STATISTICS ASSIGNMENT

By: Srajan Kumar Shukla
11705007

Overview

The pharmaceutical company Sun Pharma is manufacturing a new batch of painkiller drugs, 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 drug to completely cure the pain), as well as the quality assurance (which tells you whether the drug was able to do a satisfactory job or not).

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.
Answer:

a) The probability distribution that will accurately portray the above scenario is binomial distribution. I have selected binomial distribution because it has taken fixed number of trials. In this each trial has one out of the two outcomes, that is the result is satisfactory and the result is not satisfactory. Also it say's that it is more likely a drug is able to produce satisfactory results.
   Here are some conditions that need to be followed for us to be able to apply the distribution:
   I. Total number of trials is fixed at n
   II. Each trial is binary, i.e., has only two possible outcomes - success or failure
   III. Probability of success is same in all trials, denoted by p

b) Here:
   n = 10
   success probability (p) = 7

   Let,
   Probability(Drug will not be satisfactory) = x
   Probability(Drug will be satisfactory) = 4x
   Thus,
   4x+x = 1
   x = 1/5
   x = 0.2
   Therefor
   Probability(Drug will not be satisfactory) = 0.2
   Probability(Drug will be satisfactory) = 4*0.2

= 0.8

P(x<=3) = P(x=0) + P(x=1) + P(x=2) + P(x=3)

P(success) = 0.2

P(failure) = 1-P(success)

= 0.8

$P(x=0) = {}^{n}C_{r} (p)^{r} (1-p)^{n-r}$

$= {}^{10}C_{0} (0.2)^{0} (1-0.2)^{10-0}$

$= {}^{10}C_{0} (0.2)^{0} (0.8)^{10}$

$= 1 * 1 * (0.8)^{10}$

$= 0.10737$

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

$= {}^{10}C_{1} (0.2)^{1} (0.8)^{10-1}$

$= 10 * 0.2 * (0.8)^{9}$

$= 10 * 0.2 * 0.134$

$= 0.269$

$P(x=2) = {}^{n}C_{r} (p)^{r} (1-p)^{n-r}$

$= {}^{10}C_{2} (0.2)^{2} (0.8)^{10-2}$

$= \frac{10 * 9}{2 * 1} * (0.2)^{2} * (0.8)^{8}$

$= 45 * 0.04 * 0.168$

$= 0.302$

$$P(x=3) = {}^{n}C_{r} (p)^{r} (1-p)^{n-r}$$
$$= {}^{10}C_{3} (0.2)^{2} (0.8)^{10-2}$$
$$= \frac{10 * 9 * 8}{3 * 2 * 1} * 0.008 * 0.210$$
$$= 120 * 0.008 * 0.210$$
$$= 0.201$$

Formula:

$$P(x <= 3) = P(x=0) + P(x=1) + P(x=2) + P(x=3)$$

$$P(x <= 3) = 0.107 + 0.269 + 0.302 + 0.201$$
$$= 0.879$$

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 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. Limit your answer to 150 words.

b.) Find the required interval.

Answer

a)

The problem could be solved by using central limit theorem, which is used to estimate the population mean in form of an interval.

For sampling distribution, on the basis of theorem we can state:

(i) Sampling distribution's mean = Population mean

i.e $\mu_{\bar{x}} = \mu$

(ii) Sampling distribution's standard deviation (Standard error) $= \frac{\sigma}{\sqrt{n}}$

where, $\sigma \rightarrow$ standard deviation of original distribution

$n \rightarrow$ sample size.

Given the values of sample size, mean and standard deviation, the confidence interval for μ lies in range $\left(\bar{x} - \frac{z^* s}{\sqrt{n}}, \bar{x} + \frac{z^* s}{\sqrt{n}}\right)$

Where, $\bar{x} \rightarrow$ mean

$z^* \rightarrow$ Z-Score

$s \rightarrow$ standard deviation

$n \rightarrow$ sample size

b)

Given :- $n = 100$ $\bar{x} = 207$

$s = 65$ $z^* = \pm 1.96$

The confidence interval $= \left(\bar{x} - \frac{z^* s}{\sqrt{n}}, \bar{x} + \frac{z^* s}{\sqrt{n}}\right)$

$$= \left(207 - \frac{1.96 \times 65}{\sqrt{100}}, 207 + \frac{1.96 \times 65}{\sqrt{100}}\right)$$

$$= \left(207 - \frac{127.4}{10}, 207 + \frac{127.4}{10}\right)$$

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

$$= (194.26, 219.74)$$

Hence, margin of error $= \frac{1.96 \times 65}{\sqrt{100}}$

$$= \frac{127.4}{10}$$

$$= 12.74$$

The range is 194.26 seconds to 219.74 seconds.

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 (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 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. Also, the hypothesis test that you are conducting is the same as mentioned in the previous question - you need to test the claim whether the newer batch produces a satisfactory result.)

a) Answer: - Sample Size (n) = 10 Sample Mean (X) = 207

Sample Standard Deviation (Óx) = 65 Significance Level (5%) = 0.05

Based on question we can safely assume

Null Hypothesis (Ho) <= 200 **Alternate Hypothesis (H1) : > 200**

As Alternate Hypothesis H1 > 200 its tail test and rejection region will be on right side.

We will be using following two Hypothesis testing:

- Critical Value Method
- P-Value Method

Following would be computation of Critical value Method and P-Value Method and Hypothesis for the problem given.

$H_0 <= 200$ and $H_1 > 200$

Critical value method

$\mu = 200 \quad \sigma = 65 \quad n = 100$

$\alpha = 0.05$ i.e (5% significance level)

$\bar{x} = 207$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{65}{\sqrt{100}} = \frac{65}{10} = 6.5$$

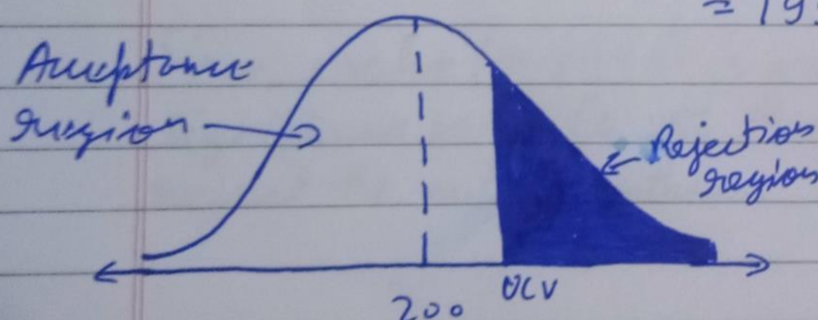
$$Z_c = (1 - 0.05) = (0.95)$$

$$\therefore Z_c = 1.644$$

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

$$UCV = 200 + (1.644 * 6.5) = 210.6912$$

$$LCV = 200 - (1.644 * 6.5) = 199.314$$



We can safely say: failing to reject null hypothesis.

P-Value method

$\mu = 200 \quad \sigma = 65 \quad \bar{x} = 207 \quad n = 100 \quad \alpha = 0.05$

$$z_{score} = \frac{(\bar{x}-\mu)}{\frac{\sigma}{\sqrt{n}}}$$

$$= \frac{207-200}{65/\sqrt{100}}$$

$$= \frac{7}{6.5}$$

$$= 1.08$$

$\therefore$ p value = 0.8599

Since, p value (0.8599) is greater than significance value (0.05), thus we fail to reject the null hypothesis.

In both the testing, we failed to reject the null hypothesis.

b) Answer: -

α=0.05
β=0.45
n=100
σ =65
$\bar{x}$=207
μ=200
A type-I error represented by α occurs when we reject true null hypothesis.
A type-II error represented by β occurs when we fail to reject null hypothesis.
In first scenario, α=0.05 and β=0.45 is the condition where type 2 error occurs frequently.
As type-I error is less dangerous then type-II it is better if we preferred both values at 0.15.

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 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:

The choice of tagline for the campaign is very subjective and difficult to predict. To resolve the issue here we use the A/B testing also known as split testing or bucket testing. A/B testing provides the better way for the two different version of same elements and can predict which option will perform better. Like the above example where we have 2 taglines proposed for the campaign by the team and we need to choose one. This is can be done with the help of A/B testing. Running an A/B test that directly compares a variation against a current experience and ask focused question about the difference of two different tagline. It is also checked to find which tagline will create more value for our customers, then the more chances of business getting success.
The A/B testing involves with the following sub tasks,
1. Collecting data
2. Identify the objective of A/B testing
3. Define and Arrive at hypothesis parameters
4. Create variations compared to the controlling entity
5. Run the tests
6. Analyze the result and derive metrics
7. Generate the Outcome
8. Repeat if needed.

If we consider the above example, let us assume and proceed with the below steps:
1. Out of 80,000 drugs, sample of 100 were taken, necessary statistical metrics were derived.
2. Objective was to observe if pain killer drug is effective enough post release to the market.
3. Based on sample sets, we defined the null and null hypothesis testing using critical values ad p- value method conclude that drug is safe to use.
4. With product being approved for launch, marketing team defines the two different tagline let's Tagline 1(which is already running in market) and Tagline 2 (the new tagline which suggested by the team).
5. Two variations will be marketed across testing group or full hypothesis testing group of full production line as needed or approved by firm.
6. Based on the sales of drugs under P1 and P2 categories, the firm will captures the conversions of respective areas as P1, P2 respectively.
7. Two proportions test will be conducted based on two sets,
8. The firm tries find the difference of letter with the former with significance level defined (ideally 5%). Here hypothesis would be H0 = Tagline 1 and H1 would not be tagline 1(old one) or New tagline is better than old tagline.
9. After getting the conclusion and plan for revise the marketing strategy.
10. Extend the experiment if it's needed further to arrive at outcomes.