Inferential Statistics and Hypothesis Testing Assignment

-Aniket Verma

SOLUTIONS

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.

The type of probability distribution that would accurately portray the above scenario is ***binomial distribution***. In order to understand the reason why we use this distribution, let us revisit the most important conditions where we use Binomial Distribution and see if they are in sync with the above mentioned scenario.

Let us assume that we have a random variable X (which in this case refers to the number of drugs that are NOT able to do a satisfactory job.

The ***binomial distribution*** describes the behaviour of a random variable *X* if the following conditions apply:

1. The probability of success is same for each outcome.

   Note the following statement – *"it is 4 times more likely that a drug is able to produce a satisfactory result than not."*

   This basically means that ***the probability of a drug passing a test is 4 times more that it failing the same.*** Hence the probability of every outcome is ***fixed*** and the probability of success outcome is 4 times more than that of failure.

2. There are only two possible outcomes (Success or Failure) and every outcome is independent.

   In the above scenario the drug can either pass or fail the test. There is no third or any intermittent outcome. Also, the probability of success / failure of one drug has no influence what so ever on the probability of success / failure of any other drug as all tests are mutually independent. Therefore this is also a ***discreet probability distribution.***

3. The number of observations if fixed.
   This also means that the sample size is fixed and hence the number of outcomes will be equal to the sample size. This is clearly explained in the above scenario as we have a fixed sample size (=10 drugs).

   Hence, we would use ***binomial distribution*** here.

b.) Calculate the required probability.

The probability that the random value X takes the value k, is given by,

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

Here the random variable X refers to the number of drugs not able to do a satisfactory job.

Sample size (n) = 10
k = 0, 1, 2, 3 (at most 3 unsatisfactory drugs).

p = probability of a drug NOT doing a satisfactory job.

From our earlier discussion, we know that, the probability of a drug doing a satisfactory job is 4 times than that of a drug doing an unsatisfactory job.

Hence, by the rule of probabilities,

$$4 * p + p = 1$$
$$5 * p = 1$$
$$p = 1/5 \text{ (or } 0.20)$$

Hence, the probability of the drug doing an unsatisfactory job is 0.20.

Substituting this values in the previous equation,

$$P(X \leq 3)$$

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

$$= \binom{10}{0}(0.20)^0(1-0.20)^{10} + \binom{10}{1}(0.20)^1(1-0.20)^9 + \binom{10}{2}(0.20)^2(1-0.20)^8 + \binom{10}{3}(0.20)^3(1-0.20)^7$$

$$= \binom{10}{0}(0.20)^0(0.80)^{10} + \binom{10}{1}(0.20)^1(0.80)^9 + \binom{10}{2}(0.20)^2(0.80)^8 + \binom{10}{3}(0.20)^3(0.80)^7$$

Using $\binom{n}{r} = \frac{n!}{r!*(n-r)!}$

$$= \frac{10!}{0!*10!}(0.20)^0(0.80)^{10} + \frac{10!}{1!*9!}(0.20)^1(0.80)^9 + \frac{10!}{2!*8!}(0.20)^2(0.80)^8 + \frac{10!}{3!*7!}(0.20)^3(0.80)^7$$

$$= (0.20)^0(0.80)^{10} + 10 * (0.20)^1(0.80)^9 + 45 * (0.20)^2(0.80)^8 + 120 * (0.20)^3(0.80)^7$$

$= 0.10737 + 0.26843 + 0.30199 + 0.20133$

$= 0.87912$ (or 0.88)

Hence, the theoretical probability that at most, 3 drugs are not able to do a satisfactory job is 0.88 or 88%.

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.

The main methodology that we would use to approach this problem is by using the ***Central Limit Theorem.***

Why do we use Central Limit Theorem?

The Central Limit Theorem (CLT) is used when we try to estimate the population mean using the sample properties.

What is the Central Limit Theorem?

*The **Central Limit Theorem** states that the sampling distribution of the sample means approaches a normal distribution as the sample size gets larger — no matter what the shape of the population distribution.*

What are the properties of CLT?

The central limit theorem says that for any kind of data, provided a high number of samples has been taken, the following properties hold true:

- Sampling distribution's mean = Population mean
- Sampling distribution's standard deviation (**standard error**)
- For n > 30, the sampling distribution becomes a **normal distribution**.

b.) Find the required range.

$$ConfidenceInterval(Range) = \mu_x \pm \frac{Z * S}{\sqrt{n}}$$

Where z = z-score corresponding to confidence level.
S = sample standard deviation.

n = sample size.
μ_x = sample mean.

z-score corresponding to 95% confidence = 1.96
S = 65
n = 100
μ_x = 207

Substituting these values in the equation, we get,

$ConfidenceInterval(Range)$

$$= 207 \pm \frac{1.96 * 65}{\sqrt{100}}$$

$$= 207 \pm \frac{1.96 * 65}{\sqrt{100}}$$

$$= 207 \pm 12.74$$

$$= (194.26, 219.74)$$

Hence, the confidence interval for 95% is (194.26, 219.74). Which means that we can say with 95% confidence that the mean would lie in the above range.

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.

Null Hypothesis: The newer batch of painkillers drugs produce a satisfactory result and passes the quality assurance test. ($H \leq 200$)

Alternate Hypothesis: The newer batch of painkillers drugs DO NOT produce a satisfactory result and passes the quality assurance test. ($H > 200$)

Using the null and alternate hypothesis we can conclude that, it's an upper tailed test. Hence the critical region will lie on the right side.

- p-value method

In the p-value test, we initially compute the z-score for the sample mean.

$$Z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

Where $\bar{x}$ = sample mean
$\mu_{\bar{x}}$ = population mean
$\sigma_{\bar{x}}$ = sample standard deviation.

Substituting the values, we get,

Z = (207 – 200) / (65 / sqrt(100))
= 1.08

Calculation the Z-value from the z-table, we get 0.8599
Hence, the probability of z being greater than 1.08 is 1 – 0.8599 = 0.1401.

Now, we know that critical region α = 5% of the total region.
However the sample mean lies in the 14.01 % region.
Hence, we fail to reject the null hypothesis.

- Critical value method:

Probability of making an error (α) = 0.05
Since this is an upper-tailed test, area under the critical region is 0.05 (only one critical region)

Area under the acceptance region = 1 – Area under the critical region
= 1 – 0.05
= 0.95

We need to determine the Zc for the Upper Critical Value at 0.95
This is equal to the Area under the curve or normal cumulative probability until this point.

From the z-table we find the Zc which will be equal to 1.645.

Hence the Upper Critical Value (UCV) = $\mu + (Zc * \sigma_{\bar{x}})$

Where μ = population mean
$\sigma_{\bar{x}}$ = sample standard deviation
= 200 + (1.645 * (65 / sqrt(100)))
= 210.69

Since the mean doesn't fall in the critical region, 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

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).

	Case 1	**Case 2**
α	0.05	0.15
β	0.45	0.15

Let us first determine what α and β are with respect to the above scenario.
Type – I error (α): When you reject the null hypothesis despite it being true.
Type – II error (β): When you fail to reject the null hypothesis despite it being false.

Taking the scenario of the new batch of drugs. Let us define the type 1 and type 2 errors in this case,

Type – I error (α): When we discard the new batch of drugs even though they produce satisfactory results. *Consequences: The company incurs a loss as manufacturing cost.*

Type – II error (β): When we do not discard the new sample of drugs even though they aren't safe. *Consequences: The health of people is at risk.*

Clearly, from the above error definitions we can see that making a type – II error is way more dangerous than making a type – I error. Hence, the goal must be to minimize the type two error and Case-2 from the above table is a relatively better option.

But, this might not be the case every time. Let us look at two examples which may correspond to Case – 1 and Case – 2.

Example – 1: Let us say that some scientists at the weather department claim that in the next 24 hrs there is a possibility of a cyclone. The administrative panel has to test the claim. Based on the results of the test, a warning will be issued to the general public to stay indoors.

Null Hypothesis: Cyclone will occur in next 24 hrs.
Alternate Hypothesis: There is no possibility of cyclone in next 24 hrs.

Type – I error: Scientists were right that the cyclone will occur but the administrative team dismissed the report and didn't issue a warning. *Consequences: The life of people is at risk.*

Type – II error: Scientists were wrong, but still the administrative team decided to raise a warning. *Consequence: A small loss of day's productivity.*

Clearly making a type – I error is more dangerous here and hence this example can fall under the scenario where we should use Case – I from the table.

There are some scenarios where we need to keep the possibility of making either of the type - 1 or type – 2 error in check.

For example: In the game of guessing, if you roll the dice and get the number you win 100/- while you lose 10/- if you roll any of the other outcomes from 1 to 6, except the number. So, it is equally costly to make either of type – 1 or type – 2 error. Hence, the idea scenario would be to make minimal or no error. However, this is an idea situation and need not necessarily occur.

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.

Solution:

Let us try to understand what A/B testing is in order to understand which option is more effective.

*A/B testing (also known as **split testing** or **bucket testing**) is a method of comparing two versions of a webpage or app against each other to determine which one performs better.*

Hence instead of using our gut feeling, we leverage concrete data and evidence to make conclusion about which version is more beneficial.

In the online ad campaign, let us assume that the two taglines are A and B.

Here A is the ***control*** and B is the ***variation.***

Our goal is to determine that which of the two taglines is more effective in attracting customers.

A/B Testing can be done in the following steps:

- Data Collection: Two sample population groups are taken. For the first group the ad campaign is run using tagline A and for the second group the ad campaign is run using tagline B.
- Setting the Target: Two determine which tagline should be used for the ad campaign.
- Hypothesis Formulation:
    - Null Hypothesis: Tagline A is at least as good as B.

    - Alternate Hypothesis: Tagline B is better than A.
- Create Variations: Using an A/B Testing Software, we can create two online ad campaigns with different taglines.
- Run the experiment: We now run the online ad campaign for both population samples using tagline A and B respectively. Here we note the following parameters:
    - Population size for each sample (number of targeted customers).
    - Conversion rate of each customer (number of people who purchased the drug after seeing the campaign).
- Obtain and access the results: Analyse the result of A/B testing to determine whether we can reject the null hypothesis or not. This can be done by comparing the p-value against the critical value. If p-value > critical value, we fail to reject the null hypothesis else we reject it.
