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.

Solution

Let's say S be the event which implies that the Drug was Good &
And B be the event which implies that the Drug was not good or bad or didn't produce a satisfactory result.

Then as per the question, **it is 4 times more likely that a drug is able to produce a satisfactory result than not.**
As per this

$$P(S) = 4P(B)$$

And we also know as these are my favorable outcomes, then as per the rule of probability

$$P(S) + P(B) = 1$$

So, after the substituting the value we get the probabilities of favorable outcomes as

$$P(S) = \frac{4}{5} \ \& \ P(B) = \frac{1}{5}$$

Let X be the event that out of 10 drugs, x number of drugs were being picked up and which didn't produce any satisfactory result. So X here is the random variable.

1. The current probability distribution is a Binomial Distribution
    a. The **total** number of trials are fixed i.e. 10
    b. Each trial is **binary**, i.e. $P(S)\ or\ P(B)$ Success or Failure
    c. The **probability of success** is **same** for each of the trial
2. Probability Calculation
    Here $n = 10$ Total number of trials and as per the formula of Binomial Distribution

$$P(X = r) = \binom{n}{r} p^r (1-p)^{n-r}$$

Now substituting the values

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

$$\Rightarrow \binom{10}{0} \frac{1^0}{5} \frac{4^{10}}{5} + \binom{10}{1} \frac{1^1}{5} \frac{4^9}{5} + \binom{10}{2} \frac{1^2}{5} \frac{4^8}{5} + \binom{10}{3} \frac{1^3}{5} \frac{4^7}{5} = \mathbf{0.879}$$ (Answer)

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.

Solution

As the time of effect is a **Continuous Random Variable.** So, the current distribution will follow a normal distribution.
So here Population mean μ is not known that we must determine,
Sample Mean $\bar{X}$= 207, Sample Size *n* = 100 (which is >30 then the sampling Distribution must be a normal distribution) &
Sample Standard deviation *S* = 65
So as the distribution follows Normal distribution, the Central limit theorem (CLT) will be determined. As per CLT

1. Population mean μ = Sample Distribution Mean
2. Sampling Distribution Standard deviation
   Standard error $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \approx \frac{S}{\sqrt{n}} = \frac{65}{\sqrt{100}} = \frac{65}{10} = 6.5$
3. Sample Size > 30

Population mean is usually denoted with some margin of error, so resultant Population mean is

$$Confidence\ Interval = \mu = (\bar{X} \pm \frac{Z^* S}{\sqrt{n}})$$

For 95% Confidence level the Z^*, Z Score associated with 95% Confidence interval is 1.96
Substituting all the values:

$$Confidence\ Interval = \mu = (207 \pm (1.96 * 6.5)) = (207 \pm 12.74)$$
$$= (\mathbf{194.26}, \mathbf{219.74})\ (Answer)$$

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.

Solution

Hypothesis Testing is needed here.
Null Hypothesis $\boldsymbol{H_0 : \mu \leq 200}$
Alternative Hypothesis $\boldsymbol{H_1 : \mu > 200}$

Sample Mean $\bar{X}$= 207, Sample Size *n* = 100 & Sample Standard deviation *S* = 65
And here $\boldsymbol{\mu = 200}$
Significance Level $\alpha = 0.05$

So as the Alternate Hypothesis has > then this is an Upper Tailed Test

Critical Value Testing

Zc (Z Score at the critical value) = Z Score (1 – 0.05 = 0.95) = (1.64 + 1.65) / 2 = 1.645 [As 0.950 is not there in the z-table.]

Upper Critical Value (UCV) = $\mu + (Zc * \sigma_{\bar{x}}) = \mu + \left(Zc * \frac{\sigma}{\sqrt{n}}\right) = \mu + \left(Zc * \frac{S}{\sqrt{n}}\right)$

$$= 200 + (1.645 * 6.5) = 210.69$$

As $\boldsymbol{\bar{X}}$**= 207 < 210.69**, hence the value is in the Acceptance region, hence we **failed to reject the Null Hypothesis**. Hence Newer batch produce satisfactory result

P-Value Testing

Z Score = $\frac{\bar{X} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{207 - 200}{6.5} = \frac{7}{6.5} = 1.076 \approx 1.08$

Z Score Associated with 1.08 is 0.8599
p-value = 1-Zscore table = 1- 0.8599 = 0.14
Since p-value of the Z-score 0.14 (14%) is greater than the significance level 0.05(5%) hence **we fail to reject the Null Hypothesis.**

Question 3

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).

Solution

Type	Case - I	Case - II
α	0.05	0.15
β	0.45	0.15

In Hypothesis Testing

- **Type I** error happens when we reject the Null Hypothesis but is true, it's denoted by α
- **Type II** error happens when we failed to reject Null Hypothesis but in actual term Null Hypothesis is False, denoted by β

Let's define the Null Hypothesis

H_o: The Drug is not harmful, it's performing as required

H_1: The Drug is Harmful, and it has adverse effects

Now we must think from some POVs

1. Consequences of Type-I and Type-II errors
2. Company's and Customer POV
3. And Drug's Criticality

For Case I

Here the $\boldsymbol{\alpha}$ is $\boldsymbol{low}, \boldsymbol{\beta}$ $value\ is$ $\boldsymbol{high}$ so here the chances are the Alternative Hypothesis will come into consideration as there is a chance of making a **Type-II error**. So, if do that what will be consequences.

- Let's say the Drug is **Painkiller** so as it's less critical drug so it's adverse effect will not be so critical, so from a customer POV there won't be so much effect as well from Company there won't be any loss. So, Type II error for critical is feasible for Company.
- Let's say if the Drug is **Heart/Cancer** specific then the Criticality is too high as it can be a matter of life and death and that point of time Customer will be affected the most and as a result Company might face a ban or closure is some legal action comes into picture. So, Type II error for critical drugs is not at all feasible for Company.

For Case II

Here the $\boldsymbol{\alpha}$ is $\boldsymbol{equal}$ to $\boldsymbol{\beta}$ $but\ is\ is\ comparitively$ $\boldsymbol{higher\ than\ the\ previous\ case,}$ so here there is a chance of doing a Type-I error, why because $\alpha \propto \frac{1}{\beta}$, let's the above two scenario

- So, when the Drug is **Painkiller** and here the criticality of drug is less, so company can again do a testing on all the drugs by doing random/stratified sampling and doing the testing on top of it. Customer won't be affected at all. Type-I is good compared to Type-II.
- But when the Drug is **Heart/Cancer** specific, then Company must spend some more fortunes for testing. As they are critical drug and if it's comes to market Company might lose the customer base and might also face ban. Type-I is good compared to Type-II.

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 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.

Solution

A/B Testing is a direct industry application, of the **two-sample proportion test sample.**

A/B testing allows company to make changes in the UIs/existing version as per the customer choice, by collecting their data, which is make them understand which is going to perform better in the outside world.

So here we need a Control group and a Variation Group.

So, we would use two tag lines and let's see what happens

1. We will apply the first Tagline and roll it out in the market/website to some customers and we would ask for the feedbacks/survey so get their sentiment, if they have positive sentiment then they will select Yes and if they have the other opinion they will Select No. So, this will be live for some time and then all the data will be collated.
2. Parallelly for a Separate Customer group of people this second Tag Line will be applied and roll out to the website and similarly we will get the customer sentiment.
3. Then we must apply the A/B test, as it's a two-sample proportion test. So, we need to define a significance level $\boldsymbol{\alpha}$ and then we would calculate the **p-value,** if the p-values comes out to be less than $\boldsymbol{\alpha}$**,** then we reject the null hypothesis, then we have to proceed with the variation group if it's the other way round then we failed to reject Null Hypothesis and hence Control Group will be the one rolled out to the population.