The Sun Pharma Case Study Assignment

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:

a) Propose the type of probability distribution that would accurately portray the above-mentioned scenario, and list out the three conditions that this distribution follows.
b) Calculate the required probability.

Answer:

a) The case described above can be modeled using the binomial distribution.

The three conditions due to which Binomial distributions was most favorable in this case are as follows:

- There are fixed number of trials (n = 10).
- Each trial is independent of others.
- Each trial can only give two possibilities, i.e., either success or failure.

Given these parameters, the binomial distribution becomes the appropriate statistical tool for finding the probability of a certain number of medicines, out of 10 samples, failing to deliver satisfactory results.

b) To find the probability that at most 3 medicines are unsatisfactory can be done by the following procedure in the next page:

Given

$P(\text{unsuccess}) = 1/5; \quad n = 10$

$q(\text{success}) = 4/5; \quad k = 3.$

To find: $P(X \leq 3)$.

Solution: Using Binomial distribution formula

ie. $P(X = k) = C(n, k) \times (P)^{(k)} \times (1-p)^{(n-k)}$

So,

$P(X=0) = C(10,0) \times \left(\frac{1}{5}\right)^{(0)} \times \left(\frac{4}{5}\right)^{10} = 0.107$

$P(X=1) = C(10,1) \times \left(\frac{1}{5}\right)^{(1)} \times \left(\frac{4}{5}\right)^{9} = 0.268$

$P(X=2) = C(10,2) \times \left(\frac{1}{5}\right)^{2} \times (4/5)^{8} = 0.302$

$P(X=3) = C(10,3) \times \left(\frac{1}{5}\right)^{3} \times (4/5)^{7} = 0.201$

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

$= 0.107 + 0.268 + 0.302 + 0.201$

$= 0.878 = \underline{\underline{87.8\%}} \Rightarrow P(X \leq 3).$

Therefore, the probability that at most 3 medicines won't work satisfactorily is ***87.8%***.

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:

a) Discuss the main methodology using which you will approach this problem. State all the properties of the required method.

b) Find the required interval.

Answer:

a) The main methodology to approach this problem is by using Central Limit Theorem, to find the range between which the potential mean might lie.

So, the central limit theorem states that for any kind of data, provided a high number of samples has been taken, the following properties hold true:

- Sampling distribution's mean (μ_x) = Population mean (μ),
- Sampling distribution's standard deviation (standard error) = $\sigma \sqrt{n}$,
- For $n > 30$, the sampling distribution becomes a normal distribution.

The formula used to find the interval is given as:

$$\text{Interval} = \left[\bar{X} - \frac{Z \times \sigma}{\sqrt{n}} \; , \; \bar{X} + \frac{Z * \sigma}{\sqrt{n}} \right]$$

b) To find the Intervals with the data given we can use the above formula as follows:

Given:

Sample size (n) = 100.

Sample mean ($\bar{x}$) = 207 seconds.

Sample Standard deviation (σ) = 65 seconds.

Confidence level = 95%

To find: Confidence Interval.

Solution: • Critical Value (Z_c) for confidence level of 95% = ± 1.96.

• Standard Error (S.E) = $\sigma / \sqrt{n}$

$= 65/\sqrt{100} = 6.5$

• Margin of Error (M.O.E) = Z_c * S.E

$= 1.96 \times 6.5 = 12.74$.

We know,

Confidence Interval = Sample mean ± M.O.E.

Confidence Interval = $(\bar{x} - M.O.E), (\bar{x} + M.O.E)$

$= (207 - 12.74), (207 + 12.74)$

$= (194.26, 219.74)$.

= (194.26 seconds, 219.74 seconds).

Hence, we can conclude that the mean might lie in between the range ***(194.26, 219.74)*** with a confidence level of 95%.

Question-3:

a) 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. Utilize 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.

b) 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.)

Answer:

a) From question we can create null and alternate hypothesis as:

Null Hypothesis (H_0): $\mu <= 200$ seconds, the medicine does a satisfactory job.
Alternate Hypothesis (H_1): $\mu > 200$, the medicine does a satisfactory job.

Since the null hypothesis is whether the value is really less than or equal to 200, we have to check whether the value stays within the acceptance region or it goes beyond the critical region (significance = 5%).

The upper tailed test will be used to check the validity of the hypothesis.
The rejection region will on the right hand side of the sampling distribution graph.

The procedure to check the hypothesis is given in the following page:

Given: $n = 100$; $\mu = 200$; $\mu_{\bar{x}} = 207$; $\sigma_{\bar{x}} = 65$; $\alpha = 0.05$.

To find: Whether to reject Null hypothesis.

Solution:

$\alpha = 0.05$ (given).

$\because$ It is an upper tailed test, we need to find Z-score for $1-\alpha$, i.e $1 - 0.05 = 0.95$.

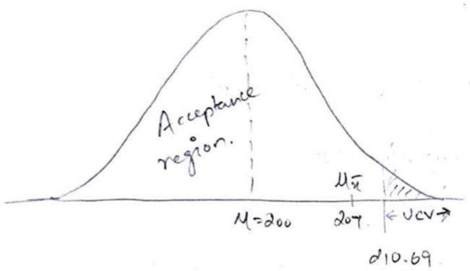
Z-score for $0.95 = 1.645$.

$$\text{Upper Critical Value (UCV)} = \left(\mu + \frac{Z \times \sigma}{\sqrt{n}}\right).$$

$$= 200 + \frac{1.645 \times 65}{\sqrt{100}}$$

$$= 200 + 10.69$$

$$= \underline{\underline{210.69}}$$



From the image above it is shown that the **sample mean value 207 seconds is less than the upper critical value of 210.69 seconds**, which indicates that the mean value lies in the acceptable region. Thus we fail to reject the null hypothesis.

For added confirmation or to understand how accurate the null hypothesis is, we can use p- value test as follow:

P-value Test:

Z score for sample mean:

$$Z = \frac{\mu_{\bar{x}} - \mu}{(\sigma/\sqrt{n})} = \frac{207 - 200}{65/\sqrt{100}}$$

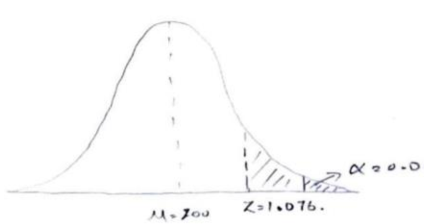
$$\therefore Z = 1.077$$

P-value for $Z = 1.077 = 0.8577$ [From Z-table]

$\because$ Since it is upper tailed test

$$\text{P-value} = 1 - 0.8577$$

$$= \underline{\underline{0.1423}}.$$



Since, the p-value **(14.23%)** is greater than significance value **(5%)**, again we can say that we have failed to reject the null hypothesis.

b) Hypothesis testing is done to check whether a claim made by a person or organization is true or false.

There are two types of errors that can be observed in hypothesis testing that are:

- Type 1(α) error = This is caused when we reject the null hypothesis even though the claim is a valid one.
- Type 2(β) error = This is due to the fact that even though the claim is invalid, we fail to reject the null hypothesis.

The choice of method depends on what errors we want to avoid and the consequences of making those errors.

- **Method 1 (α = 0.05, β = 0.45):**
  We prefer this method when we really want to minimize the chances of falsely rejecting the null hypothesis (Type I error, α). This is really important when there are serious consequences, like in situations where patient safety is at stake. With a low α value (0.05), we set a very strict standard for accepting a new batch, reducing the risk of approving a potentially ineffective or harmful medication.
- **Method 2 (α = 0.15, β = 0.15):**
  This method is preferred when the potential cost of not detecting a true effect (Type II error, β) is high, and an easier approach is acceptable. It's useful in cases where the cost of production or time constraints are high. With a higher α value (0.15), we're more lenient with potential Type II errors, making the testing process less important and suitable for situations where speed and efficiency matters.

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.

Answer:

A/B testing can be method used to determine the effectiveness of different marketing strategies, such as taglines, by comparing their performance with real users.

A/B testing is used to compare two versions of a webpage, ad, or product feature. You randomly split your audience into two groups, show each group a different version (A and B), and then analyze which one performs better. You measure user interactions like clicks, conversions, for engagement to determine which version is more effective. This helps you make data-driven decisions and optimize your content or product for improved results. Keep sample sizes meaningful, be aware of test duration, and maintain consistency in other factors for accurate insights.

Here is a stepwise procedure for conducting an A/B test for our case:

- **Define the Goal**: Figure out if Tagline A or Tagline B works better for attracting new subscribers in your ad campaign.

- **Split the Audience:** Randomly divide your audience into two groups, Group A and Group B.

- **Run the Test:** Show Tagline A to Group A and Tagline B to Group B, making sure it is done randomly and fairly.

- **Gather Data:** Collect information on user interactions, like click rates and sign-ups, from both groups.

- **Analyze the Results:** Compare how Tagline A and Tagline B performed using data analysis tools to find out which one is better.

- **Choose the Winner:** Based on the analysis, pick the tagline that does a better job of engaging users and getting new subscribers.

- **Use the Winning Tagline:** Put the winning tagline in your ad campaign to attract more subscribers.

- **Keep an Eye and Improve:** Continuously check how the chosen tagline is doing, and if things change, be ready to test or tweak it.