Inferential Statistics and Hypothesis Testing

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.

Answer:

The probability distribution that accurately portrays the above scenario is **Binomial Distribution** because the event has only two possible outcomes i.e., either the drug produces a satisfactory result or an unsatisfactory result.

The three conditions that Binomial Distribution follows are:

- The total number of trials is fixed at n and each observation is independent.
- Each trial has only 2 possible outcomes i.e. success or failure.
- The probability of "success" p is the same for each outcome.

b.) Calculate the required probability:

Answer:

Let **X** be the number of drugs that produce an unsatisfactory result after testing of 10 drugs.

Therefore, X follows a binomial distribution with n = 10. As it is specified that it is 4 times more likely that a drug is able to produce a satisfactory result than not.

Let **p** be the probability for an **unsatisfactory result.** So,

$p+4p = 1$ { The sum of the probabilities in a probability distribution is always 1. }
$\Rightarrow 5p = 1$

Therefore, $p = 1/5 = 0.2$

Probability (drug produces unsatisfactory result) i.e. p=0.2

Probability (drug produces satisfactory result) = 1-p =0.8

Therefore, the theoretical probability that at most, 3 drugs are not able to do a satisfactory job can be defined as cumulative probability of X, denoted by F(x), which is the probability that the random variable X takes a value less than or equal to x.

Therefore, $F(x) = P(X \leq x)$

$\Rightarrow F(3) = P(X \leq 3)$

$\Rightarrow F(3) = P(X=0)+P(X=1)+P(X=2)+P(X=3)$

Formula for Binomial distribution :

$$\boxed{P(X=r) = {}^{n}C_{r}\,(p)^{r}\,(1-p)^{n-r}}$$

n = no. of trials

p = probability of unsatisfactory result = 0.2

r = number of unsatisfactory result after n trials

i.e. r = 0, 1, 2, 3

$$P(X=0) = {}^{10}C_{0}\,(0.2)^{0}\,(1-0.2)^{10-0} = 0.107$$

$$P(X=1) = {}^{10}C_{1}\,(0.2)^{1}\,(1-0.2)^{10-1} = 0.268$$

$$P(X=2) = {}^{10}C_{2}\,(0.2)^{2}\,(1-0.2)^{10-2} = 0.302$$

$$P(X=3) = {}^{10}C_{3}\,(0.2)^{3}\,(1-0.2)^{10-3} = 0.201$$

$$\therefore P(X \leq 3) = P(X=0)+P(X=1)+P(X=2)+P(X=3)$$
$$= 0.107 + 0.268 + 0.302 + 0.201$$
$$= 0.878$$

Hence,

$F(3) = P(X \leq 3) = 0.878 = 87.8\,\%$

Therefore, the probability that at most 3 drugs are unable to do satisfactory job is 87.8%.

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.

Answer:

The main methodology for approaching the problem would be to use **Central Limit Theorem** in order to estimate the population mean in the form of an interval.

Central limit theorem states that no matter how the original population is distributed, the sampling distribution will follow these three properties:

- Sampling distribution's mean ($\boldsymbol{\mu_{\bar{x}}}$)= Population mean ($\boldsymbol{\mu}$)
- Sampling distribution's standard deviation (Standard error) = $\frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}$, where σ is the population's standard deviation and n is the sample size.
- For **n > 30**, the sampling distribution becomes a **normal distribution.**

❖ CLT lets us assume that the sample mean would be normally distributed with mean (μ) and therefore sampling distribution's standard deviation $\frac{\sigma}{\sqrt{n}}$ can be taken as (approx. $\frac{S}{\sqrt{n}}$) where (S) is standard deviation of the sample and (σ) is population's standard deviation. Moreover, we only know the standard deviation of the sample in the given case.

b) Find the required range.

Answer:

- Sample size (n) = 100
- Sample mean ($\mu_{\bar{x}}$) = 207
- Sample standard deviation (S) = 65

Given the sample's size, mean and standard deviation, we can say that the confidence interval for μ lies in the range of ($\mu_{\bar{x}}$ - $\frac{Z^*S}{\sqrt{n}}$, $\mu_{\bar{x}}$ + $\frac{Z^*S}{\sqrt{n}}$) i.e. the population mean and sample mean differ by a **margin of error** given by $\frac{Z*S}{\sqrt{n}}$.

Here Z^* is the Z-score associated with a y% confidence level i.e. 95% in the above case.

Z* Values for Commonly Used Confidence Levels are:

Confidence Level	Z*
90%	± 0.65
95%	± 1.96
99%	± 2.58

$\therefore$ The confidence interval is given by:

$$\left(\bar{\mu}_x - \frac{Z^* S}{\sqrt{n}} , \bar{\mu}_x + \frac{Z^* S}{\sqrt{n}}\right)$$

$$= \left(207 - \frac{1.96 \times 65}{\sqrt{100}} , 200 + \frac{1.96 \times 65}{\sqrt{100}}\right)$$

$$= (207 - 12.74 , 200 + 12.74)$$

$$= (194.26 , 219.74)$$

i.e. Margin of error corresponding to 95% confidence level = 12.74

Therefore, we can say that the **population mean (μ) lies between 194.26 seconds and 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.

Answer:

Null(H_0) and Alternate Hypothesis(H_1) are:

H_0 : $\mu \leq 200$ seconds, the time of effect to be considered as having done a satisfactory job.

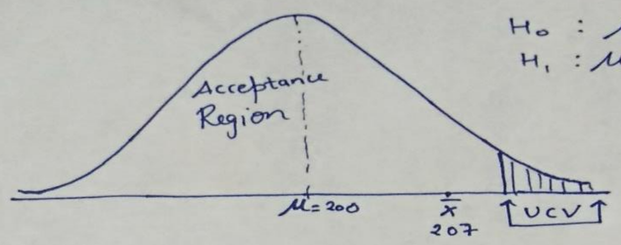
H_1 : $\mu > 200$ seconds, the time of effect to be considered as not having done a satisfactory job.

Test Type:

> sign in alternate hypothesis tells us that it would be a **One-tailed test (i.e. upper-tailed test)**. This is a directional hypothesis and therefore the **rejection region will lie on the right side** of the distribution.

- Sample size n = 100
- Assumed Sample mean (μ) = 200
- Sample mean ($\mu_{\bar{x}}$)= 207
- Sample standard deviation $\sigma_{\bar{x}}$ = 65.
- Significance level α = 5% i.e. 0.05

❖ **Critical Value Test**:



$H_0 : \mu \leq 200$
$H_1 : \mu > 200$

- $n = 100$
- Assumed sample mean $(\mu) = 200$
- Sample mean $(\mu_{\bar{x}}) = 207$
- Sample standard deviation $(S) = 65$ {Here $\sigma \approx S$}
- Significance Level $(\alpha) = 5\% = 0.05$

Cumulative probability till critical area $= 1 - 0.05$
$= 0.95$

Z score for $0.95 = 1.645$ } From Z-table

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

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

$= 200 + 10.69$

$= 210.69$

Since the **sample mean ($\mu_{\bar{x}}$) i.e. 207 seconds is less than the Upper Critical Value of 210.69 seconds,** it lies in the acceptance region. Therefore, we **fail to reject the null hypothesis** which states that the drug needs to have a time of effect of at most 200 seconds to be considered as having done a satisfactory job.

- **p-value Test:**

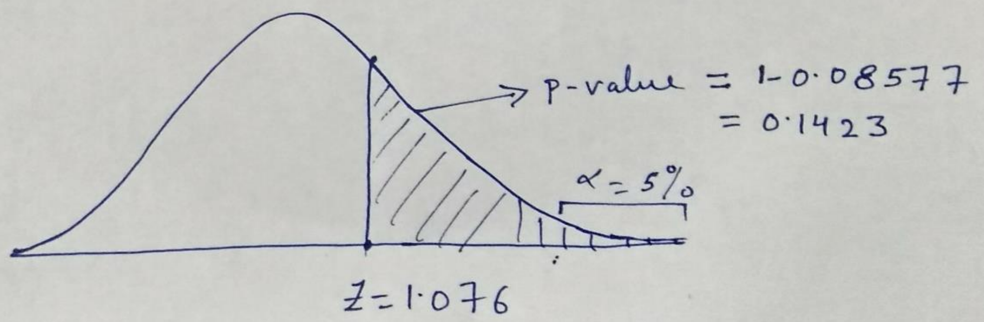
Z score for sample mean is given by:

$$Z = \frac{\bar{\mu}_x - \mu}{(\sigma/\sqrt{n})}$$

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

$$= 1.077$$

The p-value for $Z = 1.077 = 0.8577$ } from Z-table



Since the sample mean is on the right side of the distribution and this is a one-tailed test, the pvalue would be (1 - 0.8577) = 0.1423 (14.23%)

Since **the p-value is greater than the significance level (0.1423 > 0.05),** we **fail to reject the null hypothesis** which states that the drug needs a time of effect <= 200 seconds to do a satisfactory job.

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:

In hypothesis testing:

- **Type-1 (α)** error refers to the scenario where we decided to reject the NULL hypothesis even though it was TRUE

- **Type-2 (β)** error refers to the scenario where we failed to reject the NULL hypothesis when it was FALSE

There are two possible scenarios:

- **Scenario-1:** The value of α and β come out to be 0.05 and 0.45 respectively.

- **Scenario-2:** The value of α and β are controlled at 0.15 each.

We try to keep the α error limited to a very low value when there are side-effects related to acceptance of alternate hypothesis. From the scenario stated in question, if there is any side effect related to the overdose of the painkiller drug, we need to keep the α error to a minimum i.e. we are conservative in rejecting the null hypothesis.
So, if the painkiller drug has significant side effects, we would like to keep the value of α and β come out to be 0.05 and 0.45 respectively.
On the contrary, we tend to be relaxed with the α error if there were no major side effects in the painkiller drug. This would mean that we can be aggressive about changing the status quo and establish the alternate hypothesis. In such cases, our β error limit would be very restrictive.

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:

A/B testing provides a way to test two different versions of the same element and see which one performs better for a given goal. Since two taglines were proposed for the campaign, and the team is unable to decide the best of the two; we can use A/B testing to predict which tagline would be more effective in the long run.
We can show one tagline for a specific set of audience (controlled version) and another tagline for another set of audience (variant version). As different set of audience are shown different set of ads, their responses to these 2 ads can be measured statistically and thereby used to determine which ad campaign gets converted into sales and generates revenue.

Why A/B Testing:

- It Improves content engagement with the customer as A/B tests invariably make final versions of any product better for the customers.

- More conversion can be sought with minimal investment.

- A/B testing's ease of analysis makes it simple to analyze real and factual results. It's relatively easy to determine a winner and a loser based on elementary metrics.

- A/B testing allows for maximum output with minimal modifications, eventually leading to increased revenue and substantial growth.

- Any changes should be done after A/B testing as it will accurately let us know the customers receptiveness to the same.

Stepwise Procedure:

- **Performing Research:** Background research plays a critical role in A/B Testing. Figure out the problem areas using surveys, analytics tools etc. in order to

- **Collect Data:** Enough data should be collected to generate effective insights from the collected data like detailing low conversion rates or high drop-off rates areas.

- **Set Business Goals:** Setting conversion goals helps in understanding what the objective is. After that we find the metrics that determine which ad campaign will result in more sales and generate greater revenue.

- **Formulate Hypothesis:** Based on the above research insights, a hypothesis should be built. The hypothesis can be arrived at by determining statistics on the user behaviour on the type of ad watched, etc.

- **Determine the test sample size:** Appropriate calculation of sample size should be done so that the test results are statistically significant.

- **Determining the time-span:** The recommended testing time should be aptly formulated. Once the duration is decided, the test shouldn't be stopped before the duration ends. In this case since there are 2 taglines, the test duration can be calculated keeping in mind the expected change in the conversion rate.

- **Testing and drawing conclusions:** Wait for the stipulated time for achieving statistically significant result. Once the testing is done, analyze the test results and, if it succeeds, choose that tagline. If the test remains inconclusive, we should try to implement them in subsequent tests.