Inferential and Hypothesis Testing Assignment

Name	**Rahul Kumar**
Email	**Rahul.cs068@gmail.com**
Phone	**+91-9538109454**
Course	**PGDDS**
Batch	**Dec 2018**

Contents

Problem Statement 3
Comprehension 3
Question 1: 3
Answer 1: 3
Question 2: 7
Answer 2: 7
Question 3: 9
Answer 3: 9
Question 4: 14
Answer 4: 14

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.

Answer 1:

To portray the above scenario, I would propose **Binomial distribution probability** which is more suitable looking to the problem statement.

The formula for finding binomial probability is given by –

P(X=r) = nCr * (p)^r * (1−p)^n−r

There are some conditions that need to be followed in order to be able to apply the formula.

- 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

Now in this case the three conditions of binomial distribution will be calculated as below:

1) Total no of trials
   - We are required to find the theoretical probability that at most, 3 drugs are not able to do a satisfactory job, this means **X <= 3**
   - So, the value of **r** can be 0,1,2,3

   Hence, Total no of trials **= 4**

2) Each trial is binary, i.e., has only two possible outcomes - success or failure
   - The trial here is binary in nature as the drug quality check results either in satisfactory or un-satisfactory, so only two possible outcomes.
   - The problem statement says "*the quality assurance check 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*"
   - This derives the meaning that if '**t**' number of drugs is resulting into unsatisfactory result, then **4*t** will result into satisfactory on quality check.

   So, this gives us idea to calculate the probability of satisfactory or un-satisfactory or say Success or failure of trials.

   Probability of un-satisfactory drug = t
   Probability of satisfactory drug = 4t

   Consider total number of drugs = (t+4t) = 5

   ⇨ Probability of un-satisfactory drug = t → 1/5 → **0.2**
   ⇨ Probability of satisfactory drug = 4t → 4/5 → **0.8**

3) Probability of success is same in all trials, denoted by p

- Yes, now we have same probability for success and failure in all the trials

   ⇨ Probability of success = **0.8**
   ⇨ Probability of failure = **0.2**

   Let's create probability distribution table

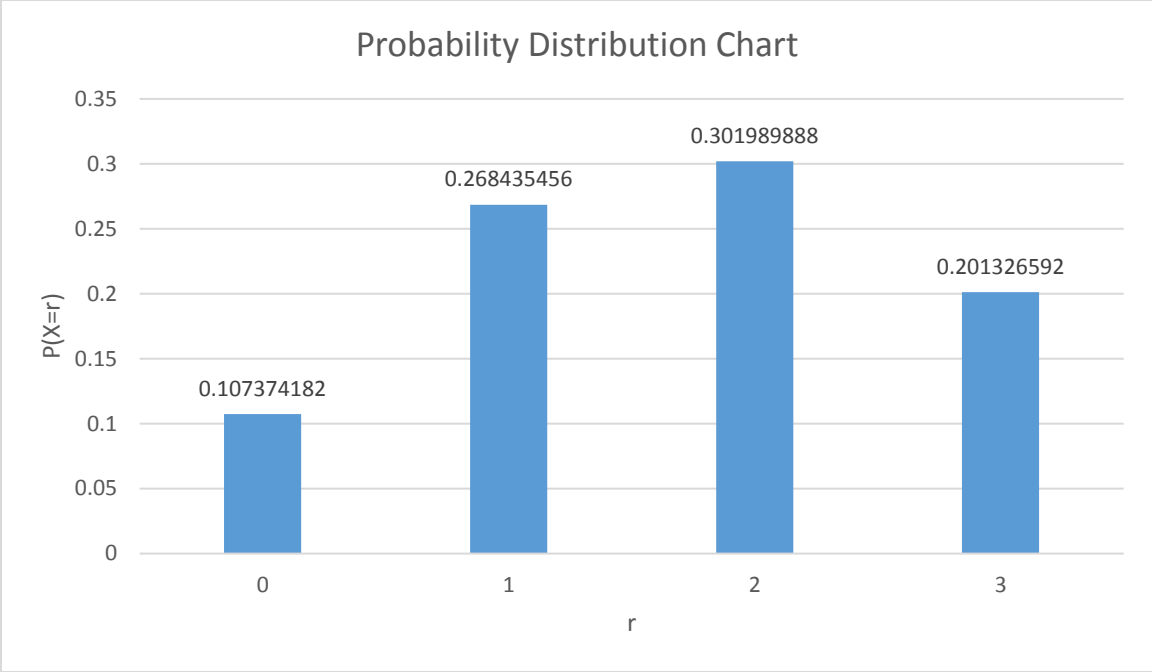
   Given, Sample size i.e. **n = 10**

n	r	n-r	n.C.r	(P)^n	(1-p)^n-r	nCr * (p)^r * (1−p)^(n −r)
10	0	10-0 = 10	10.C.0 = 1	0.8^10 = 0.107374182	0.2^0 = 1	0.107374182
10	1	10-1 = 9	10.C.1 = 10	0.8^9 = 0.134217728	0.2^1 = 0.2	0.268435456
10	2	10-2 = 8	10.C.2 = 45	0.8^8 = 0.16777216	0.2^2 = 0.04	0.301989888
10	3	10-3 = 7	10.C.3 = 120	0.8^7 = 0.2097152	0.2^3 = 0.008	0.201326592

Will keep the distribution table in simple form

X = r	P(X=r)
0	0.107374
1	0.268435
2	0.30199
3	0.201327

To visualize the probability in easy form, will draw a chart using these values:



Note: Above diagram is created using "chart in Microsoft word feature"

Now, finally to find the required probability of given problem statement is

P(X<=3) = P(X=0) + P(X=1) + P(X=1) + P(X=3)
P(X<=3) = 0.107374182 + 0.268435456 + 0.301989888 + 0.201326592
P(X<=3) = 0.879126118
P(X<=3) = **0.8791**

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.

Answer 2:

The methodology I would approach for this problem would be **Central limit theorem**.

The properties for CLT is:

1. Sampling distribution's mean ($\mu_{\bar{X}}$) = Population mean (μ)
2. Sampling distribution's standard deviation (Standard error) = $\frac{\sigma}{\sqrt{n}}$, where σ is the population's standard deviation and n is the sample size
3. For n > 30, the sampling distribution becomes a normal distribution

According to third property of CLT, it is clear in this example will become a normal distribution as sample size given is more than 30 i.e. 100.

Let's gather some information given in the problem statement:

Sample Size (n) = **100**
Sample mean ($\bar{X}$) = **207**
Standard Deviation (σ) = **65**
Population size (N) = 80000
Population mean (μ) = not known, can be calculated using confidence level
Confidence Level = 95 %

Calculating additional parameters which would help to find the confidence interval

Standard Error = σ/√n = 65/√100 = 65/10 = **6.5**
Confidence level = 95 % = **0.95**
Z* for confidence level of 95% = **1.96**

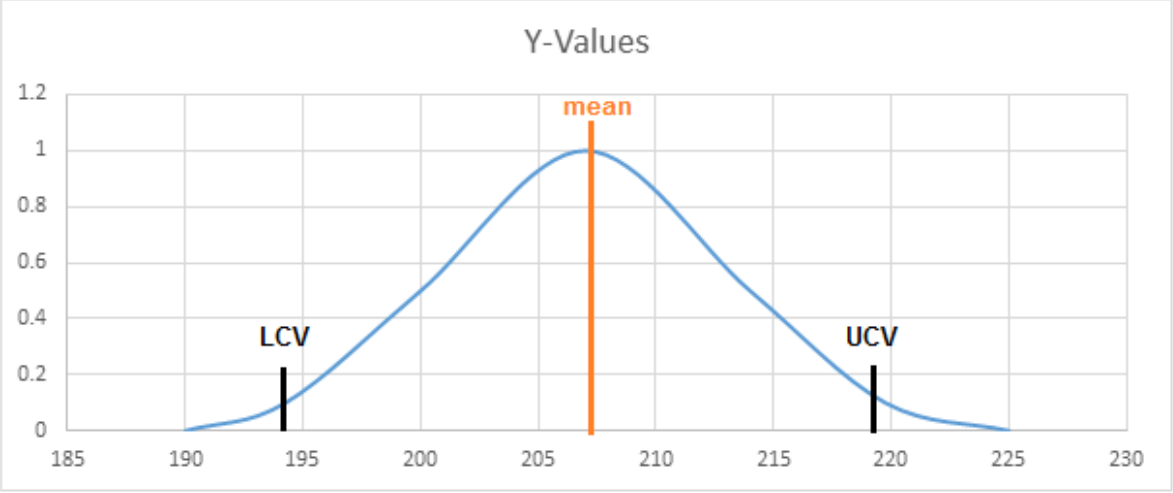
Now to find the solution of problem i.e. confidence interval, we should do

Confidence interval = $(\bar{X} - \frac{Z^* S}{\sqrt{n}}, \bar{X} + \frac{Z^* S}{\sqrt{n}})$,

Lower Confidence Limit = $\bar{X} - \frac{Z^* S}{\sqrt{n}}$;
= 207 – (1.96 * 65)/√100
= 207 – (1.96 * 65)/10
= 207 – 127.4)/10
= 207 – 12.74
= 194.26

Upper Confidence Limit = $\bar{X} + \frac{Z^* S}{\sqrt{n}}$
= 207 + (1.96 * 65)/√100
= 207 + (1.96 * 65)/10
= 207 + (127.4)/10
= 207 + 12.74
= 219.74

Hence the required range in which the population mean lie with a 95% confidence is (**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)

As first step, let's formulate the two hypotheses testing i.e. null and alternate hypotheses for the given situation

Null Hypothesis → The painkiller drug doing a satisfactory job with time of effect of at most 200 seconds
Alternate Hypothesis → The painkiller drug doing a satisfactory job with time of effect of more than 200 seconds

Null Hypothesis $\mathbf{H_0: \mu <= 200}$
Alternate Hypothesis $\mathbf{H_1: \mu > 200}$

The information given in problem statement
Sample Size (n) = **100**

Sample mean ($\bar{X}$) = **200**
Standard Deviation (σ) = **65**
Significance level = 5 % = **0.05**

Confidence Level = 100 – 5 = 95 % = 0.95

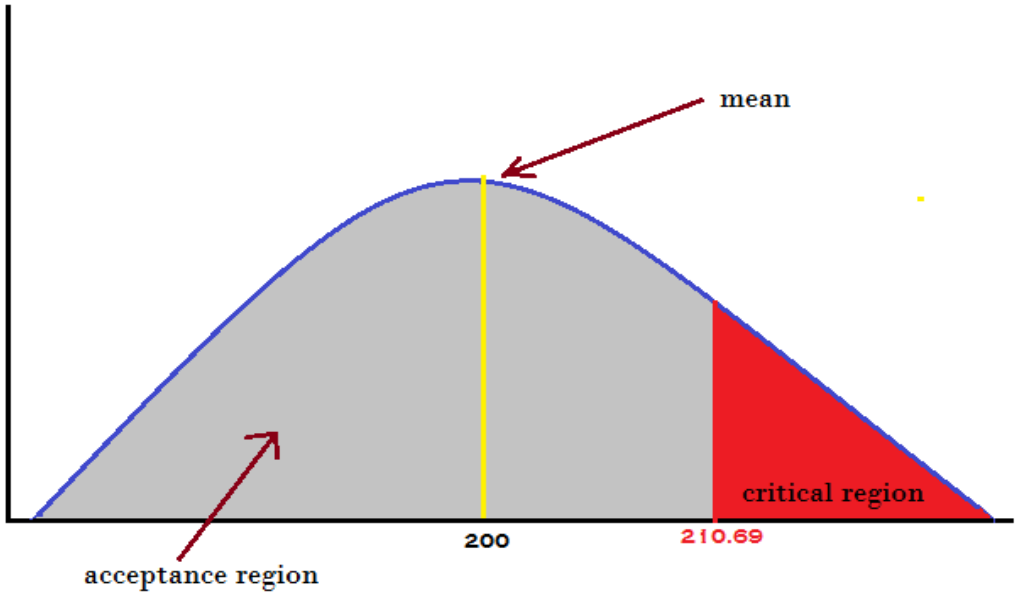
Here, the value of α is 0.05 (of 5%), so the area of the critical region would be 0.05

And since this is an upper-tailed test, the critical region is only on the right-hand side of the distribution,

So,
UCV (cumulative probability of that point) would be= 1-0.05 = **0.95**
So, this is the area cover under acceptance region.

Z value for 0.95 = **1.645**

Upper Critical Value = $\bar{X} + \frac{Z^* S}{\sqrt{n}}$
= 200 + (1.645 * 65)/√100
= 200 + (1.645 * 65)/10
= 200 + 10.6925
= **210.69**



If sample mean lies in the acceptance region, we fail to reject the null hypothesis because it is not beyond the critical point and we can consider that sample mean is equal to the population mean statistically.

The UCV values for this test is 210.69 the sample mean in this case is 200 seconds, this means sample mean is within acceptance region, hence, it implies that we **fail to reject the null hypothesis**.

Now, let check the decision according to **p value method**

Value of $\alpha = 0.05$

$Z = (207 - 200)/6.5 = 7 / 6.5 = 1.076923$

Z score for 1.076923 = 0.8577

So, P value = 1 - 0.8577 = **0.1423**

Since p value i.e. 0.1423 is greater than 0.05 value, hence decision drawn is **fail to reject the null hypotheses**

So, from both the method we could find result as "fail to reject the null hypothesis"

b)

A type I-error represented by α occurs when you reject a true null hypothesis.

A type-II error represented by β occurs when you fail to reject a false null hypothesis.

From the problem statement we could find

Case 1

Type 1 (α) = 0.05
Type 2 (β) = 0.45

Case2

Type 1 (α) = 0.15
Type 2 (β) = 0.15

Now let's check the total error for both the cases:
In first case, alpha value is very less (0.05) but total error is high (0.45 + 0.05) = 0.5
In Second case, alpha and beta is same, so total error is (0.15 + 0.15) = 0.3, which is less than case1.

Example 1
To prove case 1 is good, will take example from a separate industry

Construction Model of a bridge is correct.

Type I error: Predicting that the model is correct when it is not.

Type II error: Predicting that a model is not correct when it is correct.

In this case, Type I error is a huge risk. It could mean building the bridge which is faulty and can lead to a bridge collapse. Type II error is less serious as we can go through more models and still find a correct one.
Hence as we do have alpha as less and beta as more, so these scenario fits suitable here.

Example 2
Case 2 is in actual good for Sun pharma comparatively to Case 1.
The best example would be if there are two drugs and are critical for a patient. So In this case, looking at alpha and beta value, result will be good. And any increase in any value either alpha or beta cannot be risked.

Some additional details:
The (1-α) is the probability of correct decision and it correlates to the concept of 100 (1- α) % confidence interval used in estimation.
The (1- β) is the probability of correct decision and also known as 'Power of the test'. Since it indicates the ability or power of the test to recognize correctly that the null hypotheses in false, therefore, we wish a test that yields a large power.

Type I errors are generally considered more serious than Type II errors. The probability of a Type I error (α) is called the significance level and is set by the experimenter. There is a tradeoff between Type I and Type II errors. The more an experimenter protects himself or herself against Type I errors by choosing a low level, the greater the chance of a Type II error. Requiring very strong evidence to reject the null hypothesis makes it very unlikely that a true null hypothesis will be rejected. However, it increases the chance that a false null hypothesis will not be rejected, thus lowering power. The Type I error rate is almost always set at .05 or at .01, the latter being more conservative since it requires stronger evidence to reject the null hypothesis at the .01 level then at the .05 level.

We will take three scenarios to understand how the problem statement and situations are so important in this context. Let us revisit the example above first.

Scenario/Problem Statement 1: Providing access to an asset post a biometric scan.

Type I error: Possibility of rejection even with an authorized match.
Type II error: Possibility of acceptance even with an unauthorized match.

Here in this scenario, Type I error is not an issue. The person can do a scan again and it should work correctly. We would not want a Type II error happening. It would mean that an unauthorized user has access to the asset.

Scenario/Problem Statement 2: Construction Model of a bridge is correct.

Type I error: Predicting that the model is correct when it is not.
Type II error: Predicting that a model is not correct when it is correct.

In this case, Type I error is a huge risk. It could mean building the bridge which is faulty and can lead to a bridge collapse. Type II error is less serious as we can go through more models and still find a correct one.

Scenario/Problem Statement 3: Medical trials for a drug which is a cure for Cancer
Type I error: Predicting that a cure is found when it is not the case.
Type II error**:** Predicting that a cure is not found when in fact it is the case.
In this case, Type I error is not an issue. It could be corrected later with more trials. Type II error is more serious as it could be discarded as no cure and a cure can save millions of lives.

As you can see depending on the problem statement **Type I error can be better than Type II error or vice versa.** The chances of committing these two types of errors are inversely proportional, decreasing Type I error rate increases Type II error rate, and vice versa. Risk of committing a Type I error is represented by your alpha level (*p* value below which you reject the null hypothesis. Read more here). The commonly accepted $\alpha = .05$ means that we will incorrectly reject the null hypothesis approximately 5% of the time. To decrease your chance of committing a Type I error, we can make your alpha (p) value more strict. Alternately, we can also increase the sample size.

Similarly, if our dataset is medical in nature, we have to side more on the side of caution. On the other hand, something like spam filtering requires a different approach to these errors.

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 4:

A/B testing is a method of comparing two versions of an event against each other to determine which one performs or attracts the customer better.

It is also known as split testing or bucket testing.

Running an AB test that directly compares a variation against a current experience lets us ask focused questions about changes to the event, and then collect data about the impact of that change.

This testing takes the guesswork out of taglines for ad campaign optimization and enables data-informed decisions that shift business conversations from "we think" to "we know." By measuring the impact that changes have on metrics, we can ensure that every change produces positive results.

Let’s see, how A/B testing works

In an A/B test, we took two taglines for ad campaign. Then, half of our team is shown the one version of tagline for ad campaign and another half of team are shown with the second tagline of ad-campaign.

As both teams are served either the control or variation, their engagement with each experience is measured and collected in an analytics dashboard and analyzed through a statistical engine. We can then determine whether changing the experience had a positive, negative, or no effect on user behavior.

Why we should use A/B Test

A/B testing allows individuals, teams, and companies to make careful changes to their user experiences while collecting data on the results. This allows them to construct

hypotheses, and to learn better why certain elements of their experiences impact user behavior. In another way, they can be proven wrong—their opinion about the best experience for a given goal can be proven wrong through an A/B test.

More than just answering a one-off question or settling a disagreement, AB testing can be used consistently to continually improve a given experience, improving a single goal like conversion rate over time.

For this example, company's marketing team want to decide on best tagline for their online ad campaign. In order to achieve that goal, the team would try A/B testing to the tagline for ad campaign.

The team may test making some changes to the tagline as well. Testing one change at a time helps team to pinpoint which changes had an effect on their customer's behavior, and which ones did not. Over time, marketing team can combine the effect of multiple winning changes from experiments to demonstrate the measurable improvement. And can decide out of two tagline, which would be more effective and can be used.

By testing ad copy, marketers can learn which version attracts more clicks. By testing the subsequent landing page, they can learn which layout converts visitors to customer's best. The overall spend on a marketing campaign can actually be decreased if the elements of each step work as efficiently as possible to acquire new customers.

A/B Testing Process

The following is an A/B testing framework we can use to start running tests:

Collect Data: The analytics will often provide insight into where we can decide on our goal or may be optimizing.

Identify Goals: Goals can be anything from clicking a button or link to product purchases and e-mail signups. Here, in this example goal is to decide which tagline is better for online ad campaign which could attract more customers.

Generate Hypothesis: Once we've identified a goal we can begin generating A/B testing ideas and hypotheses for why we think that will be better than the other version. Once we have a list of ideas, prioritize them in terms of expected impact and difficulty of implementation.

Create Variations: Using A/B testing, we can make the desired changes or tests to an element of your experience. This might be changing anything.

Run Experiment: Kick off the experiment and wait for team to participate! At this point, user experiment will be randomly assigned to either the control or variation of experience. Their interaction with each experience is measured, counted, and compared to determine how each performs.

Analyze Results: Once experiment is complete, it's time to analyze the results. A/B testing software will present the data from the experiment and will show the difference between how the two versions of taglines performed, and whether there is a statistically significant difference.

----- End of File -----