sign, it is an example of an upper-tailed test with the critical region lying only on the right-hand side.



Therefore, we will calculate **only the Upper Critical Value (UCV)** from the table using the formula:

Critical Value (CV) = μ + (Z_C * S / $\sqrt{n}$)

where (Z_C * S / $\sqrt{n}$) is also called the standard error, which showsthe maximum deviation allowed with a certain confidence level.

Using the given sample data values, we have:

Critical Value = 210.69 seconds

As a result, the acceptance region corresponds to all the values less than 210.69 seconds.

Now, we see that the upper critical value comes out to be 210.69 seconds whereas the sample mean was 207 seconds.

Since the sample mean value of 207 seconds lies within the acceptance region, **we fail to reject the null hypothesis.**

Hence, we can safely say that the painkiller has a time-effect of at most 200 seconds on a 95% confidence level.

Second hypothesis testing method:

We will use the given value of sample data as follows:

S = 65 seconds (same as population S.D.)

μ (M) = 200 seconds

$\bar{X}$ = 207 seconds (Sample Mean)

n = 100

Significance value = 5% (0.05)

Now, we shall find the Z-value correspondingto

the given sample data by using the formula:

$$z = \frac{x - M}{\frac{\sigma}{\sqrt{n}}}$$

Putting the given values of sample data in the formula, we have:

Z-value = 1.077

Now, we will use the Z-table to find the Z-score (Z_s) correspondingto the given Z-value of 1.08 (approximated), which comes out to be 0.86

Next, we find the corresponding p-value, which is nothing but the probability of the null hypothesis being correct.

So, the p-value is given by:

p-value = 1 - Z_s

Hence, the p-value obtained is 0.14

Upon comparing the p-value with the significant value of 0.05, we observe that the p-value is higher than the significant value. **Therefore, we fail to reject the null hypothesis.**

Again, we can safely say that the painkiller has a time effect of at most 200 seconds on a 95% confidence level.

A Type – I error occurs when you reject a true null hypothesis. It is denoted by **'α'.**

In this case, a **Type - I** error occurs when you conclude that the pain-killer has a time effect of more than 200 seconds when it actually has a time effect of less than or equal to 200 seconds.

Therefore, the Type–I error would suggest that the pain-killer has quality issues in terms of its efficiency even when it works perfectly fine**.** As a result, a greater value of 'α' would heavily cost the company in terms of the resources spent to minimize this error. However, a lower value of 'α' would not have a considerable impact on the company's expenses.

A Type – II error occurs when you fail to reject a false null hypothesis. It is denoted by **'β'**.

In this case, a **Type - II** error occurs when you conclude that the painkiller has a time effect of at most 200 seconds when it actually has a time effect of more than 200 seconds.

Therefore, the Type – II error would suggest that the painkiller works fine even when it is not efficient. As a result, a greater value of ‘β’ would have a serious regulatory issue as the product may prove tobe inefficient even in critical health conditions. However, a lower value of ‘β’ would not have a considerable effect on the patient’s health.

Based on the reasons mentioned above, we will look at the following cases:

Case 1: When the values of α and β come out to be 0.05 and 0.45,respectively.

In this case, the value of α is 0.05 (small) and the value of β is 0.45

Example: A hypothesis test for Case 1 values will be preferred when the cost is very high when rejecting a batch of painkillers.

Case 2: When the values of α and β are controlled at 0.15 each.

In this case, the value of α is 0.15 and the value of β is 0.15 (lower than before).

Example: A hypothesis test for Case 2 values will be preferred when safety is a concern.

Question 4:

Once one 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 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.

Solution:

A/B testing is used to test two different versions of the same product and then compare their performances. Since the marketing team is divided into two groups, A/B testing would be the most suitable in this case as it will help us to decide which option should be used.

The following stepwise procedure can be followed:

1. The two marketing teams will formulate one tagline each. The taglines will differ in terms of keywords and colour formatting used.
2. Both teams will test their taglines for a similar amount of time since both are relatively new.
3. Next, we will analyze the response of the customers to bothoptions.
4. The tagline which shows a higher response rate of customers will certainly be more effective. Hence it will be adopted.