STATISTICS ASSIGNMENT

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.

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, which exactly satisfies the given condition, is Binomial distribution. Because it satisfies the three conditions that, the distribution follows.

i. Fixed number of trails (N=10)
ii. Each trail is independent of others.
iii. There are only two outcomes (success or failure)

b) *Calculate the required probability.*

Answer:

Let S= be the event of drug with satisfactory results

NS = be the event of drug with not so satisfactory result.

Given, P(S) = 4 times more likely to produce satisfactory result:

Say if x is the probability of producing not satisfactory results, then, 4x is the probability of producing satisfactory results.

P(S) =4x; P (NS) =x

We know the Probability of producing not so satisfactory drug is:

P(S) + P (NS) = 1

=> 4x +x =1

Therefore, our probability of success P is, probability of drug not producing satisfactory result

= > P =1/5; 1-P = 4/5

Now, the formula for finding Binomial probability is below.

$\boxed{\text{STATISTICS ASSIGNMENT}}$

$\boxed{\text{Question 7(b)}}$ Answer: Calculate the Probability?

formula to calculate Binomial Distribution

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

we have $P = 1/5 = 0.2$;

$1-P = 4/5 = 0.8$;

to calculate the probability of 3 drugs not providing Satisfactory results, lets assume 'x' is the event & 'x' be the no of drugs not satisfying the satisfactory results.

Now, we need to calculate Probability of 3 $\boxed{\text{atmost 3 drugs}}$ not producing satisfactory result.

i.e $P(X \leq 3)$.

The probability of atmost 3 drugs not producing satisfactory results is given by?

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

Applying in the Binomial distribution formula; where $\boxed{n=10}$; $\boxed{P=1/5}$ $\boxed{1-P=4/5}$

$P(X=0) = {}^{10}C_{0}\,(1/5)^{0}\,(\frac{4}{5})^{10-0} = 1\times 1\times 0.1073741 = 0.1073741824$

$P(X=1) = {}^{10}C_{1}\,(1/5)^{1}\,(4/5)^{10-1} = 10\times 0.2\times 0.134217 = 0.268435456$

$P(X=2) = {}^{10}C_{2}\,(1/5)^{2}\,(4/5)^{10-2} = 45\times 0.04\times 0.167772 = 0.301989888$

$P(X=3) = {}^{10}C_{3}\,(1/5)^{3}\,(4/5)^{10-3} = 120\times 0.008\times 0.209715 = 0.201326592$

$\boxed{P(X\leq 3) = 0.879126 \simeq 0.88}$

This is the theoritical probability of atmost 3 drugs not able to do a Satisfactory Job.

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.

2 (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 mail methodology or approach to the problem would be Central Limit theorem. We use Central limit theorem since we cannot find the mean and St. Deviation of population due to time/money constraints. That is using CLT we infer the population mean from the sample mean.

1. Properties of CLT:

    a. Sampling Distribution mean (μx) = Population mean (μ).

    b. Standard Error= σ/√n (σ is population deviation and n is no of samples)

    c. For n>30, sample distribution becomes normal distribution. i. (In our case n=100 which is >30.)

Therefore, we take a small sample of the population; calculate the mean and St. Deviation with some margin of error. Now in the given problem, sample size, mean time and St. Deviation are given and we have to calculate the confidence interval in which the mean lies for confidence level 95%.

2(b) Find the required range.

$\boxed{\text{Question 2(b)}}$ → find the required range.

(Sol'n)

from given problem we have,

Sample mean $(\bar{x}) = \cancel{10.0}$ 207 sec

Sample Size $(n) = 100$

Sample St. deviation $(s) = 65$ sec

formula to calculate Confidence Interval $= \left(\bar{x} - \dfrac{z^* s}{\sqrt{n}},\ \bar{x} + \dfrac{z^* s}{\sqrt{n}}\right)$

$\boxed{z^* \text{ for confidence interval } 95\% = 1.96}$

Substituting the values $= \left(207 - 1.96 \times \dfrac{65}{\sqrt{100}},\ 207 + 1.96 \times \dfrac{65}{\sqrt{100}}\right)$

$= (207 - 1.96 \times 6.5,\ 207 + 1.96 \times 6.5)$

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

$\boxed{\therefore \text{ Confidence Interval } = (194.26,\ 219.74)}$

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: 3(a)

1. Formulate Null and Alternate hypothesis: In the given question, the claim is new batch time of effect is at most 200sec to be considered as drug doing satisfactory job. Therefore, below hypothesis is now constructed as follows.

    a. Null Hypothesis- H0: P (x ≤ 200 sec)

    b. Alternate Hypothesis – H1: P(x> 200 sec)

2. Calculate using two Hypothesis testing methods:

    1. **Critical value method.**
    2. **P-Value method.**

3. Given values:

    a. Sample size n=100
    b. Sample mean (x)= 207 sec
    c. Sample St. Deviation = 65 sec.
    d. Significance level α=5%

Critical value Method:

To make a decision using critical value method after formulating the hypothesis is as follows:

1. Calculate the value of Zc from given value of α=5%.
2. Calculate the critical values from Zc from Z table
3. Make a decision based on sample mean value x referring to critical value.

STATISTICS ASSIGNMENT

Question 3(a): Calculate two hypothesis testing methods.

Answer ⇒ (1) **Critical Value Method**

we have, Null Hypothesis : H_0 : $P(X \leq 200sec)$

Alternate Hypothesis : H_1 : $P(X > 200sec)$

(i) Calculate value of Z_c (Z critical) for significance level $\boxed{\alpha = 5\%}$

$\alpha = 5\% = 0.05$

Here the critical region of critical point is on the right side i.e 0.05 & since it is a Upper tailed test as Alternate hypothesis H_1 : $(P(X) > 200 sec)$.

∴ cumulative probability of critical point

$= 1 - 0.05 = 0.950.$

(ii) Also, from Z table for 0.950 is not there, so we have to take a Average of close values like $0.9495 = 1.64$ & $0.9505 = 1.65$, which account for $\boxed{1.645}$ ⇒ $\left(\frac{1.64 + 1.65}{2} = 1.645\right)$

$\boxed{Z_c = 1.645}$

(iii) Now the critical value can be calculated from below to make a decision is,

formula of $C.V = \mu + Z_c (\sigma/\sqrt{n})$

In the formula $\sigma_{\bar{x}}$ i.e Sample st. deviation is given & not for population. In such cases we can approximately assume population st. deviation as sample st. deviation

i.e $\sigma \simeq \sigma_{\bar{x}} = 65$ sec.; $\mu = 200$ sec $Z_c = 1.645$

$CV = 200 + 1.645 \times \frac{65}{\sqrt{100}} \Rightarrow 200 + 10.69 = \boxed{210.69}$

$\boxed{CV = 210.69}$

Decision: Since the sample mean is less than the critical value we fail to reject the null hypothesis

P-Value Method:

To make a decision using P-Value method after formulating hypothesis H0 and H1, below are the calculations performed:

1. Calculate the value of the Z-score:
2. Calculate the P-Value from cumulative probability of the given Z-score using Z-table.
3. Make a decision based on the P-value referring to significance level α.

STATISTICS ASSIGNMENT

② P- Value Method

There are three steps involved in making a decision using p-value method:

① Calculate Z score :-

for Sample mean ($\bar{x}$) 207 sec,

formula to find Z score = $\dfrac{(\bar{x}-\mu)}{(\sigma/\sqrt{n})}$

$$\text{Zscore} = \frac{(207-200)}{(65/\sqrt{100})} = \frac{7}{6.5} = 1.077 \simeq 1.08$$

$$\boxed{\text{Z Score} = 1.08}$$

Notice that as the sample mean lies on the right side of the hypothesis mean, the Z-score is positive.

② Find P-value :-

The p-value for the Z score 1.08 from Z-table fo 1.0 on the vertical axis & 08 on horizontal axis is $\boxed{0.8599}$

- Since the Sample mean lies on right side of the distribution & it is a Upper tailed test $(H_1 : P(x>3))$, the p-value wouldbe $1 - 0.8599 = 0.1401$

$$\boxed{P = 0.1401}$$

③ Make a decision :-

Since p-value is greater than the significance level $(0.1401 > 0.05)$. Higher p-value gives less evidence to reject null hypothesis.

Decision : We fail to reject the null hypothesis

Question: 3(b)

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:

From the question below table with two cases can be taken:

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

Hypothesis. From the given problem:

1. Type I error: Reject a true Null hypothesis
2. Type II error: Fail to reject a false Null

Case 1: α = 0.05; β=0.45

We can consider a condition where the criticality of the drug is low:

1. I.e. the consequences of committing type-I error is low to consumer (which also means consumer do not have huge health side effects in committing a type-II error),
2. However, the company incur huge losses by committing type-1 error on the production, then we can prefer Case –I.

Case 2: α = 0.15; β =0.15

In case 2 we can consider a condition where the criticality of the drug is high, we need to keep both type-I and type-II error low.

1. I.e. the consequences of committing a type-II error (fail to reject false null hypothesis) is high to the consumer as well as the company as the drug will have severe side effects on consumers and company faces serious consequences from regulatory authorities.
2. In addition, consequences of committing a type-I error is also high to company where the company incurs huge losses on the production.

Summary:

In my opinion, I would prefer **case II where both α and β are 0.15** in this drugs scenario.

The reason being:

I. Here in this case the error value is minimum for both the type of errors, which would be beneficial for both the company as well as the consumer.
    a. That is type –I error (α) – Rejecting a true null hypothesis would incur loss to the company, and type-II error (β) - Fail to reject a False Null hypothesis would incur loss to consumer.

Therefore, keeping both minimum would be ideal for both company and consumer.

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 widely used in scenarios where we have to compare two versions of entities such as taglines, colors, webpages, etc. while deciding a marketing strategy.

Firstly, since there are two tag lines proposed by the marketing team for launching of the new drug in to the market, we need to identify which tag line is performing better. In such cases to identify, which tag line is getting consumer acceptance and conversion rate to business to the company, we use A/B testing.

Step 1:

We run both the tagline ads on social media platform like Facebook, Twitter etc. Divide the consumer traffic seeing tagline -1 into Group A and the consumer traffic seeing tagline -2 into Group B.

1. Group-A with tag line 1
2. Group-B with tagline 2

Step 2:

To find out which tagline performed better with the consumers. We will now need to see the frequency of customers in Group 1 and Group 2 and see who actually showed interest in buying the drug.

Step 3:

We now have the frequencies of both the taglines to get a statistical evidence
Of which tagline is performing better, we use Two-proportion test.

Step4:

In two-proportion test using XLT in excel, we can calculate A/B testing statistical to determine which tagline is performing better based on the frequency of user traffic we have.

We need to formulate our Null and alternate Hypothesis, which would be as follows.

Null Hypothesis H0-Tagline 1 = Tagline 2 (which means both are performing equally better)

Alternate Hypothesis: H1: Tagline 1 ≠ Tagline 2

We will then need to select two-proportion test in XLT, provide Frequency and sample size for both the groups and set the alternate hypothesis and determine which tagline is performing better or in other words if our null hypothesis is rejected or fail to reject.