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:

Total population of the new batch is 80,000. Since we cannot perform our analysis on this huge amount of data at once, we can take a small sample from this complete population data. For example take a sample of 10

Given:
If “x” is denoted as probability of an unsatisfactory result then “4x” is denoted as probability of a satisfactory result,
Hence, we can say that, probability of
satisfactory + unsatisfactory = 1
⇒ **x + 4x = 1** ⇒ x = ⅕ ⇒ x= 0.2 and 4x = 0.8

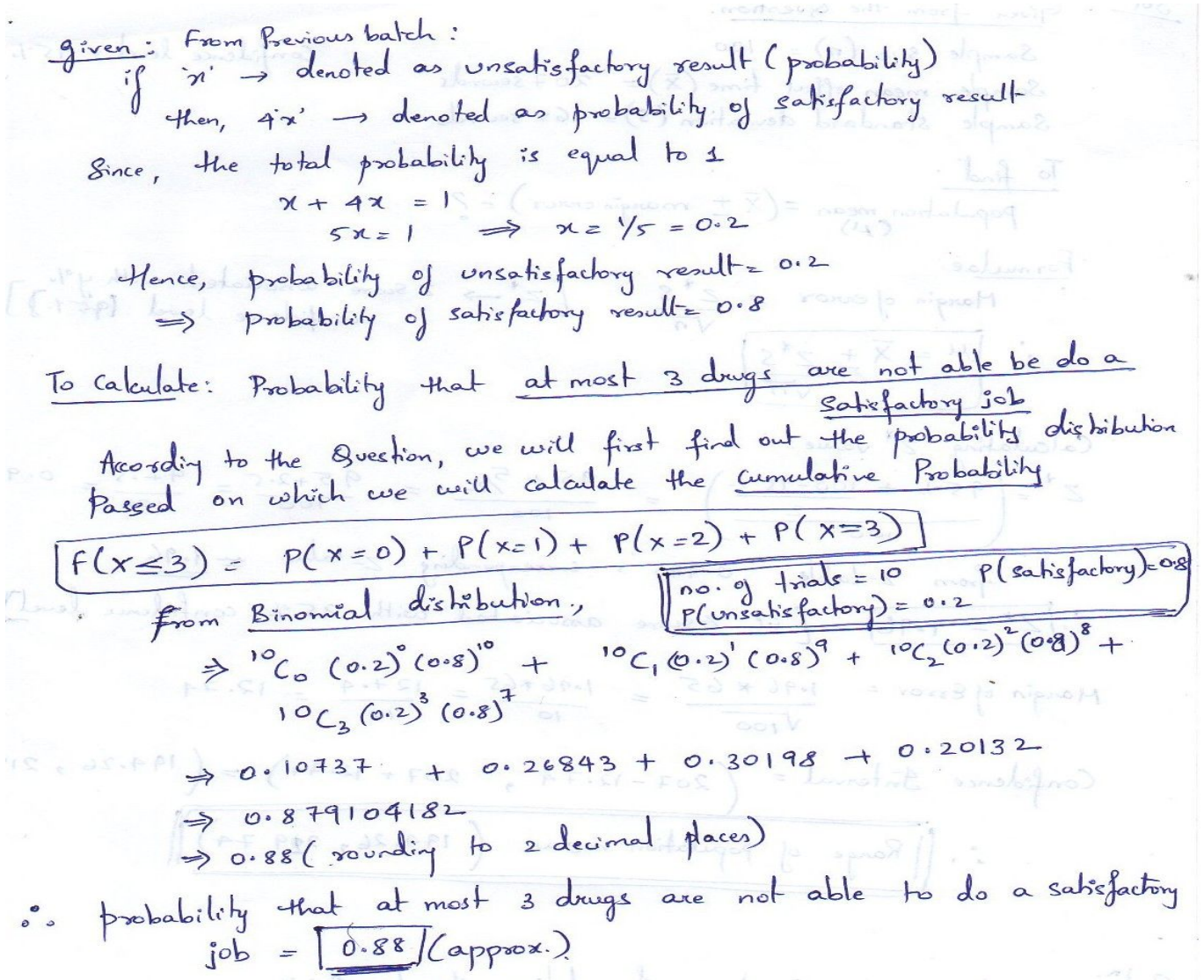
That is, from the previous batches of drugs, **Probability of a satisfactory result = 0.8** while the **Probability of unsatisfactory result =0.2**

a). **Binomial Distribution** would accurately portray the above scenario, this is mainly because **Binomial distribution is applicable in situations where there are fixed number of “Yes” or “No” questions with the probability of “Yes” or “No” remaining the same for all the questions.** In this case it is either “Satisfactory” or “Not Satisfactory” hence it is binary.

The three conditions that this distribution follows are:

- Total number of trials is fixed at “n”
- Each trial is binary i.e has only two outcomes (Success or Failure)
- Probability of success is same in all trials, denoted by “p”
- P(X=r) = P(Getting ‘r’ successes in ‘n’ trials) = nCr (p^r) [(1-p)^(n-r)]

First calculate the Probabilities and then calculate the Cumulative probability for at most 3 drugs which are not able to do a satisfactory job.



Conclusion:

Therefore, the probability of at most 3 drugs that are not able to do a satisfactory job = 0.879(approx.)

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

The Main Methodology:

- Take the sample size ‘n’
- Find the Mean and Standard deviation of the sample (which is already given)
- Compute the confidence interval for a y% (in this case 95%) confidence level

However these steps cannot be completed without the concept of CLT.

Following are the properties of CLT:

- Sampling distribution’s mean = population mean
- Sampling distribution’s standard deviation (standard error) = population’s standard deviation/ square root(sample size)
- For sample size greater than 30 the sampling distribution becomes normally distributed.

Hence, Central Limit Theorem lets us assume that the sample mean would be normally distributed with the mean and standard deviation (sigma/square root of sample size ~ approx S/square root of sample size) → Using which it is possible to find the Margin of Error, confidence Level etc.

Given from the Question:

Sample size (n) = 100 — Confidence level = 95%

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

Sample standard deviation (s) = 65 seconds

To find:

population mean (μ) = ($\bar{x} \pm$ margin error) = ?

Formulae:

Margin of error = $\frac{z^* s}{\sqrt{n}}$ [$z^* \rightarrow$ z score associated with y% confidence level [95%]]

$\therefore \boxed{\mu = \bar{x} \pm \frac{z^* s}{\sqrt{n}}}$

Calculating z^* value.

$z^* = \left(\frac{95\% + \frac{100-95\%}{2}}{100}\right) = \frac{95 + 5/2}{100} = \frac{95+2.5}{100} = \frac{97.5}{100} = 0.975$

from z-table, 0.975's corresponding z-value $\fallingdotseq$ 1.96

$\therefore \boxed{z^* = 1.96}$ [i.e z-score associated with 95% confidence level]

Margin of Error = $\frac{1.96 * 65}{\sqrt{100}} = \frac{1.96*65}{10} = \frac{127.4}{10} = 12.74$

Confidence Interval = $(207 - 12.74, 207 + 12.74) = (194.26, 219.74)$

$\therefore$ Range of population mean $(194.26, 219.74)$

Conclusion:

Therefore, Range of Population Mean: (194.26, 219.74)

Question 3:

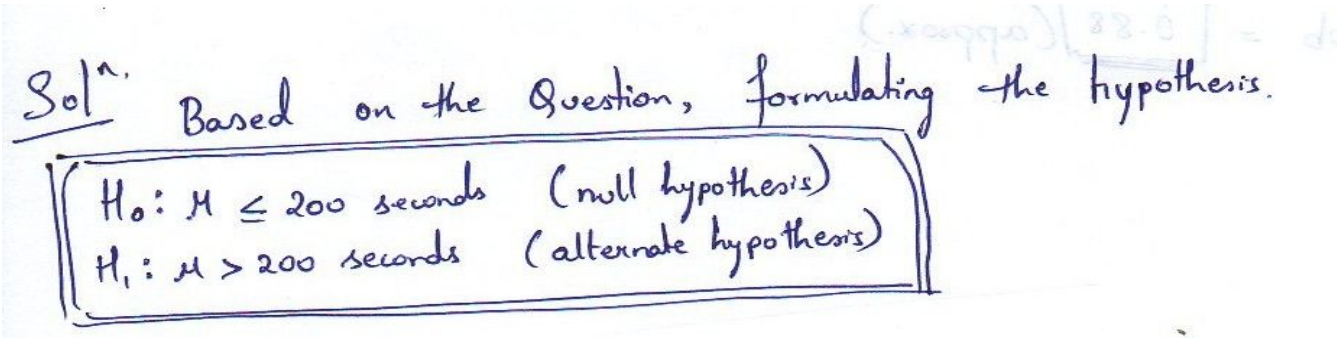
a) The painkiller drug needs to have a time of effect of at most 200 seconds to be considered as having done a satisfactory job. Given the same sample data (size, mean, and standard deviation) of the previous question, test the claim that the newer batch produces a satisfactory result and passes the 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.

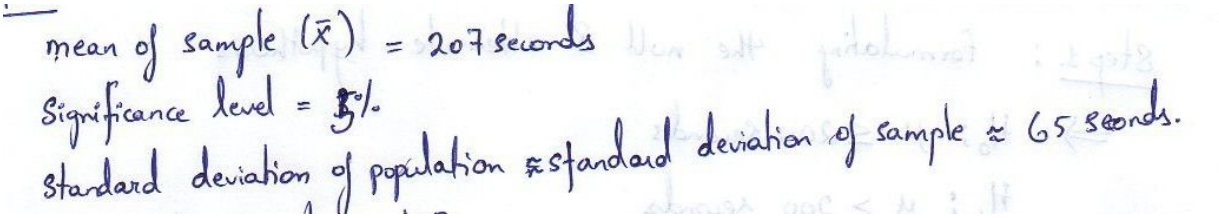
Now, a different sampling procedure(with different sample size, mean, and standard deviation) is proposed so that when the same hypothesis test is conducted, the values of α and β are controlled at 0.15 each. Explain under what conditions would either method be more preferred than the other, i.e. give 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

3a).

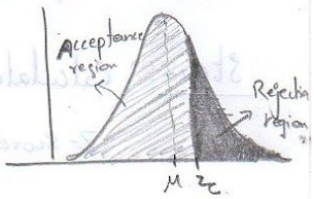


Given:



Method 1: Critical Value Method

Size of the sample $n = 100$



Method 1: Critical Value Method.

Step 1: Formulating the Null & Alternate hypothesis.

⇒ null hypothesis: $H_0: \mu \leq 200$ seconds
alternate hypothesis: $H_a: \mu > 200$ seconds

Step 2: Calculate the value of z_c from the given significance level

⇒ for 5% significance level, we would have 1 critical region, with the total area of 0.05, as it is a "1-tailed test" and the critical region is on the right side, more specifically, it is an "upper-tailed test"

∴ which means "0.05" is the total area of rejection,

Hence, total area of acceptance $= 1 - 0.05 = 0.950$

z-score corresponding to 0.950 [from z-table] $= 1.645$

Step 3: Calculating Critical Value.

⇒ $C.V. = \mu + \left(z_c * \frac{\sigma_{\bar{x}}}{\sqrt{n}}\right)$ [we use '+' sign since the critical value is on the right hand side]

$= 200 + \left(1.645 * \frac{65}{\sqrt{100}}\right) = 200 + 10.6925 = 210.6925 = 210.69$

Step 4: Make the decision on the basis of the value of the sample mean ($\bar{x}$) with respect to the critical value calculated.

⇒ $C.V. = 210.69$ seconds. $\bar{X} = 207$ seconds.

Since, sample means falls in the area of acceptance, hence,

"we fail to reject the null hypothesis"

Method 2: p-Value Method

Method 2: p-value method.

Step 1: formulating the null & alternate hypothesis

$\Rightarrow H_0: \mu \leq 200$ seconds

$H_1: \mu > 200$ seconds

Step 2: Calculate the z-score.

$$\text{Z-score} = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{207 - 200}{65/\sqrt{100}} = \frac{7 * 10}{65} = 1.0769$$

Step 3: find p-value.

Cumulative probability = (z-value of 1.0769) = 0.8577

Since, it is a 1-tailed test:-

p-value = 1 - cumulative probability

= 1 - 0.8577 = 0.1423. (approx.)

Step 4: Make the decision based on p-value and significance level.

p-value = 0.1423 significance level = 0.05

Since, p-value > significance level.

i.e. $0.1423 > 0.05 \Rightarrow$ "We fail to reject the null hypothesis"

Conclusion:

Hence, we can conclude from Critical Value Method as well as p-Value Method that we would fail to reject the null hypothesis.

3b).
H0:- The drug produces satisfactory result
H1:- The drug doesn't produce satisfactory result

The term **alpha represents Type-1 error** which implies that even though the null hypothesis is True, but we reject the null hypothesis, that is, in this case even **if the drug produces satisfactory result but we conclude that drug doesn't produce a satisfactory result.**

Whereas the term **beta represents Type-2 error** which implies that even though the Hypothesis was false, we fail to reject the null hypothesis, that is in this case even **if the drug doesn't produce satisfactory result but we conclude that it produces satisfactory result.**

From the above we can say that making Type-1 error would be very inefficient than making Type-2 error, since the consequences of concluding that the drug doesn't produce satisfactory result even though it does produce satisfactory result will lead to false measurement taken further, like increasing the dosage of the painkiller, even though the amount taken was sufficient for the body which further leads to severe side effects.

Hence we need to make the probability of type-1 error as low as possible and increase the probability of type-2 error.

Case 1:
Alpha = 0.05 an Beta = 0.45
This case concludes that the probability of having type-2 error is more than the probability of making type-1 error which is a safe step although.
Hence, in other words, probability of saying that the drug produces satisfactory result even though it doesn't is safe as it will have less consequences that the drug will not be efficient for the person.
But this will be **more dangerous in terms of the Company**. Since the probability of making type-2 error is so high that the consumer will be disappointed with the drug and may shift to another drug that may give more efficient results and hence the company would be losing their customers, leading to **loss of the company**.

Case 2:
Alpha = 0.15 and Beta = 0.15
This case has equal probability of having both types of errors. Although this is not as safe as the previous step. The consequences are that it would be **dangerous for the Consumer**. As we would be making type-1 error more.

We can conclude that, Case 1 is Dangerous for the Company. Where as Case-2 is Dangerous for the Consumer.

The Company could use Case 2 for conducting the hypothesis testing. As it may favour Consumer and Company equally.

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.

Answer

A/B Testing is like a research method which allows us to take up **two different variants** of any particular application for example webpage and then **find out which among them is more effective** for the on-going use-case. **---> What is A/B Testing**

Hence, to conduct A/B testing we require two sets of versions to be tested for example A version and B version. For each version equal number of visitors would be exposed. At the end, we can check as to which page has got more of the targeted actions like registrations, purchases or page views etc. based on which kind of webpage that needs to be used in future will be decided upon. **---> How to do A/B Testing**

Hence we can say that A/B testing helps in understanding the changes that needs to be made which will lead to more profit to the company. A/B testing is a way to do Hypothesis Testing in other terms. **---> What can we infer from the results of A/B testing.**

Procedure of A/B testing:

1. The first and foremost step of A/B testing is Formulating the Hypothesis
2. Identify Targets
3. Select One test Item
4. Determine the test sample size
5. Determine the duration of the experiment
6. Finally do A/B testing

For example, according to the question, the marketing team needs to sell the newly processed batch after Quality test, since two taglines are proposed and two teams were assigned one each to do A/B testing, following steps can be taken:

- Hypothesis would be to attract a new set of customers for the product
- Marketing team needs to get the data on the target customers for the product
- They would have to select one quality of the analysis that may be considered for the evaluation purpose it could be “Price”, “Competitors”, etc.
- Determine the sample population size they would like to choose to test this upon.
- Also determining the duration of this experiment.
- Then try and perform A/B testing.

At the end of this experiment, they will be able to get an idea on how the product will perform and what could be some changes that may be made to improve the consumption of the product.