# STATISTICS ASSIGNMENT

QUESTION 1:

a) T-Distribution would accurately portray the scenario for the following reasons
    1. The sample observations were obtained randomly.
    2. The sample size is 10 which is less than 30.
    3. The standard deviation of the population is unknown.

b) Ho: $\mu$ <=3,
   H1: $\mu$ >3,

   Probability that a drug is able to produce satisfactory result(S) = 4* Probability that a drug is not able to produce satisfactory result (NS)

   As the total probability is always equal to 1,

   NS + S = 1
   NS + (4*NS) = 1
   5NS = 1
   NS = 1/5
   NS = 0.2
   S= 4*NS = 4*0.2 = 0.8

   Sample size (n) = 10
   Sample mean ($\bar{x}$) = 0.3
   Population mean ($\mu$) = 0.2
   Degrees of freedom (df) = n-1 = 10-1 = 9

   Standard deviation of the sample can be calculated by using the below formula

Standard Deviation = Standard Error = $\sqrt{\frac{P(1-P)}{n}}$

Here Standard Error = $\sqrt{\frac{S * NS}{n}} = \sqrt{\frac{0.8 \times 0.2}{10}} = 0.126$

Observed t-value (t) = $\frac{\bar{x} - \mu}{S_{\bar{x}}} = \frac{0.3 - 0.2}{0.126} = \frac{0.1}{0.126}$

$= 0.793.$

Probability from T table based on the T – Statistic (0.793) and the Degrees of freedom (9) is 0.775

Therefore the probability that at most 3 drugs are not able to do a satisfactory job is 0.775.

QUESTION 2:

a) Sample size (n) = 100
   Sample mean (x̅) = 207
   Sample standard deviation(S) = 65
   Confidence level = 95%
   Z* for 95% confidence level is 1.96

   As the sample size is greater than 30, on applying the Central Limit Theorem for this scenario, the sampling distribution becomes normally distributed

   Thus the following properties of normal distribution can be applied.

- The normal curve is bell shaped.
- The total area under the curve is equal to one.
- The distribution is symmetrical about its mean, median and mode which are coincidental at the center.
- The normal curve approaches, but never touches, the x-axis.
- Between μ − σ and μ + σ the graph is concave down and elsewhere the graph is concave up. The points at which the graph changes concavity are called inflection points.

b)

The range in which the population mean might lie can be calculated using the below formulae.

$$\text{Confidence Interval} = \bar{x} - \frac{Z^{*}S}{\sqrt{n}},\ \bar{x} + \frac{Z^{*}S}{\sqrt{n}}$$

$$= 207 - \frac{1.96 \times 65}{\sqrt{100}},\ 207 + \frac{1.96 \times 65}{\sqrt{100}}$$

$$= (194.26,\ 219.74)$$

Therefore the range in which the population mean might lie is 194.26 to 219.74

QUESTION 3:

a) Sample size (n) = 100

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

Sample standard deviation(S) = 65

Confidence level = 95%

Population mean ($\mu$) = 200

Significance level ($\alpha$) =0.05

Beta ($\beta$) = 0.45

1) Critical Value Method:

   H0: $\mu$<=200

   H1: $\mu$>200

   As the alternate hypothesis is directional and is an upper tailed test, the rejection region lies on the right side.

   Acceptance Region = (1- $\alpha$) = (1-0.05) = 0.95

   Zc = 1.645 based on the above cumulative probability for the acceptance region, 0.95.

   The Upper Critical Value can be calculated using the below formula

$$\text{Critical Value (U.C.V)} = \mu + (Z_c \times \sigma_{\bar{x}})$$

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

$$\text{U.C.V} = 200 + (1.645 \times 6.5) = 210.692$$

As sample mean ($\bar{x}$ = 207) is less than the Upper Critical Value (U.C.V = 210.692), the sample mean lies in the acceptance region so, we fail to reject the null hypothesis.

2) P – Value Method:

   H0: μ<=200

   H1: μ>200

   As the alternate hypothesis is directional and is an upper tailed test, the rejection region lies on the right side.

   Acceptance Region = (1- α) = (1-0.05) = 0.95

   The Z score of the sample mean can be calculated using the following formula.

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}, \text{ where } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$\text{Here } \sigma_{\bar{x}} = \frac{65}{\sqrt{100}} = 6.5$$

$$\therefore Z = \frac{207-200}{6.5} = 1.076$$

   Cumulative probability for the above Z Score = 0.8577(from Z table).

   As it is one tailed test, P Value = 1 – 0.8577 = 0.1423.

As the P Value (0.1423) is greater than the Significance Level ($\alpha = 0.05$), the sample mean lies in the acceptance region so, we fail to reject the null hypothesis.

b)

Method 1: $\alpha = 0.05$, $\beta = 0.45$
Total Error = 0.05+0.45 = 0.5

Method 2: $\alpha = 0.15$, $\beta = 0.15$
Total Error = 0.15+0.15 = 0.3

Scenario 1: Preferring Method 1 over Method 2

Consider the scenario where the patient prefers low dosage drugs over the time of effect based on his medical condition (As high dosage drugs are not safe for the patient).

Assuming that Sun Pharma pain killer drugs are very safe to use for the patients because of its composition compared to the pain killer drugs produced by other companies. In this scenario, we need to use Method1 for the following reasons:

- Probability of committing Type 1 Error is low compared to Method 2 so there would be less chances of rejecting the null hypothesis when it's true. So, there is high probability of drugs passing the quality assurance test.

- Probability of committing Type 2 Error is high compared to Method 2, as the patient prefers low dosage drugs over time of effect, even if the drugs with more time of effect pass the quality assurance, it would not affect the patient negatively as he prefers low dosage drugs over time of effect.

Scenario 2: Preferring Method 2 over Method 1

Consider the scenario where the patient cannot compromise on the time of effect as it is unsafe for him if the pain killer drug takes more time to show its effect, based on his medical condition
In this scenario, Method 2 is preferable for the following reasons:

- Probability of committing Type 1 Error is high compared to Method 1 so there would be more chances of rejecting the null hypothesis when it's true. This would in no way harm the patient.

- Probability of committing Type 2 Error is low compared to Method 1, as the patient needs the drug to show the effect in time, less number of drugs would pass the quality assurance if the time of effect is more in Method 2 compared to Method 1.

Conclusion:
So, based on the medical condition of the consumer base that will be using the medicines, Sun Pharma has to decide upon the method to be used for the Hypothesis Testing

- If the patients prefer low dosage drug over the time of effect, it can go for Method 1 even though the total error is high compared to Method 2.
- If the time of effect of the drug is very important for the patients based on their medical condition, it can go with Method 2.

<u>QUESTION 4:</u>

Why:

We need to use A/B testing to plan an effective online ad campaign because of the following reasons:

- It reduces the time and cost of advertising. If we know which tagline works better we can avoid spending time and money on the other tagline for advertising, as it results in lower conversion.
- It increases the conversion rate. If we know which tagline works better we can spend more time and money on it to achieve better conversion and optimize the experience for improved results.
- Helps to understand the customers better. It will help in understanding the customers' behavior which in turn helps in designing more improved methods of advertising.

How:

- First we need to decide if it is an on-site or off-site test. On-site tests refer to marketing using the website Off-site tests refer to any other marketing outside of the website.
- Identify the conversion goals like clicking a button or a link, purchasing the product etc.

- As two taglines were proposed for the campaign, the traffic should be split into 50 – 50 and the two taglines should be assigned randomly to the traffic.
- Timing plays an important role in the ad campaign results. Run the two variations of the taglines simultaneously .If not, it will be difficult to make out if the difference in the performance is due to the timing of the campaign or due to the different taglines used.
- Give enough time for the test to produce useful data. If the traffic is less it takes long to produce required data.
- The tagline which performs better becomes the “Control”.
- Repeat this experiment on a different traffic or with a new “Variation” of tagline to improve the performance.
- Test one change at a time; it helps in identifying the changes which resulted in better performance.
- Following the above steps helps in optimizing the experience for a desired outcome and in taking steps for making a marketing campaign more effective.