V.K. Sham Kumar
Dec 23rd 2018

Statistics Assignment

Problem Statement:

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).

Questions

Q1. 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.

"Cumulative Probability Distribution" portrays the above scenario more accurately.

Conditions:

- Total number of trials is fixed at n
- Each trial is binary, i.e., has only two possible outcomes - success or failure
- Probability of success is same in all trials, denoted by p

b.) Calculate the required probability.

n = 10

Probability of producing satisfactory results = 4/5

Probability of not producing satisfactory results = 1/5

r = 3

$P(x=r) = n \, C \, r(p)r(1-p)n-r$

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

$= 10 \, C \, 3(1/5)\verb|^|3(1 - 1/5)\verb|^|10\text{-}3$

= **0.8791 or 87.9%**

Q2. 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.

We have been provided the Population total, which is 80,000. We now need to find the range of the Population mean with 95% confidence Interval. We can approach this problem using the Central Limit Theorem. Using this method, we can find the Population mean with margin of errors.

Properties of this methodology is as follows:

- Sampling distribution mean (Mu X bar) = Population mean (Mu)
- Sampling distribution standard deviation = Sigma/Sqrt(n)
- For a sample > 30, the sampling distribution becomes a normal distribution

b.)Find the required range.

Sample Size (n) = 100
Sample Mean ($\mu \bar{X}$) = 207
Standard Deviation (σ) = 65
Confidence Interval = 95%

Range = ($\bar{X}$ - Z*S/√n, $\bar{X}$ + Z*S/√n)

Range = (207 – 1.96*65/√100, 207 + 1.96*65/√100)

Population Mean Range = 194.26 to 219.74

Population Mean Range lies between 194 and 220 seconds

Q3. 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.

Critical Value Method:

Null Hypothesis H0 <= 200

Alternate Hypothesis H1 > 200

μ = 200

n = 100

$\bar{X}$ = 207

σ = 65

α = 5%

Type of Test = One Tailed Test

Cumm. Probability of the Critical Point = 0.950

Zc = 1.645

CV = μ + Zc * ($\sigma/\sqrt{N}$)

   = 200 + 1.645(65/$\sqrt{100}$)

   = 210.69

Since $\bar{X}$ (207) is less than the critical value (210.69), $\bar{X}$ lies in the acceptance region and we fail to reject the null hypothesis

P Value Method:

Null Hypothesis H0 <= 200

Alternate Hypothesis H1 > 200

Z score for the sample mean of 207 = ($\bar{X}$ - μ) / ($\sigma/\sqrt{n}$)

                  = (207 – 200)/(65/$\sqrt{100}$)

                  = 1.08

P Value for 1.08 = 0.8599

        = 1 – 0.8599

        = 0.1401

Since the P value (0.1401) is greater than the significance level (0.05), we fail to reject the null hypothesis.

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.

In the above case of α 0.05 and β 0.45,

- Type 1 error occurs when we reject the null hypothesis stating that the mean time of effect of the drug is > 200 seconds, when it is actually not.
- Type 2 error occurs when we fail to reject the null hypothesis stating that the mean time of effect of the drug is < = 200 seconds, when it is actually not.

Situation for Sample 1 (α 0.05 and β 0.45):

When a scientist conducts a research to compare 2 medicines and check if the effect of both the medicines are same and there is no change between them. In this situation,

- H0 will be "Both the medicines effect are the same"
- H1 will be "Both are not the same"
  - Type 1 error occurs if the scientist rejects H0 stating that the drugs are not the same, when they are actually the same.
  - Type 2 error occurs, when the scientist fails to reject the H0 stating that the drugs are the same, when it is actually not.

For this situation, we should prefer the first sample with α 0.05 and β 0.45. Because we would have failed to reject the null hypothesis stating that they are same(P value of 0.14 > significance level of 0.05). In this case, even if Type 1 error occurs, it would not have caused a big harm to the patients, because both the medicines are the same.

Similarly, Type 2 error would have created a potential harm to the patients as both the drugs are different.

Situation for Sample 2 (α 0.15 and β 0.15):

When a Jury is going to give the judgement of a Criminal trial. Here:

- H0 will be "Defendant is innocent"
- H1 will be "Defendant is not innocent"
  - Type 1 error occurs if the Jury convicts the defendant even though he is innocent.

- Type 2 error occurs, when you fail to reject a True null hypothesis. Jury let's the accused walk free

For this situation, we can prefer the sample 2, because we would have rejected the null hypothesis and letting the defendant walk free.

In this case, Type 1 error occurs when the Jury rejects the True null hypothesis and convicts the defendant, even when he/she is innocent. This would have been a gross miscarriage of justice.

Q4.

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 statistically prove that any new ideas/strategies will have an impact to the process or not. It helps to learn better why a certain elements will have an impact in the user behaviour. Testing one changes at a time will help us to pin point which changes had an effect on the user behaviours.

By testing these 2 taglines, marketers can learn which one among the two will attract more clicks and conversion rates. Without testing, the new Tag lines could have been a lot of risk in losing the customers. This would have been a huge loss in terms of revenues and acquiring new customers. Hence, A/B testing can be used to make sure that the right Tag line is selected and deployed to the entire audience.

Procedure:

- Collecting Data:
  - This step helps in identifying the pages or areas in the web page which has the high visibility rates, so that the new TAG lines can be launched.
- Identify the goals:
  - This step helps us to identify the metrics to determine the success of the Tests. E:G – More number of conversion rates, More clicks to the link, etc.
- Run Experiment:
  - Kick off the experiment by launching these 2 tag lines and randomly assigning the customers to one of these at a time. Their interactions with each experience is measured, counted and compared to determine how each performs.
- Analyse Results:
  - Once the experiment is completed, it's time to analyse the results. We can set a specific time period (30 Days, 60 Days etc) to do the testing. There are

many software which will present the data from the experiment and show the difference between how the two versions performed. If one of the Tag line is the winner, launch the winning creative.