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

a) The binomial distribution is a probability distribution that summarizes the likelihood that a value will take one of two independent values under a given set of parameters or assumptions. The underlying assumptions of the binomial distribution are that there is only one outcome for each trial, that each trial has the same probability of success, and that each trial is mutually exclusive, or independent of each other.

The binomial distribution is a common discrete distribution used in statistics, as opposed to a continuous distribution such as the normal distribution. This is because the binomial distribution only counts two states, typically represented as 1 (for a success) or 0 (for a failure) given a number of trials in the data. The binomial distribution therefore represents the probability for x successes in n trials, given a success probability p for each trial.

The binomial distribution is a discrete probability distribution used when there are only two possible outcomes for a random variable: success and failure. The binomial distribution assumes a finite number of trials, n. Each trial is independent of the last.

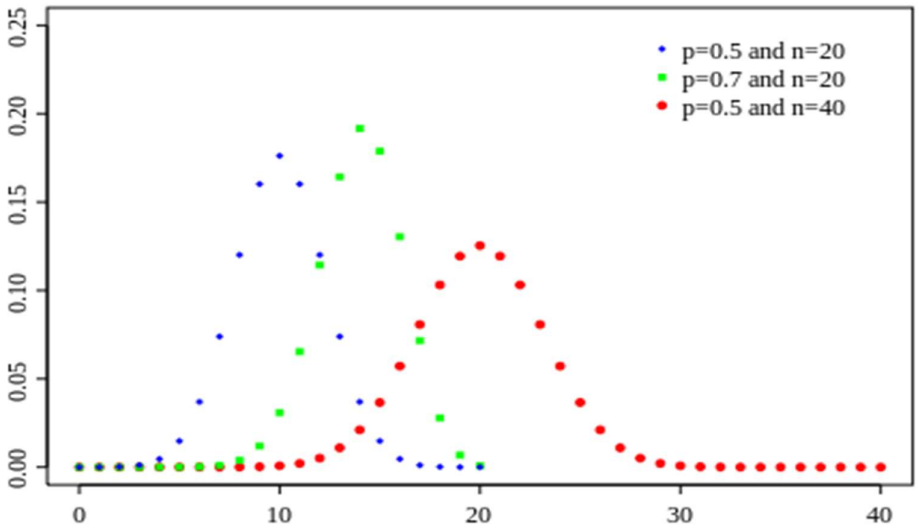


Fig-1

fig-1 shows binomial distribution with different trial & probability combination.

For large values of n, the distributions of the count X and the sample proportion are approximately normal. The mean and variance for the approximately normal distribution of X are np and np(1-p), identical to the mean and variance of the binomial(n,p) distribution.

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

1: The number of observations n is fixed (here n=10)
2: Each observation is independent.
3: Each observation represents one of two outcomes (here "success" or "failure").
4: The probability of "success" p is the same for each outcome.

These conditions are met, in above scenario that is that at most 3 drugs are not able to do a satisfactory job out of 10. Drugs. That means it is not passed the test then *X* has a binomial distribution with parameters *n* and *p*, abbreviated *B(n,p)*.

It would be Binomial distribution. A **binomial distribution** can be thought of as simply the probability of a SUCCESS or FAILURE outcome in an experiment or survey that is repeated multiple times. The binomial is a type of distribution that has **two possible outcomes** (the prefix "bi" means two, or twice). Taking a above scenario test could have two possible outcomes: that is pass or fail.

BINOMIAL PROBABILITY DISTRIBUTION

x	P(X=x)
0	$^nC_0(p)^0(1-p)^n$
1	$^nC_1(p)^1(1-p)^{n-1}$
2	$^nC_2(p)^2(1-p)^{n-2}$
3	$^nC_3(p)^3(1-p)^{n-3}$
.	.
.	.
.	.
.	.
n	$^nC_n(p)^n(1-p)^0$

So, the formula for finding binomial probability is given by –

$P(X=r)= nCr(p)r(1-p)n-r$

Where n is no. of trials, p is probability of success and r is no. of successes after n trials.

Binomial distribution is clearly depicts in above scenario.

Three conditions that this distribution follows.

1.Total number of trials is fixed at n.

2.Each trial is binary, i.e., has only two possible outcomes - success or failure.

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

b) **A cumulative binomial probability** refers to the probability that the binomial random variable falls within a specified range (e.g., is greater than or equal to a stated lower limit and less than or equal to a stated upper limit).

These conditions are met, in above scenario that is that at most 3(P(X<=3)) drugs are not able to do a satisfactory job out of 10 drugs.

Cumulative probability of x, generally denoted by **F(x)**, is the probability of the random variable X, taking a value lesser than x. Mathematically speaking, we'd say –
F(x) = P(X < =x)

It states that the probability of drugs producing satisfactory result =4/5 and probability of drugs not producing satisfactory result =1/5 during the quality assurance of previous batch of product

If X is defined as the number of sample drugs found to be not able to do a satisfactory job. after testing 10 samples, then X would follow a binomial distribution with n = 10 and p = 0.2. So, the probability asked in the question becomes P(X<=3) = P(X=0) + P(X=1) + P(X=2)+ P(X=3)= 10C0(.02)0(0.8)10+ 10C1(0.2)1(0.8)9+ 10C2(0.2)2(0.8)8 +10C3(0.2)3(0.8)7

= 0.1073+0.2684+0.3019+0.2013=87.89=87.89%

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.

a.) Above problem main methodology is **Central Limit Theorem** used. The sampling distribution, which is basically the distribution of sample means of a population, has some interesting properties which are collectively called the central limit theorem, which states that no matter how the original population is distributed.

The central limit theorem states that if we have a population with mean μ and standard deviation σ and **take** sufficiently large random samples from the population with replacement , then the distribution of the sample means will be approximately normally distributed.

Central Limit Theorem and Statistical Inferences- **Central Limit Theorem** (CLT) is an **i**mportant result in statistics, most specifically, probability theory. More specifically, as

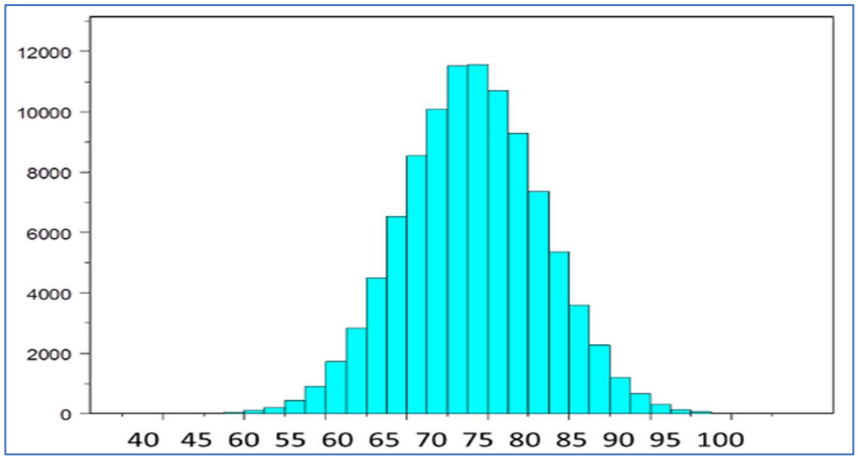
sample sizes become larger, the mean distribution that is measured from repeated sampling will reach the normal limits.

The sampling distribution will follow these properties –

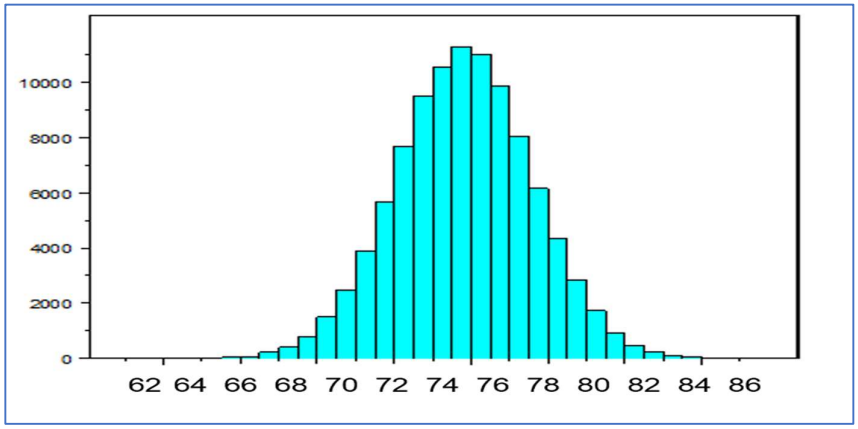
1. Sampling Distribution's Mean ($\mu\bar{X}$) = Population Mean (μ).
2. Sampling Distribution's Standard Deviation (Standard Error) = $\sigma\sqrt{n}$, where σ is the population's standard deviation and n is the sample size.
3. We can use the CLT if n is large enough.

It works normal population and Skewed Distribution.

The figure below illustrates a normally distributed characteristic, X, in a population in which the population mean is 75 with a standard deviation of 8.



If we take simple random samples of size n=10 from the population and compute the mean for each of the samples, the distribution of sample means should be approximately normal according to the Central Limit Theorem. Note that the sample size (n=10) is less than 30, but the source population is normally distributed, so this is not a problem. The distribution of the sample means is illustrated below. Note that the horizontal axis is different from the previous illustration, and that the range is narrower.



The mean of the sample means is 75 and the standard deviation of the sample means is 2.5, with the standard deviation of the sample means computed as follows:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{10}} = 2.5$$

If we were to take samples of n=5 instead of n=10, we would get a similar distribution, but the variation among the sample means would be larger. In fact, when we did this we got a sample mean = 75 and a sample standard deviation = 3.6.

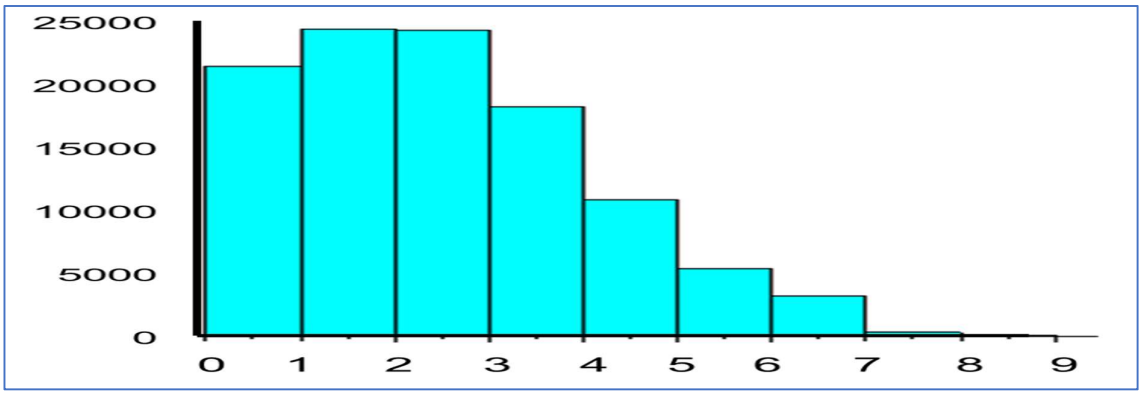


Fig shows Skewed Distribution

To summarise, the notation and formulae related to samples, populations and sampling distributions are –

Population/Sample	Term	Notation	Formula
Population ($X_1, X_2, X_3, \ldots\ldots, X_N$)	Population Size	N	Number of items/elements in the population
	Population Mean	μ	$\frac{\sum_{i=1}^{i=N} X_i}{N}$
	Population Variance	σ^2	$\frac{\sum_{i=1}^{i=N}(X_i - \mu)^2}{N}$
Sample ($X_1, X_2, X_3, \ldots\ldots, X_n$) (Sample of Population)	Sample Size	n	Number of items/elements in the sample
	Sample Mean	$\bar{X}$	$\frac{\sum_{i=1}^{i=n} X_i}{n}$
	Sample Variance	S^2	$\frac{\sum_{i=1}^{i=n}(X_i - \bar{X})^2}{n-1}$
Sampling Distribution of the Sample Mean ($\bar{X}_1, \bar{X}_2, \bar{X}_3, \ldots\ldots, \bar{X}_k$) (k Sample Means)	Sampling Distribution's Size	No convention (We have used k, but that is not a norm)	
	Sampling Distribution's Mean (mean of sample means)	$\mu_{\bar{X}}$	$\mu_{\bar{X}} = \mu$
	Sampling Distribution's Standard Deviation	S.E. (Standard Error)	S.E. = $\sigma/\sqrt{n}$

So, to summarise, let's say we have a sample with sample size n, mean \bar{X} and standard deviation S. Now, the y% confidence interval (i.e. the confidence interval corresponding to y% confidence level) for μ would be given by the range:

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

where, Z* is the Z-score associated with a y% confidence level. In other words, the population mean and sample mean differ by a margin of error given by $\frac{Z^*S}{\sqrt{n}}$.

Z* values are given below:

Confidence Level	Z*
90%	±1.65
95%	±1.96
99%	±2.58

The three steps to calculate require range

1.First, take a sample of size n

2, Then, find the mean $\bar{X}$ and standard deviation S of this sample

3. the confidence interval for the population mean μ, , is given by $(\bar{X} - \frac{Z^*S}{\sqrt{n}}, \bar{X} + \frac{Z^*S}{\sqrt{n}})$

We cannot finish step 3 without CLT. CLT lets us assume that the sample mean, would be normally distributed, with mean u and standard deviation $\frac{\sigma}{\sqrt{n}}$ (approx. $\frac{S}{\sqrt{n}}$).

b) Find the required range.

The formula for estimation is:

$\mu = \overline{X} \pm 1.96\,\sigma / \sqrt{n}$ (confident level 95%)

Calculation

$\mu = \overline{X} \pm 1.96\,\sigma / \sqrt{n}$

Where $\overline{X}$: =sample mean ,n=no. of the sample , Z = Z statistic determined by confidence level,

σ =standard deviation

$\overline{X}$: σ =207, =65 n=100

$\mu = 207 \pm 1.96*65/\sqrt{100}$)

$\mu = 207 \pm 1.96*6.5$

$\mu = 207 \pm 12.74$

Result

$\mu = 207$, 95% CI [194.26, 219.74].

Uses 95% confident that the population mean (μ) falls between 194.26 and 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.

Answer

We are using two hypothesis methods to describe above problem.

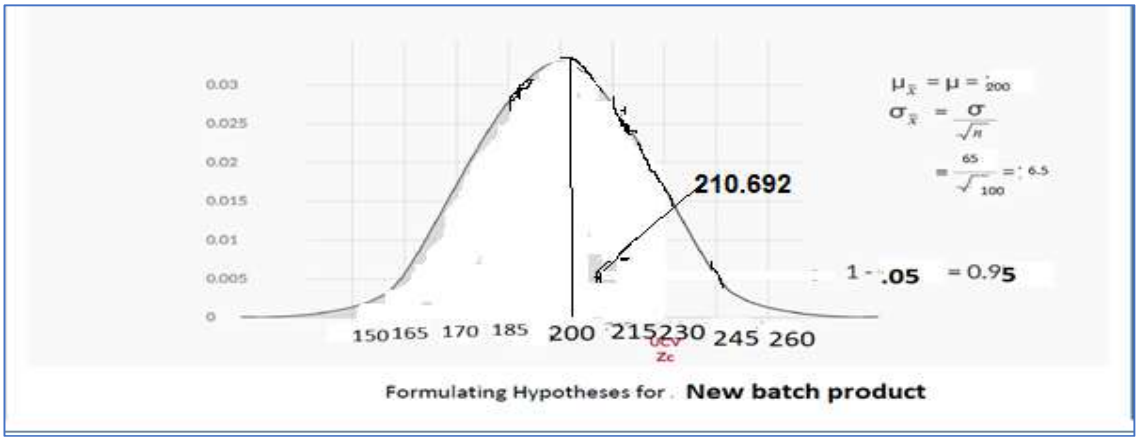
1; **Critical Value Method**
2. **p-value Method**

Critical Value Method:

The **critical value approach** involves determining "likely" or "unlikely" by determining whether or not the observed test **statistic** is more extreme than would be expected if the null hypothesis were true. That is, it entails comparing the observed test **statistic** to some cut-off **value**, called the "**critical value**."

The **critical region** is the **region** of values that corresponds to the rejection of the null hypothesis at some chosen probability level.

Find below snap shot of the distribution.



From the above problem we get

μ = 200 σ = 65 N (Sample size) = 100

Zc=1-.05=0.95

0.95, Z value is 1.645

Hypothesis:

Formulating the hypothesises this **upper-tailed test**, which would be:

H_0: μ ≤ 200 (null hypothesis) H_1: μ > 200(Alternative hypothesis)

'> 'in H_1 → Upper-tailed test → Rejection region on **right side** of distribution

The steps to follow to make a decision using the critical value method are as follows:

1.Calculate the value of Zċ from the given value of α (significance level).

2.Calculate the critical values from the value of Zċ.

3.Make the decision on the basis of the value of the sample mean x with respect to the critical values.

It is **an Upper-tailed test**, In this problem, the area of the critical region beyond the only critical point, which is on the right side, is 0.05. So, the cumulative probability of the critical point (the total area till that point) would be 0.95).

The z-score for 0.95 is 1.645.

So, the Zc comes out to be 1.645. Now, find the critical value for the given Zc and make the decision to accept or reject the null hypothesis.

Calculation:

μ = 200 σ = 65 N (Sample size) = 100

–

x= 207

The critical value can be calculated from μ + Zc x (σ/√N)

200 + 1.645(65/√100) = 210.692

Decision: Since 207 ($\bar{x}$) is less than 210.692, $\bar{x}$ lies in the acceptance region and fail to reject the null hypothesis.

Critical value = 210.692 and Decision = Fail to reject the null hypothesis

So new batch produces a satisfactory result and passes the quality assurance test.

p-value Method:

A P-value measures the strength of evidence in support of a null hypothesis. Suppose the test statistic in a hypothesis test is equal to *K*. The P-value is the probability of observing a test statistic as extreme as *K*, assuming the null hypothesis is true. If the P-value is less than the significance level, we reject the null hypothesis.

The **p-value** is the level of marginal significance within a **statistical** hypothesis test representing the probability of the occurrence of a given event. The **p-value** is used as an alternative to rejection points to provide the smallest level of significance at which the null hypothesis would be rejected.

Statistical hypothesis testing is used to determine whether the result of a data set is **statistically significant**. This test provides a p-value, representing the probability that random chance could explain the result; in general, a p-value of 5% or lower is considered to be **statistically significant**.

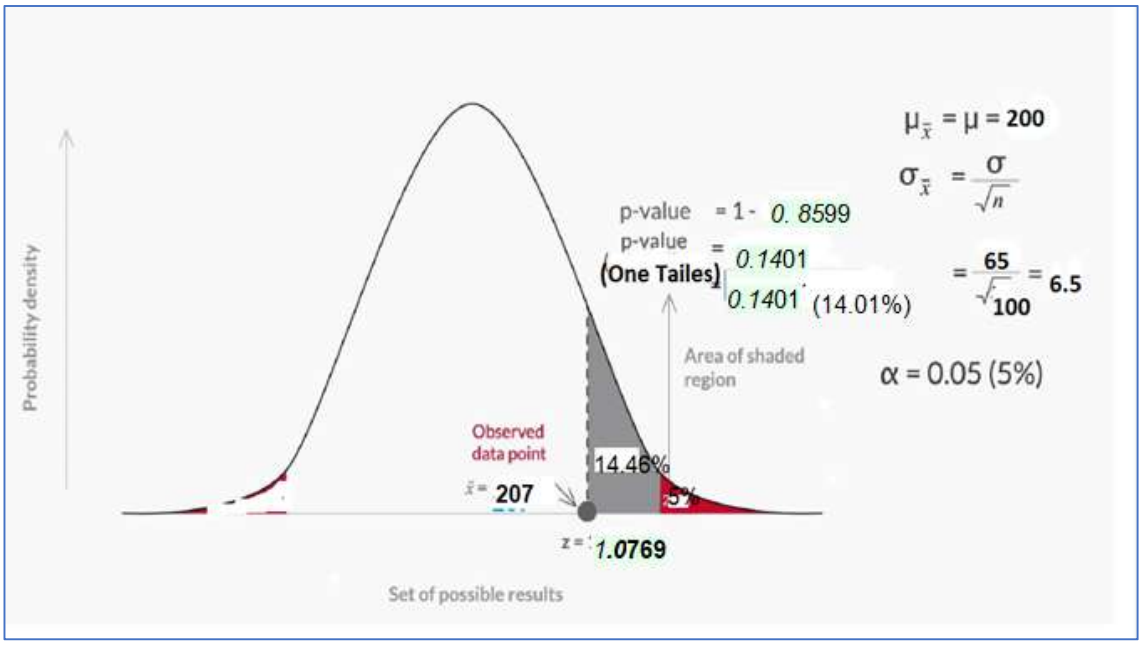


Figure 1 – Interpretation of p-value

Hypothesis: Formulating the hypothesises for this **upper-tailed test**, which would be:

H_0: μ ≤ 200 (Null hypothesis) and H_1: μ > 200(Alternative hypothesis)

'> 'in H_1 → Upper-tailed test → Rejection region on **right side** of distribution

Steps to evaluate p-value method

1.Calculate the value of z-score for the sample mean point on the distribution

2.Calculate the p-value from the cumulative probability for the given z-score using the z-table

3.Make a decision on the basis of the p-value (multiply it by 2 for a two-tailed test) with respect to the given value of α (significance value).

Calculation:

Calculate the z-score for sample mean ($\bar{x}$) = 207

From the above problem we get

μ = 200 σ = 65 N (Sample size) = 100

formula: ($\bar{x}$ - μ) / (σ /√N). This gives (207 - 200) / (65 /√100) = (7) / (65 / 10) = 1.0769, Since the sample mean lies on the right side of the hypothesised mean of 200, the z-score comes out to be positive.

The value in the z-table corresponding to 1.08 on the vertical axis and 0.08 on the horizontal axis is 0. 8599. Since the sample mean is on the right side of the distribution and this is an one-tailed test the p-value would be (1 - 0. 8599) = 0.1401

Decision:

Here, the p-value comes out to be 0.1401 Since the p-value is more than the significance level (0.1401>0.05) and higher p-value gives higher is the probability of failing to reject a null hypothesis.

A large p-value (> 0.05) indicates weak evidence against the null hypothesis, so you fail to reject the null hypothesis.

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

p-value = 0.1401 **and Decision = Fail to reject the null hypothesis**

So new batch produces a satisfactory result and passes the quality assurance test.

Question

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.

Answer

Types of errors:

Type-I error - Occurs when we reject a null hypothesis even when it is true

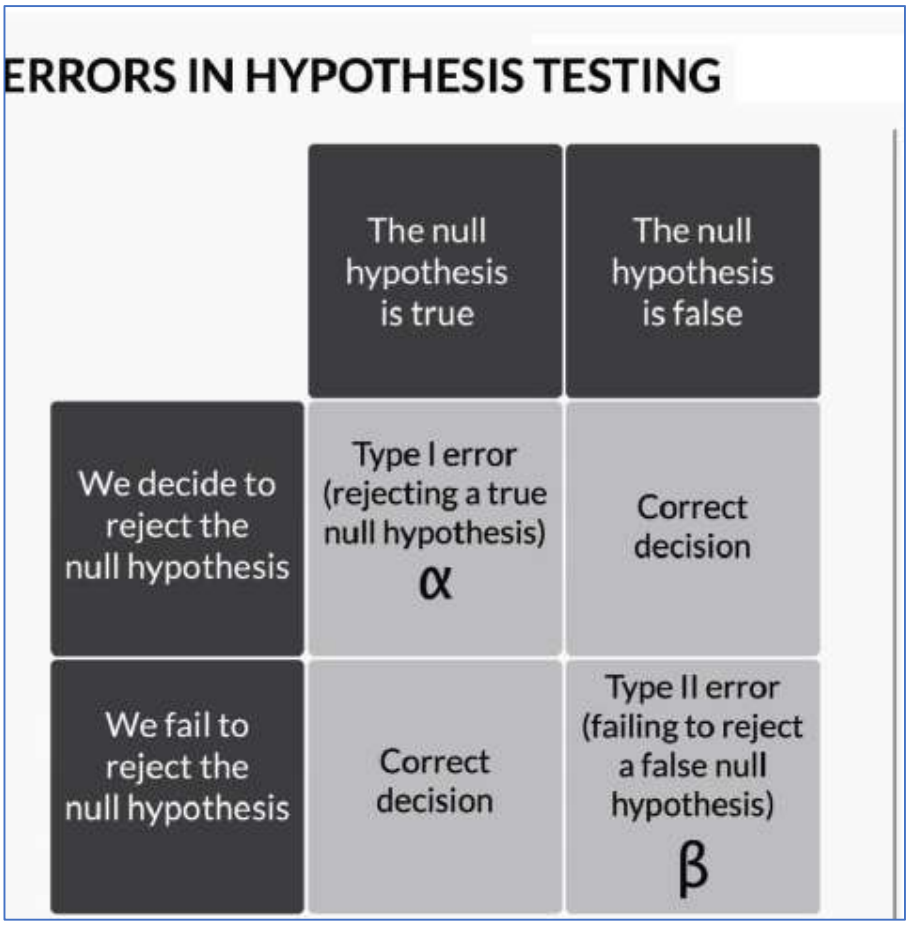
- Its probability is represented by α

Type-II error - Occurs when we fail to reject the null hypotheses even though it is false

- Its probability is represented by β

Relation between α & β is it inversely proportional.

Hypothesis: H_0: μ ≤ 200 (null hypothesis) H_1: μ > 200(Alternative hypothesis)



Types of errors in Hypothesis Testing

Case-1 ***type-1 error & probability***(α) ***=5% & type-2 error & probability***(β) ***=45%***

(1) The desired *α* level, that is, we willingness to commit a Type I error.(5%)
(2) The desired power or, equivalently, the desired *β* level, that is, we willingness to commit a Type II error.(45%).

Therefore, it is in testing, at the $\alpha = 0.05$ level, the null hypothesis H_0: $\mu <= 200$ against the alternative hypothesis that H_A: $\mu > 200$.

As it is always the case, we need to start by finding a threshold value *cv*, such that if the sample mean is larger than *cv*, we'll reject the null hypothesis:

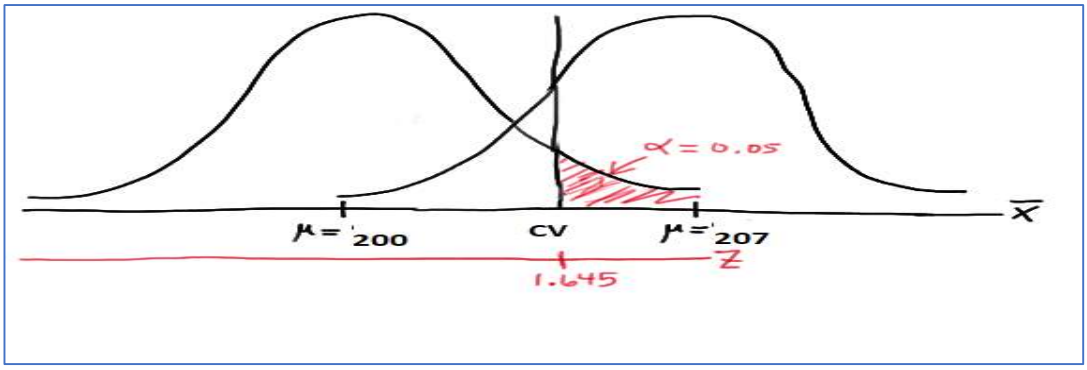


Fig-1

Fig=1 shows $\alpha = 0.05$ where $\mu = 200$ z value=1.645

That is, in order for our hypothesis test to be conducted at the $\alpha = 0.05$ level, the following statement must hold (using our typical *Z* transformation):

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

But, that's not the only condition that *cv* must meet, because *cv* also needs to be defined to ensure that our power is 0.45 or, alternatively, that the probability of a Type II error is 0.45. That would happen if there was a 45% chance that our test statistic fell short of *cv* when $\mu = 207$, as the following drawing illustrates in blue:

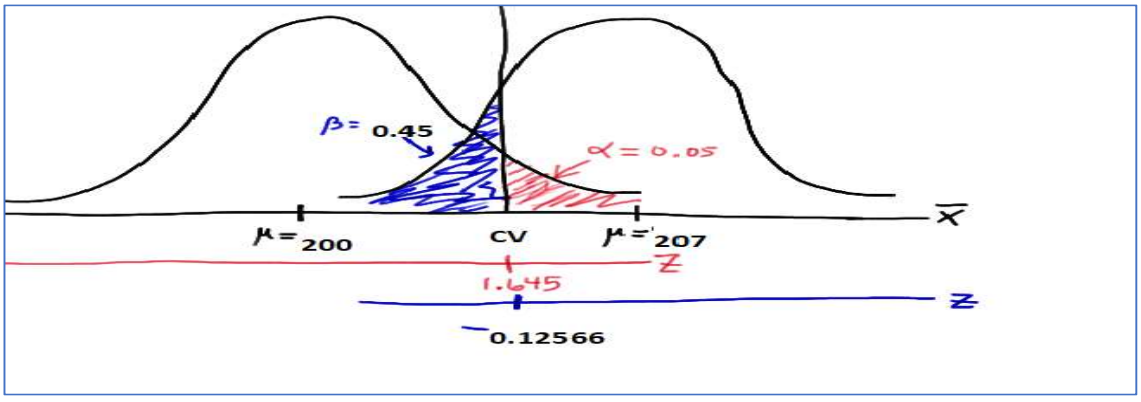


Fig-2

This illustration suggests that in order for our hypothesis test, the following statement must hold (using our usual *Z* transformation):

$$cv=207-0.126(65\sqrt{100})$$

Now that we know we will set $n = 100$, we can solve for our threshold value *cv*:

$$cv=200+1.645(\frac{65}{\sqrt{100}})$$
$$cv=200+1.645(6.5)= 210.692$$

So, in summary, if the medical researcher collects data on n = 100 samples and fail to reject the null hypothesis H_0: μ <=200 if the 100 sample time of effect is greater than 210.692 seconds, he will have a 5% chance of committing a Type I error and a 45% chance of committing a Type II error if the population mean μ were actually 207. because we don't have enough evidence to deny H0.

A Type I error is when we reject a true null hypothesis. Lower values of alpha(α)make it harder to reject the null hypothesis, so choosing lower values for alpha(α) can reduce the probability of a Type I error. The consequence here is that if the null hypothesis is false, it may be more difficult to reject using a low value for alpha(α). So using lower values of alphaα can increase the probability of a Type II error.

So above scenario it is seen lower significant level (5%) it is harder to reject null hypothesis.

So choosing lower values for alpha(α) can reduce the probability of a Type I error.

So we will conduct hypothesis testing with different set of scenario.

We now conducting testing below scenario where α & β controls by 15%..

Case-2 : *type-1 error & probability(α) =15% & type-2 error & probability(β) =15%*

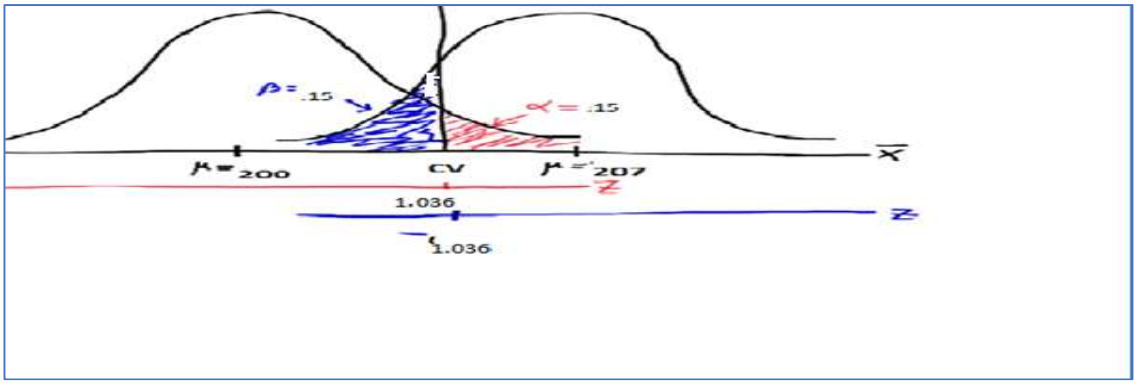


Fig-3

(1) The desired α level, that is, we willingness to commit a Type I error.(15%)
(2) The desired power or, equivalently, the desired β level, that is, we willingness to commit a Type II error.(15%).

That is, in order for our hypothesis test to be conducted at the α = 0.15 level, the following statement must hold (using our typical Z transformation):

$$cv=200+1.036\ (65\sqrt{100}$$

But, that's not the only condition that c*v* must meet, because *cv* also needs to be defined to ensure that our power is 0.15 or, alternatively, that the probability of a Type II error is 0.15. That would happen if there was a 15% chance that our test statistic fell short of *c*v when $\mu = 207$, as the following drawing illustrates in blue:

This illustration suggests that in order for our hypothesis test, the following statement must hold (using our usual *Z* transformation):

$$cv=207-1.036(65\sqrt{100})$$

Now that we know we will set $n = 100$, we can solve for our threshold value *c*:

$$cv=200+1.036(65\sqrt{100})cv=(200+6.734)=206.734$$

So, in summary, if the medical researcher collects data on $n = 100$ samples and reject the null hypothesis H_0: $\mu <=200$ if the 100 sample time of effect is greater than 206.734 seconds, he will have a 15% chance of committing a Type I error and a 15% chance of committing a Type II error if the population mean μ were actually 207.

As 207>206.734 out of critical region.(rejecting a true null hypothesis)

committing type I error has been considered more risky, and thus more strict control of type I error has been performed in statistical inference

A Type II error is when we fail to reject a false null hypothesis. Higher values

of alpha(*α)* make it easier to reject the null hypothesis, so choosing higher values

for alpha(*α*)can reduce the probability of a Type II error. The consequence here is that if the

null hypothesis is true, increasing alpha(*α)* makes it more likely that we commit a Type I

error (rejecting a true null hypothesis).

Scenario 1 works good when significant level less (5%) and less evident & less no of

samples.

Scenario 2 works good when significant level more (15%) and more evident & more no of

samples.

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.

Answer

Why A/ Testing A/B testing is a direct industry application of the two-sample proportion test sample .Team can't take decision which tagline is better A/B testing is introduced.

Two taglines options are as follow

Option1: Tagline A in landing page (old feature)-Control Group

Option2 : Tagline B in Landing Page(new feature)-Variation Group

Thus A/B testing is a method of comparing two versions of a sample against each other to determine which one performs better. AB testing is essentially an experiment where two or more variants of a sample are shown to users at random, and statistical analysis is used to determine which variation performs better for a given conversion goal.

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 behaviour. 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 instance, the pharmaceutical company Sun Pharma may want to improve their sales lead quality and volume from campaign landing pages. In order to achieve that goal, the team would try A/B testing changes to the headline, visual imagery, form fields, call to action, and overall layout of the page

Testing one change at a time helps them pinpoint which changes had an effect on their visitors' behaviour, and which ones did not. Over time, they can combine the effect of multiple winning changes from experiments to demonstrate the measurable improvement of the new experience over the old one.

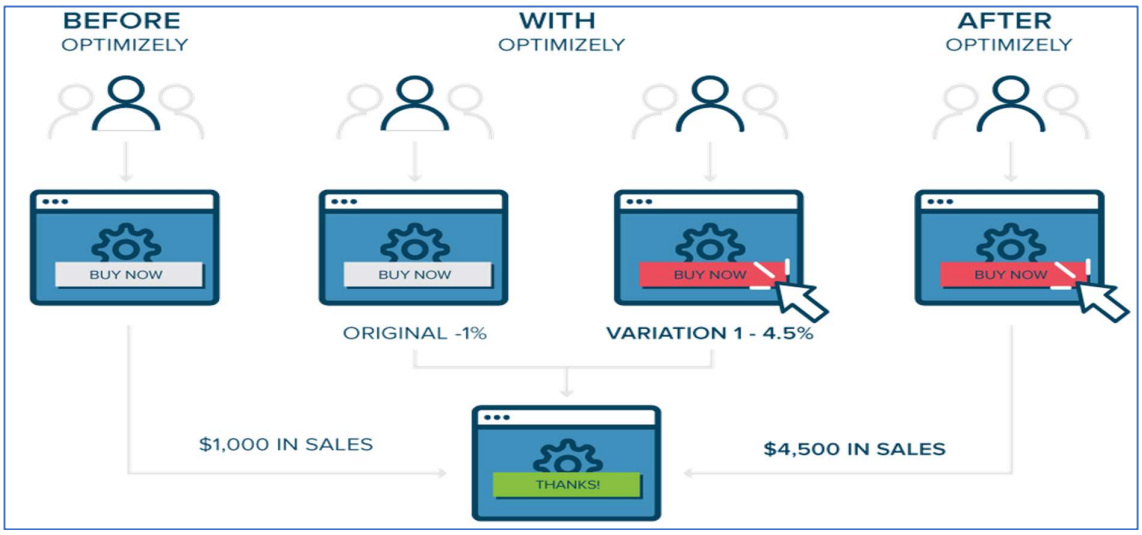
This method of introducing changes to a user experience also allows the experience to be optimized for a desired outcome and can make crucial steps in a marketing campaign more effective.

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 customers 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 demonstrates the impact of new time of effect features or changes to a user experience. Product onboarding, user engagement, modals, and in-product experiences can all be optimized with A/B testing,

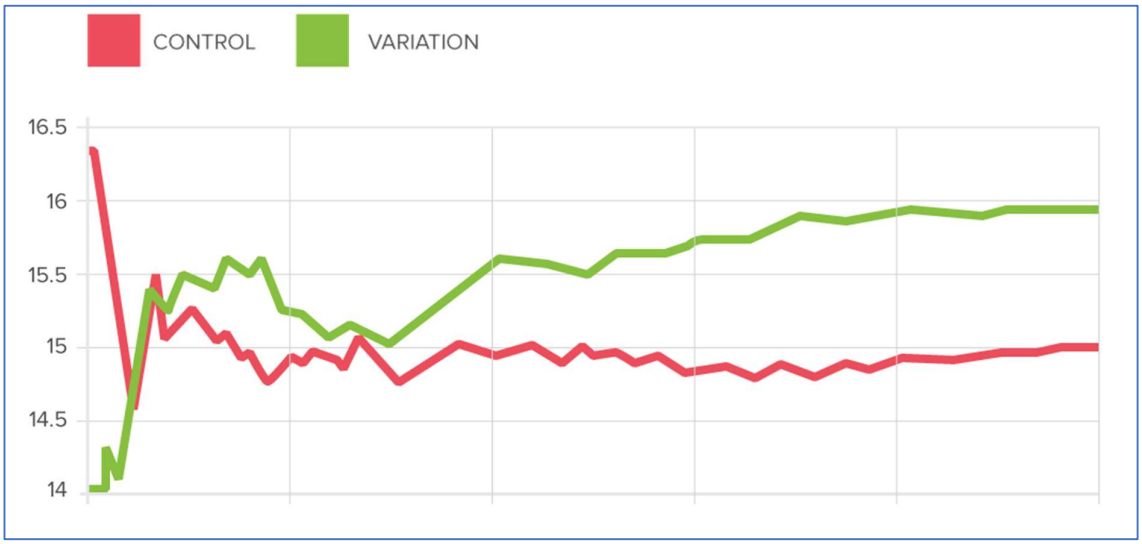
How A/B Testing Works

In an A/B test, we take a webpage or app screen and modify it to create a second version of the same page. This change can be as simple as a single headline or button, or be a complete redesign of the page. Then, half of the traffic is shown the original version of the page (known as the control) and half are shown the modified version of the page (the variation).

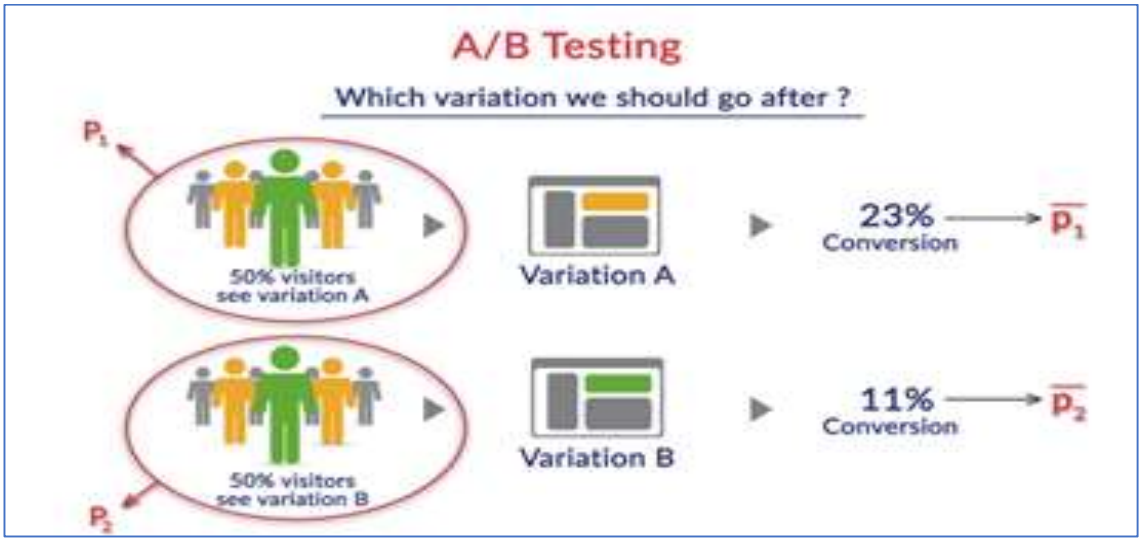


As visitors are served either the control or variation, their engagement with each experience is measured and collected in an analytics dashboard and analysed through a statistical engine. You

can then determine whether changing the experience had a positive, negative, or no effect on visitor behaviour.



AB testing is 2 population proportion testing.



In AB testing is we divided all the visitors in to two categories. One is experienced control group A while other in experiences variance B. We get two population P1 & P2. Now Both P1 & P2 some proportion of customers would positively converted for the company. Now we get proportion of $\overline{P_1}$ & $\overline{P_2}$ .Objective is to difference between $\overline{P_1}$ & $\overline{P_2}$ arrive a conclusion there is a significant different between $\overline{P_1}$ & $\overline{P_2}$ & control A & variation B.

We use two versions of apply now button with option-1 & option2.Option1 contains old features and known as Control group. Option-2 contains new features and known as Variable group.

Collect Data:

Now we collect data from web-site for proportion testing. Frequency is captured in both cases in which people fill data. Following data captured for option1 & option2

Option-1	Option-2
1.Sample size-2620 2.Frequency-618	1.Sample size-2380 2.Frequency-601

So fill up data for option-1 is 23.5% & Option-2 is 25%.Now we 've to test hypothesis whether new features are good or bad.

Option1: Tagline A in landing page (old feature)-Control Group

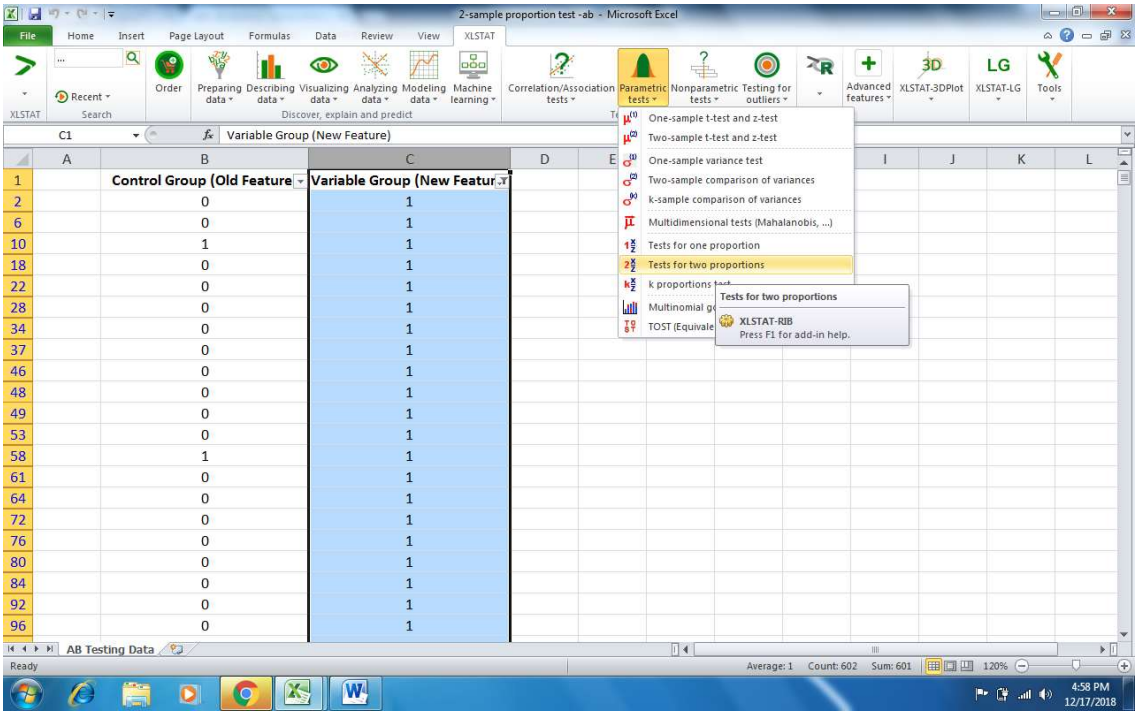
Option2 : Tagline B in Landing Page(new feature)-Variation Group

Generate Hypothesis Convention of null hypothesis is old features are better than new features. That means conversion of control group is better than conversion of experiment group. We 've to invalidate this hypothesis & prove new features which we introduce in new version are working good.

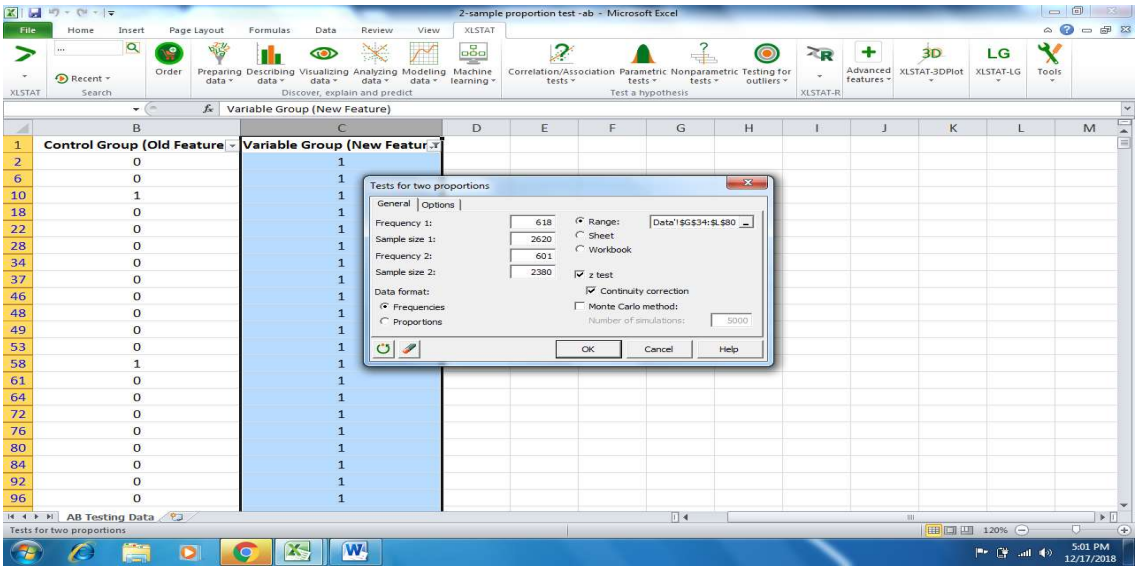
Now we present above scenario by use EXSTAT to perform A/B testing. Although other tools are(Optimizely, R) available in the market.

Steps for testing

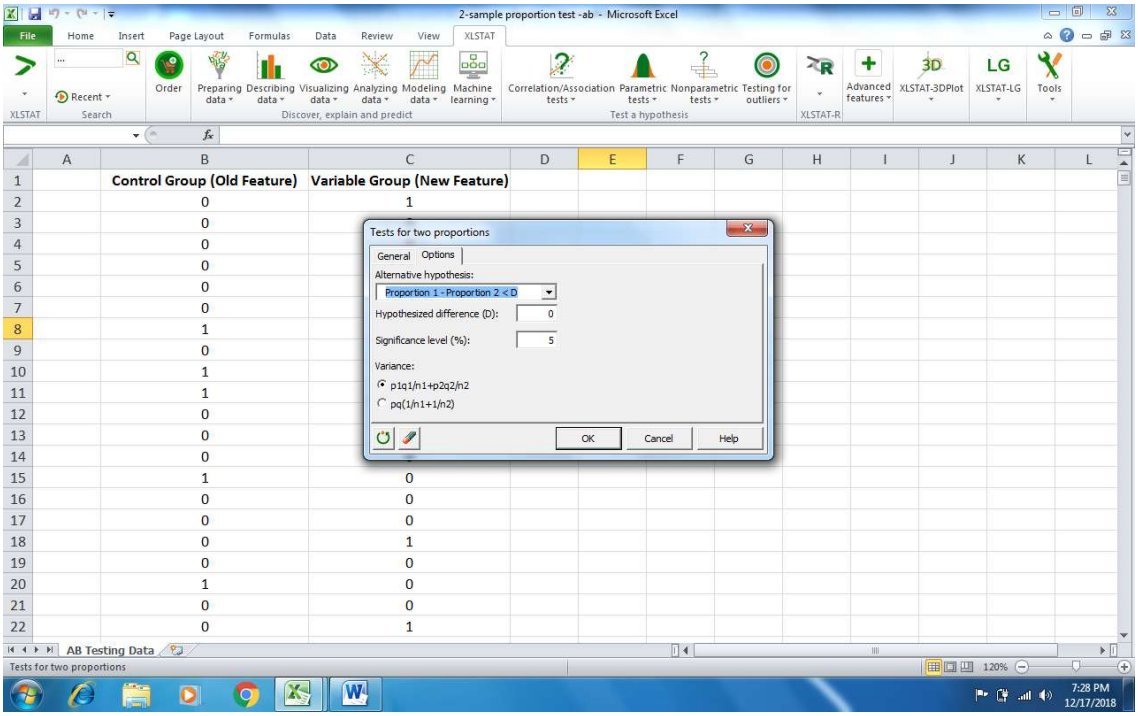
1.Choose parametric test for two proportion from EXSTAT.



2.Given require parameters(sample size & frequency for option1 & option2)



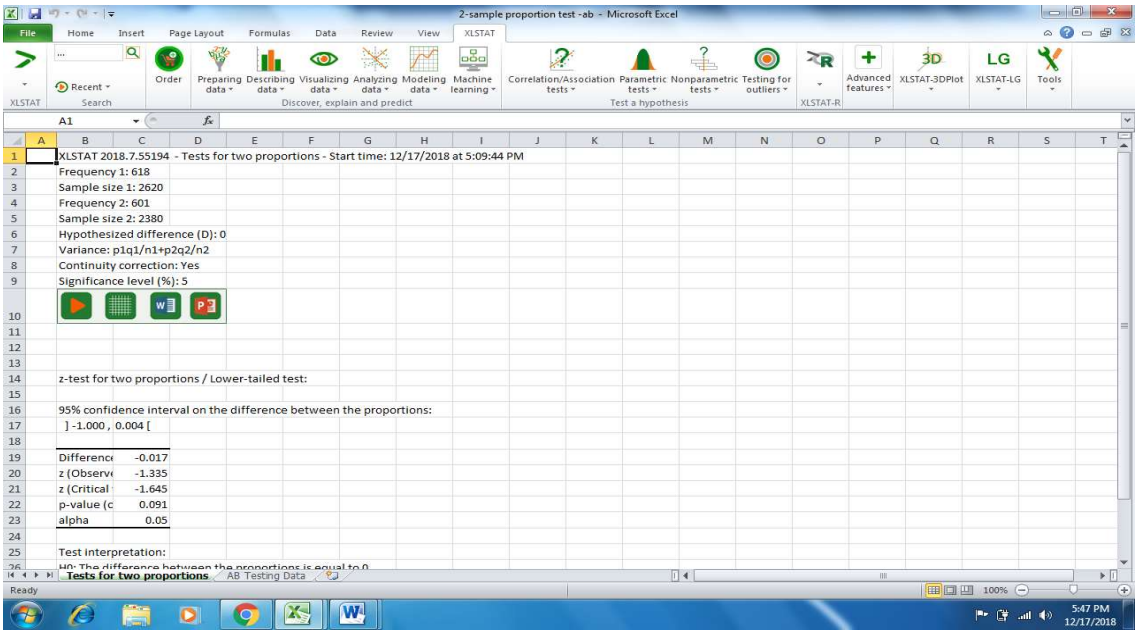
3.Choosing option proportion-1 -proportion-2<D, Hypothesized difference(D):0, Significant level 5% & variance.

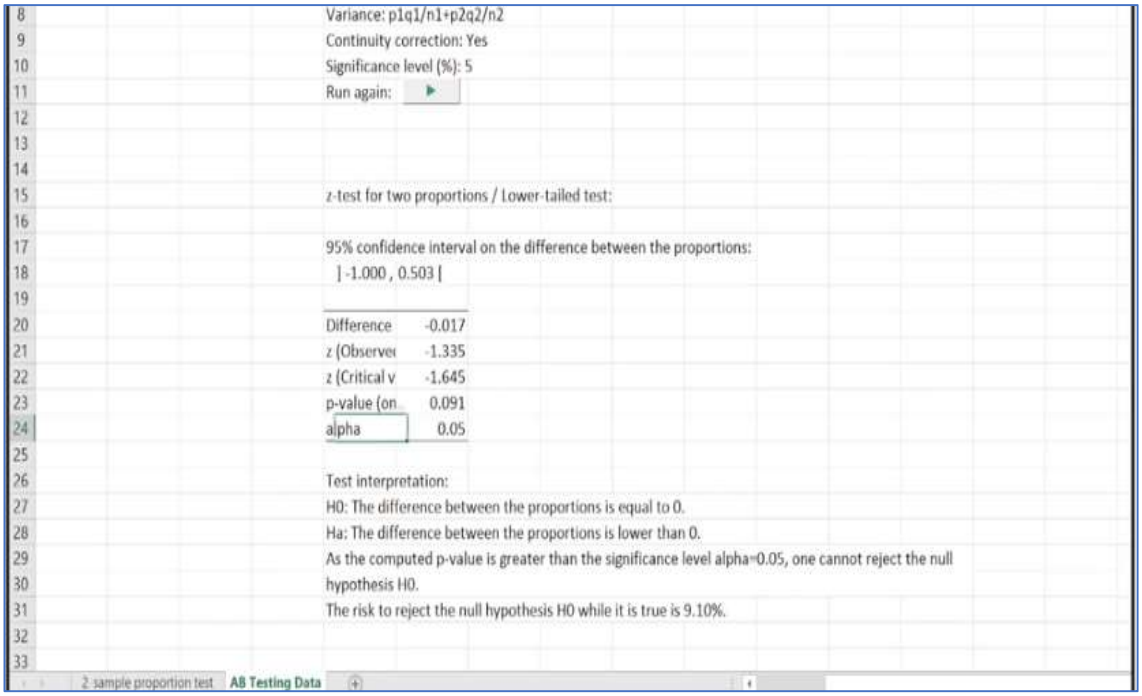


Alternative hypothis (proportion-1 -proportion-2<D)that means proportion-2 is better than proportion-1.

Null hypothesis is proportion-1>=proportion2 which means old features is as good as new feature, even better.

Getting the result





We can make a decision on a test based on the p-value and the significance level of the test. As the p-value (0.091) is greater than the significance level (0.05), we fail to reject the null hypothesis. In this example of A/B testing, the null hypothesis is that the old feature is better or as good as the new feature.

Although we saw initially experimental group 's conversion is25% in comparison of old features 23%, We can't reject null hypothesis.

In summary result is as below

	Visitors	Conversions
Control	2620	618
Variation	2380	601

Significant At	
90% confidence:	YES
95% confidence:	NO
99% confidence:	NO
Z-score	-1.36759
P-value	0.091

Conversion Rate	**Standard Error**
23.59%	0.83%
5.25%	0.89%

90% Conversion Rate Limits	
From	*To*
22.22%	24.96%
23.78%	26.72%

95% Conversion Rate Limits	
From	*To*
21.96%	25.21%
23.51%	27.00%

It shows 90% confident level shows new Tagline(Tagline B) is better than old tagline.

But 95% & 99% confident level is not statistically show enough significant difference & idea is to continue testing with more users.

Since we made experiment live. We get 10,20 or 100 people per day. Second day another 100 people & 25 days we will get 2600 people for testing,

For new features, we may wait for few more days like 2 weeks for getting more data. Then only we can get statistically significant level where we can accept or reject the null hypothesis.

A/B Testing Process:

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

- **Collect Data:** Our analytics will often provide insight into where we can begin optimizing. It helps to begin with high traffic areas of our site or app, as that will allow us to gather data faster. Look for pages with low conversion rates or high drop-off rates that can be improved.

- **Identify Goals:** Our conversion goals are the metrics that we are using to determine whether or not the variation is more successful than the original version. Goals can be anything from clicking a button or link to product purchases and e-mail signups.

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

- **Create Variations:** Using our A/B testing software (like Optimizely), make the desired changes to an element of our website or mobile app experience. This might be changing the color of a button, swapping the order of elements on the page, hiding navigation elements, or something entirely custom. Many leading A/B testing tools have a visual editor that will make these changes easy. Make sure to QA your experiment to make sure it works as expected.

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

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

.