Statistics Assignment

VinayKumar V

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.

For this problem, its **Binomial Distribution**, it's a discrete probability distribution and its statistical experiment has the following properties,

1. Consists of n repeated trails
2. Each trial can result in just two possible outcome (which is binary form), we call one of these outcomes as success (I am taking Not Satisfactory with Drug as success) and failure (Satisfactory with Drug)
3. The probability of success is denoted by p (in this problem success is getting not satisfactory with drug) and failure denoted q=1-p (which indicates satisfactory with drug)
4. The trails are independent, that is outcome of one trail does not affect the outcome of the other trail

Binomial Distribution: $P(X = r) = n_{C_r} p^r q^{n-r}$

where '**r**' is number of success which are required

'**n**' is number of independent trails

'**p**' is probability of success in each trail = 0.2 (which represents not satisfactory (1/5 times))

'**q=1-p**' is probability of failure in each trail = 1-0.2 = 0.8 (which represents satisfactory (4/5 times))

b.) Calculate the required probability.

1 b. probability that at most, 3 drugs are not able to do a satisfactory job

i.e $P(r \leq 3) = P(r=0) + P(r=1) + P(r=2) + P(r=3)$

$$P(r \leq 3) = \left[{}^{10}C_0 (0.2)^0 (0.8)^{10-0}\right] + \left[{}^{10}C_1 (0.2)^1 (0.8)^{10-1}\right]$$

$$+ \left[{}^{10}C_2 (0.2)^2 (0.8)^{10-2}\right] + \left[{}^{10}C_3 (0.2)^3 (0.8)^{10-3}\right]$$

$$P(r \leq 3) = 0.107 + 0.268 + 0.301 + 0.201$$

$$P(r \leq 3) = 0.877$$

$$\simeq 87.7\%$$

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)

Sampling Distribution and the Central Limit theorem:-

properties of sampling distributions of sample Means.

(1.) The mean of the sample means, $\mu_{\bar{x}}$ is equal to the population mean, i.e., $\boxed{\mu_{\bar{x}} = \mu}$

(2) The standard deviation of the sample means $\sigma_{\bar{x}}$ is proportional to σ:

Standard error of Mean $\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$ [$\because$ when $n \leq 0.05N$]

2 b.

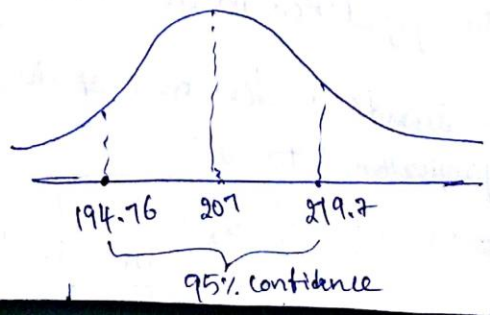
- Sample of 100 drugs was taken
  i.e $n = 100$
- The Mean time of effect was 207 seconds
  i.e $\bar{x} = 207$ sec
- Standard deviation coming to 65 seconds
  $\sigma = 65$ sec
- Confidence level 95%
  $\alpha = 95\%$ $(\because Z_\alpha = 1.96)$
- Confidence Interval for true population Mean is
  (where C.I = Confidence Interval)

$$C.I = \left(\bar{x} - Z_\alpha \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_\alpha \frac{\sigma}{\sqrt{n}}\right)$$

i.e $\left(\bar{x} - Z_\alpha \frac{\sigma}{\sqrt{n}}, \bar{x} + Z_\alpha \frac{\sigma}{\sqrt{n}}\right)$

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

Confidence Interval for population mean $= (C.I) = (194.76, 219.74)$



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 hypothesis test conditions (sample size, mean, and standard deviation), the value of α and β come out to 0.05 and 0.45 respectively.

Now, a different sampling procedure 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.

3

a.

- The pain killer drug needs to have a time of effect at most 200 seconds to be considered as having done a satisfactory job

  i.e $\mu \leq 200$ → Satisfactory

  $\mu > 200$ → NOT satisfactory

i) ~~Method 1: p-value~~ Method :-

$n = 100$
$\bar{\mu} = 207$ sec
$s = 65$ sec
$\mu = 200$ sec

- Null Hypothesis (H_0) : $\mu \leq 200$

  (effect of drug at most 200 sec which is satisfactory)

- Alternative Hypothesis (H_1) : $\mu > 200$

  (drug not satisfactory)

  (Right tailed test)

- Level of significance $\alpha = 0.05$ (5%)

- $Z_c = \dfrac{\bar{\mu} - \mu}{\sigma/\sqrt{n}} = \dfrac{207 - 200}{65/\sqrt{100}} = 1.08$

$Z_c = 1.08$

$Z_{table} = 0.859$

$p = 1 - 0.859 = 0.141 \simeq 14\%$

$\therefore 14\% > 5\%$

failed to reject Null Hypothesis (H_0)

i.e which is satisfactory

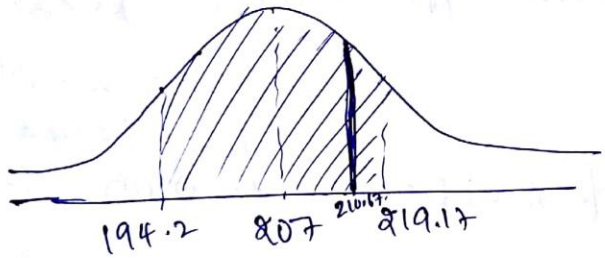
Method 2:- Critical Values Method:-

- Null Hypothesis (H_0): $\mu \leq 200$ (satisfactory)
- Alternative Hypothesis (H_1): $\mu > 200$ (not satisfactory)
  (Right tailed Test)
- Level of significance $\alpha = 0.05$

$$\text{Critical value} = \mu + Z_\alpha \frac{\sigma}{\sqrt{n}}$$

$$= 200 + 1.64 \times \frac{65}{\sqrt{100}}$$

Critical Value = 210.67



∴ The critical value lies in Acceptance region

Sample Mean 207 is less than critical value.

∴ ~~We reject~~

- We failed to reject null Hypothesis (H_0)

3b.

If $\alpha = 0.15$, $\beta = 0.15$

(i) Critical Value Method

For $\alpha = 0.15$ $Z_c = 1.036$

$\therefore$ Critical value $= 200 + Z_c \times \frac{\sigma}{\sqrt{n}}$

$= 200 + 1.036 \times \frac{65}{\sqrt{100}}$

$= 206.734$ sec

$\therefore$ we failed to reject null hyp.

(ii) P-value Method

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{207 - 200}{65/\sqrt{100}} = 1.0769$$

P-Value $= 1 - 0.8592 = 0.1408$

$= 14.08\%$

$\therefore 14.08\% < 15\%$ $(\because 15\% = 0.15)$

$\therefore$ we reject the null hypothesis H_0.

- Scenario 1: Let take, we have $\alpha = 0.05$ and $\beta = 0.75$

In the event that season of Impact is over 0.05 ($\alpha = 5\%$) then ordinary time of impact it might be not kidding influence to patients, so sun pharma should be watchful that season impact will be on limit, $\beta = 75\%$ won't influence to patients, so time of impact will be in the scope of -75% to 15%

- Let's take $\alpha = 15\%$ and $\beta = 15\%$ (equal Values), on the off chance that understanding had mishap or something not clear, it won't influence, therefore Expand the portion of parameter

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 for its existing 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.

A/B Testing is used to make business decisions based on the results derived from data, instead of just making predictions. It allows you to create variations of your website or app and then helps you to confirm or discard your decision to make changes.
This testing allows you to optimize your site or app in such a way that it increases the conversion rates. A higher conversion rate means getting more value from your existing users instead of having to pay more on traffic acquisition. It helps you to use mathematical data and statistics to determine the direction of your product variations (tag line in this testing).
Here, we have to do A/B testing as it can be used to decide which option (tag line) is more effective. A/B Testing also enables you to gain maximum from your existing traffic on a webpage. The cost of increasing your conversion is minimal as compared to the cost of setting up the traffic on your website. The ROI (return on investment) on A/B Testing is huge, as

1. Determine your goal

Before you create a test, you need to know what, exactly, you're hoping to accomplish.

2. Decide which page to test

Once you've determined a goal you want to accomplish, you'll need to select a page to start optimizing.

3. Select elements to A/B test

At this point, you should know which page on your site you'll be testing, and what you hope to accomplish with your test.
Now, you need to decide exactly *what* you want to test on that page.

4. Design your test

By this point, you've set a goal, selected a page, and determined what you want to test.

5. Let your test accumulate data

Once you've launched your test, you need to let it run for a long enough period of time to collect significant data.
This is often the most difficult part of running an A/B test for many site owners.

6. Document or implement your results

Regardless of your results, you have some work to do at the end of each test.