Statistics Assignment- Solutions

Problem Statement

Comprehension: -

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 01 – a):-

1. Looking at the above scenarios, where the sample is taken 10 times.
2. It produces only two possible outcomes. The drug will either produce the satisfactory result or not.
3. All the sample results are independent to each other. It means result of previous drug testing will not impact the result of next drug.
4. Probability of getting satisfactory result is 4 time more likely than will not produce the satisfactory result. So the **P(success) = 4/5 = 0.8** and **P(Not Success) = 1- 0.8 = 0.2** (As sum to total probability of outcome is 1)

Conclusion:- From the above 3 mentioned points, it's a type of **Binomial Distribution.**

Answer 01 – b)

As it is a Binomial Distribution, the probability of the binomial distribution is given as for a binomial experiment consisting of n trials and result in x number of success. If the Probability of the success on an individual trial P, Then the binomial probability would be given as :-

$$b(x;n,p) = {}^{n}C_{x} \times p^{x} \times (1-p)^{n-x}$$

$$= \left(\frac{n!}{(n-x)!\,x!}\right) \times p^{x} \times (1-p)^{n-x}$$

Probability (Drug will do the satisfactory) = 4 * Probability (Drug will not do the satisfactory) --- **From the question**

The sum of probability of all the possible outcomes is 1. It means
Probability (Drug will do the satisfactory) + Probability (Drug will not do the satisfactory) = 1 (Probability Theorem)

Let's suppose Probability (Drug will not do the satisfactory) = **X**

So, Probability (Drug will do the satisfactory) = 4X

4X + X = 1 (from the probability Theorem)
X = 1/5 = 0.2

Probability (Drug will not do the satisfactory) **= 0.2**
Probability (Drug will do the satisfactory) = 4*0.2 = **0.8**

To Find the at most, 3 drugs are not able to do a satisfactory job.

P(X<=3) = P(X=0) + P(X=1) + P(X=2) + P(X=3) where X is the random variable of a drug will not able to do a satisfactory job.

So P(X=0), it's probability that all trials went successful and we didn't have any drug which failed to perform satisfactory. Number of trails (n) = 10

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

$$P(x=0) = {}^{n}C_{0} (P)^{0} (\bar{P})^{10-0}$$

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

$$= 1 \times 1 \times 0.8^{10}$$

$$\boxed{= 0.107}$$

P(X=0) = 0.107

So P(X=1), it's probability of getting 1 drug which not able to perform satisfactory in our sample trials. Number of trails (n) = 10

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

$$= \frac{10!}{9!\,1!} \times 0.2 \times (0.8)^{9}$$

$$= 10 \times 0.2 \times 0.134$$

$$\boxed{= 0.269}$$

P(X=1) = 0.269

So P(X=2), it's probability of getting 2 drug which not able to perform satisfactory in our sample trials. Number of trails (n) = 10

$$P(X=2) = {}^{10}C_2 (0.2)^2 (0.8)^{10-2}$$

$$= \frac{10!}{8!\,2!} \times 0.04 \times 0.168$$

$$= 45 \times 0.04 \times 0.168$$

$$\boxed{P(X=2) = 0.302}$$

P(X=2) = 0.302

So P(X=3), it's probability of getting 3 drug which not able to perform satisfactory in our sample trials. Number of trails (n) = 10

$$P(X=3) = {}^{10}C_3 (0.2)^3 (0.8)^{10-3}$$

$$= \frac{10!}{7!\,3!} \times 0.008 \times 0.210$$

$$= \frac{10 \times 9 \times 8}{3 \times 2} \times 0.008 \times 0.210$$

$$= 120 \times 0.008 \times 0.210$$

$$\boxed{P(X=3 = 0.201}$$

P(X=3) = 0.201

Total Probability P(X<=3) = P(X=0) + P(X=1) + P(X=2) + P(X=3)

As, we got P(X=0) = 0.107

P(X=1) = 0.269

P(X=2) = 0.302

P(X=0) = 0.201

P(X<=3) = 0.107 + 0.269 + 0.302 + 0.201

= 0.879 (Approx. 88 %)

The Required Probability is 0.879 (Approx. 88 %).

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.

Answer 02 -A :-

The above problem we have given that we sample of 100 drugs and the mean value of the sample is 207 seconds with the standard deviation to 65 sec. So we have all the information about the sample size and we need to find the mean of the population. This is the case **Central Limit Theorem**, which states that no matter how the original population is distributed, the sampling distribution will follow the below properties.

1. Sample mean of the distribution is equal to population mean.
2. Sampling distribution's standard deviation (Standard Error) = $\frac{\sigma}{\sqrt{n}}$, where sigma is the population's standard deviation and n is the sample size.
3. For n (sample size) > 30, the sampling distribution becomes a normal distribution.

Answer 02 – B

Given :-

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

Sample Size (n) = 100

Sample Standard deviation ($\sigma_{\bar{x}}$) = 65

Confidence level = 95 %

Z value associated with = ± 1.96

As per Central limit theorem,

Sample mean ($\bar{x}$) = population mean (μ)

$\boxed{\mu = 207}$

Population Mean range = $\mu \pm$ Margin of Error

~~μ =~~

Margin of Error = ?

As per central limit theorem,

$$\text{Standard of Error} = \frac{\sigma}{\sqrt{n}} \approx \frac{S}{\sqrt{n}}$$

$$= \frac{65}{\sqrt{100}} = \frac{65}{10} = 6.5$$

Standard Error = 6.5

Margin Error = Z × Standard Error

$$= Z \times \left(\frac{\sigma}{\sqrt{n}}\right)$$

Z associated with 95% confidence level $= \pm 1.96$

Margin Error $= \left(\pm 1.96 \times 6.5 \quad \cancel{= 1.96 \times 6.5}\right)$

$= \pm 12.74$

The Required Range of population mean $= (207 - 12.74,\ 207 + 12.74)$

$= (194.26,\ 219.74)$

The required range of population is (194.26, 219.74)

Question 03

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).

Answer 03 – A

Q3 – A

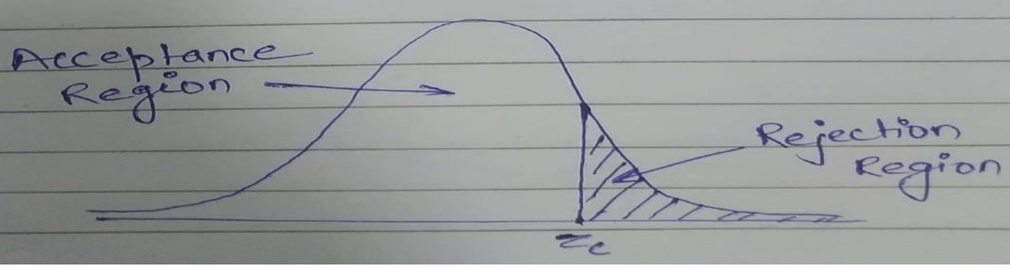
Given :- Sample Size (n) = 10
Sample mean ($\bar{x}$) = 207
Sample Standard deviation ($\sigma_{\bar{x}}$) = 65

According to question,

Null hypothesis (H_0) $\leq$ 200 sec
Alternate hypothesis (H_1) $>$ 200

Significance level (α) = 5% = 0.05

Looking at the alternate hypothesis $H_1 > 200$ it's ~~tapper~~ upper tail test.
It means the rejection region on right side of distribution



If the critical value of the hypothesis testing lies in the region of Rejection then we would reject the null hypothesis.

If the critical value of the hypothesis testing Lies in the acceptance region then we would fail to reject the null hypothesis.

Methods of hypothesis testing :-

1. Critical value Method
2. p - value Method

Critical value Method :-

Step 01 :- Calculate the Z_c value for the given α (significance level)

The cumulative probability of the critical point (the total area till that point) would be 0.950

So, looking at the z table,
z score for 0.950 is 1.645

$$\boxed{z_c = 1.645}$$

Step 02 :- Find critical value

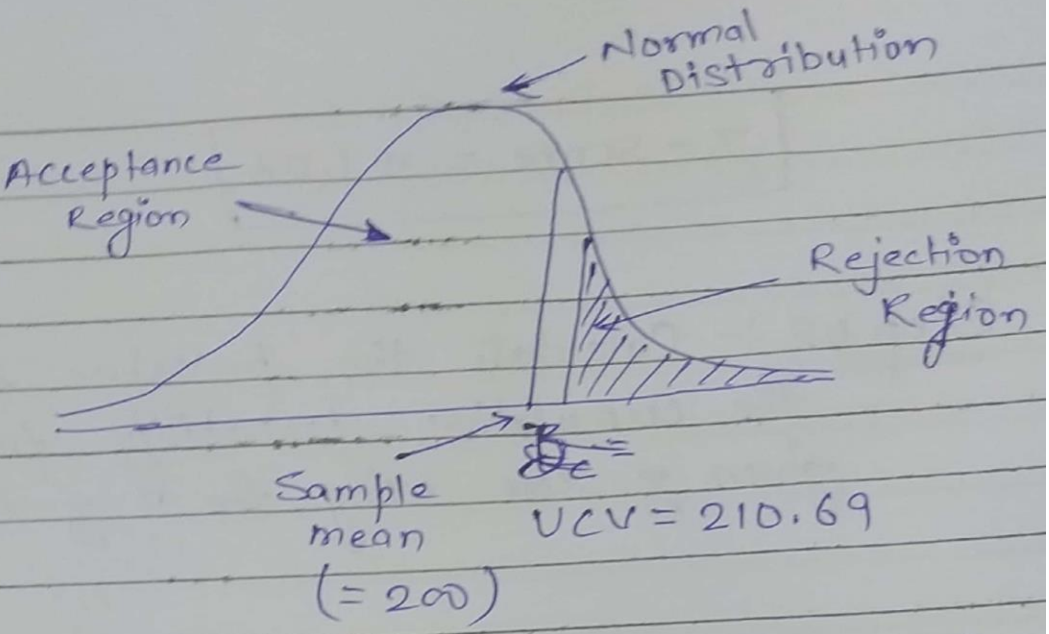
$$UCV = \mu + Z_c \times \left(\frac{\sigma}{\sqrt{n}}\right)$$

$$= 200 + 1.645 \times \left(\frac{65}{\sqrt{100}}\right)$$

$$= 200 + 1.645 \times 6.5$$

$$= 200 + 10.69$$

$$\boxed{UCV = 210.69}$$



As sample mean(200) lies in the acceptance region, so we are Failed to Reject the null hypothesis

2. p- Value Method :-

Step 01:- Calculate the value of z score for the sample mean point on the distribution

$$z\text{-score} = \frac{(\bar{x} - \mu)}{\frac{\sigma}{\sqrt{n}}}, \quad \begin{pmatrix} \bar{x} = 207 \\ \mu = 200 \\ \sigma = 65 \\ n = 100 \end{pmatrix}$$

$$= \frac{207 - 200}{65/\sqrt{100}} = \frac{7}{6.5}$$

$$= +1.08$$

$\boxed{z\text{-score} = +1.08}$

Step 02 :- Calculate the p-value from the cumulative probability for the given z-score using the z-table.

The value from z table corresponding to +1.08 is 0.8599

$\boxed{p\text{-value} = 0.8599}$

Step 03 :- Make a decision on the basis of p-value.

$\alpha = 0.05$ (Given)

So, p-value is greater than Significance value (α). ($p\text{-value} > \alpha$)

We failed to Reject the null hypothesis

From the results of both the hypothesis testing, it's we failed to reject the null hypothesis.

Answer 03- B:-

In the first scenario, alpha is 0.05 and beta is 0.45, is the condition where we type 2 error frequently. On the later both the values are at same value, hence chances of type 1 error increases compared to the former.

As type 1 error less dangerous than type 2, latter with both values at 0.15 preferred.

Question 04

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.

Answer 04:-

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's also checked the 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. Analyse the result and derive metrics
7. Generate the Outcome
8. Repeat if it's 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.