Question-1

The quality assurance checks on the previous batches of drugs found that — it is 4 times more likely that a drug can 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 would portray the above scenario is the Binomial Distribution.

The various conditions this distribution follows are:

- The total number of trials are fixed.
- Each trial is binary, i.e. it has only two possible outcomes(Success or Failure).
- The probability of success is same for all the trials.

b) Calculate the required probability.

According to the statement "it is 4 times more likely that a drug is able to produce a satisfactory result than not.",

Probability of the drug being not satisfactory = x

Probability of the drug being satisfactory = 4x

Total probability = 1

x+4x=1

5x=1

x=0.2

Probability of the drug being not satisfactory -> **p(ns) = 0.2**

Probability of the drug being satisfactory -> p(s) = 4*x=s*0.2=0.8 => **p(s) = 0.8**

Sample size -> **n=10**

For finding the theoretical probability that at most, 3 drugs are not able to do a satisfactory job,

F(3)=P(X<=3)=P(X=0)+P(X=1)+P(X=2)+P(X=3)

Using the formula for binomial distribution,

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

F(3) = $\binom{10}{0}0.2^0 0.8^{10}$+ $\binom{10}{1}0.2^1 0.8^9$+ $\binom{10}{2}0.2^2 0.8^8$+ $\binom{10}{3}0.2^3 0.8^7$

F(3)=(1*1*0.107) + (10*0.2*0.134) + (45*0.04*0.167) + (120*0.008*0.209)

F(3) = 0.107+0.268+0.3006+0.200

F(3) = 0.8756

The theoretical probability that at most 3 drugs are not able to do a satisfactory job is 0.8756

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.

Answer:

We can use the Central Limit Theorem to approach this problem.
The properties of this theorem are:

- The sampling distribution's mean is equal to the population mean. $\mu_{\bar{x}} = \mu$
- Sampling distribution's standard deviation or the standard error = $\frac{\sigma}{\sqrt{n}}$ where σ is the population's standard deviation and n is the sample size.
- For n>30, the sampling distribution becomes a normal distribution.

b.) Find the required range.

According to the statement,
Sample size, n=100
Sample mean, $\bar{X} = \mu_{\bar{x}} = 207$
Sample standard deviation, $\sigma_{\bar{x}} = 65$ = S
Confidence Interval=95%=0.95
For 95% confidence interval, Z* = 1.96
Finding the interval for the sample with 95% confidence interval,
μ = $(\bar{X} - \frac{Z*S}{\sqrt{n}}, \bar{X} + \frac{Z*S}{\sqrt{n}})$
μ = (207-12.74,2.7+12.74)
μ = (194.26,219.74)

The above is the confidence interval for 95% confidence.

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:

According to the problem statement,

Null Hypotheses: $H_0 \Rightarrow \mu_x \leq 200$
Alternate Hypotheses: $H_1 \Rightarrow \mu_x > 200$
Sample Size, n=100
Sample Mean, $\mu_x = 207$
Sample standard deviation, $\sigma_x = 65$
Significance level, α=5%=0.05

Firstly, let us use the Critical Value method to test the hypothesis.
The test would be a right tailed test.
Acceptance region will be 1-0.05=0.95=95%
For 95%, Z_c=1.645
$\mu_{\bar{x}} = \mu = 200$
$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{65}{10} = 6.5$
Critical Value = $\mu + (Z_c * \sigma_{\bar{x}})$
Upper Critical Value= 200+(1.645*6.5) =210.69
Lower Critical Value= 200-(1.645*6.5) =189.31
207(sample's mean) lies within the critical value region, we fail to reject the null hypotheses.

Now let us use the p-Value method to test the hypotheses.
$\mu_{\bar{x}} = \mu = 200$
$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{65}{10} = 6.5$
$Z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{207-200}{6.5} = 1.07$
The cumulative probability of 1.07 is 0.8577 from the Z – table.
p-Value=1-0.8577=0.1423.

As the p-Values is greater than α (0.05), we fail to reject the null hypotheses.

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:
Type 1 error occurs when we reject a true null hypothesis. Smaller values of α make it harder to reject the null hypothesis. So, choosing lower values for α can reduce the probability of a Type I error. But β should also be small because there might be cases when the null hypotheses might be false but we fail to reject it. The probability of such event should be very low. 0.15 is less that 0.45. So, I would prefer the second sampling procedure.

Failing to reject the null hypotheses when it is false would be a very dangerous scenario as it will impact the lives of many people. So, keeping the value of β as low as possible decreases this risk.

While doing a quality control survey in a noodles factory, our null hypothesis would be that the maximum percentage of lead content in the noodles is 6%. Since we want to maximize the company profits and we do not want to discard too many packed units, we set the value of α low but now very low. This is the producer's risk. But if we are a consumer, and don't want to get damaged products, and the null hypothesis would be false. Though we want to reduce the risk of customer, we keep a value of β as low as possible. So, compromising with the above two situations, keeping the values of α and β same can help.

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 a method of comparing two options against each other to determine which one performs better. Before releasing the product to the market, we need to decide which option would work better for the whole population. Before releasing it to the overall population, better is we check the effect on a smaller population. The results are then analyzed, and the better option is imposed on the overall population.
The process of A/B Testing for the campaign:

1) Create the two taglines that are supposed to be tested in the campaign.
2) Randomly select places and place ads for the campaigns. The number of ads for both the taglines should be the same. Both the ads should be placed in the areas with same population region.
3) Analyze the results of the campaigns for both the options.
4) Use the ad that gave best results.