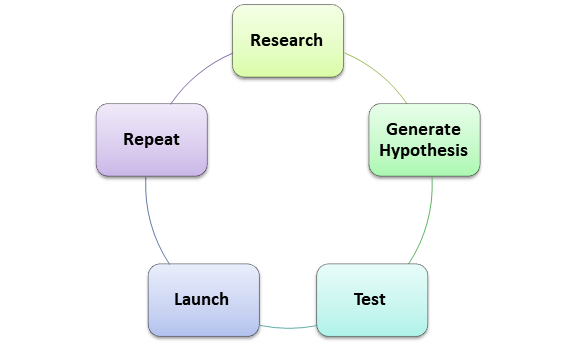
**A/B Testing :**



what variant is in A/B testing?

Variant is a change which we are planning to test with comparison to the default website and whichever turned out to be better we launch it.

Your goal is to work the company understand if they :

1. Should implement the new page,
2. Or keep the old page
3. Or perhaps run the experiment longer to make their decision.

How do we run these experiments?

* We basically design a **conversion funnel** which keeps an accountability of **click through rates** and **conversion rates** for each stage . We define all the steps which takes us to our end target.
* The conversion funnel is a way to understand where exactly the users are dropping off and how can we improve there customer experience for more retention or conversions.

# **What is A/B Testing?**

A/B testing is a general methodology used online when testing product changes and new features. You take two sets of users, show one set of users the changed product (experiment group) and show the second set of users the original product or set of features (control group). You then compare the two groups to determine which version of the product is better.

A/B Testing is important because it:

* Removes the need for guessing or relying on intuition
* Provides accurate answers quickly
* Allows for rapid iteration on ideas
* Establishes causal relationships — not just correlations

**How do we understand whether it is even worth ?**

1. The changes made should be statistically significant.
2. The success of the change is measured by user experience which is basically the click-through-rate (measures the usability), click-through-probability (measures the impact), number of active users on the default and modified web page.

**Metrics**

Metrics help us in determining whether the variation is better or the default product. Any of the metric which we chose in an experiment is based on the user experience. One of the metrics can be the cookies which gives us the user information. While few other metrics include number of active users, number of clicks, number of page views etc.

These Metrics are broadly divided into two categories:

1. Sanity Checks (Invariant Metrics) : The metrics we chose here are **not to be changed** in the overall experiment i.e. between the control and the experiment group.
2. Evaluation Metrics : These metrics are chosen as per the **business goals** that we aim to achieve. These metrics should see a **change** in the overall experiment i.e. between the control and the experiment group.

## ****Why do we need to draw samples from population ?****

Statistical power is the probability that the test rejects the null hypothesis when it should be rejected. It is basically 1 minus beta. A common value for statistical power is 0.80 (so beta is 0.20).

In order to obtain meaningful results, we want our test to have sufficient statistical power. And, sample size influence statistical power. For example, when comparing two means, the follow formula can be used to calculate statistical power. As sample size increases, the statistical power increases. Therefore, for our test to have desirable statistical power (usually 0.80), we want to estimate the minimum sample size required.

<https://towardsdatascience.com/required-sample-size-for-a-b-testing-6f6608dd330a#:~:text=Sample%20Size%20for%20Comparing%20Two,actually%20the%20number%20of%20days>**.**

**Formulating Hypothesis :**

## Null Hypothesis: It states there is no difference between the control and experiment group which means:

## The retention rates are same in both the groups.

1. There is no statistical significant result.

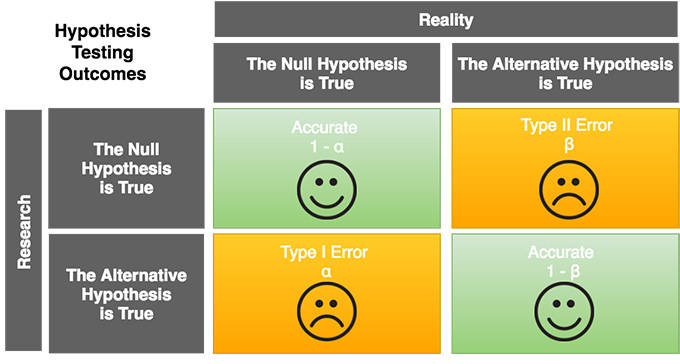
Alternate Hypothesis: There is difference between the control and experiment group which means:

1. Retention rates are different in both groups.
2. It gives Statistical Significant result

Type I and II Error :

There are two types of errors that may occur in our hypothesis testing:

* Type I error: We reject the null hypothesis when it is true. That is we accept the variant B when it is not performing better than A
* Type II error: We failed to reject the null hypothesis when it is false. It means we conclude variant B is not good when it performs better than A
* Concluding that there is a significant difference between the tested rates when in fact it isn’t.



You avoid both of these errors when calculating your sample size.

* To avoid type I errors, you specify a significance level when calculating the sample size.
* To avoid type II errors, you set the power at 0.8 or 0.9 if possible when calculating your sample size, making sure that the sample size is large enough.

When calculating the sample size, you will need to specify the **significance level, power and the desired relevant difference between the rates** you would like to discover**. minimal desired effect(MDE)**

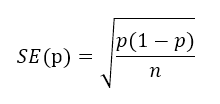
**SAMPLE SIZE**

To have enough statistical power, as well as statistical significance, we need to choose an appropriate sample size. Choosing the appropriate value of alpha(significance level) and beta(statistical power) will give us the exact sample size we need for each variation. Try using this [online calculator](http://www.evanmiller.org/ab-testing/sample-size.html) which gives you a better picture of sample size.

**CONFIDENCE INTERVAL**

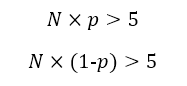
A confidence interval contains the range in which the mean is expected to fall. Confidence interval is calculated using the proportion and the standard deviation, where proportion is the probability of a sample of an event and standard deviation gives us the spread of the curve.

Here, we considered a binomial distribution and a 95% confidence interval. For a binomial distribution , the **standard error** is calculated using this equation:



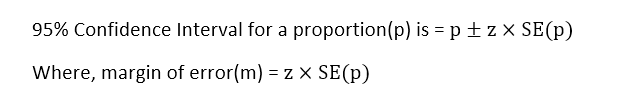
where, p is the the proportion of the event/outcome in a sample and n is the size of sample.

But, when the sample size is large, we prefer using normal distribution. One way to check whether it is a normal distribution or not is:



Calculating the confidence Interval:

The click through probability gives us the proportion of the users who clicked/the number of users . This estimated value acts as the center value of the confidence Interval.

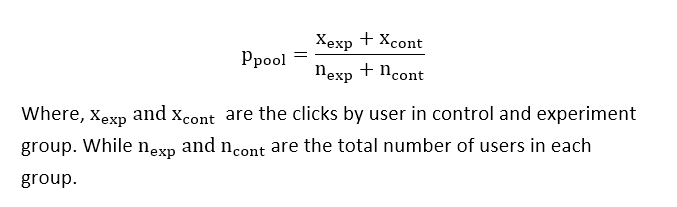


The value of z in this case would be 1.96.

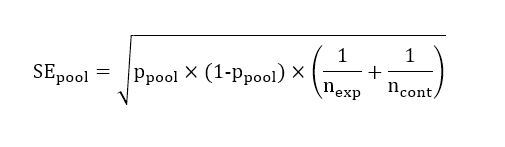
**COMPARISON OF SAMPLES PROPORTION**

When the two estimated proportions are different i.e. control and experiment proportion, it may be due to a difference in the populations or it may be due to chance. It is important to understand which group will give better results i.e. better statistically significance.

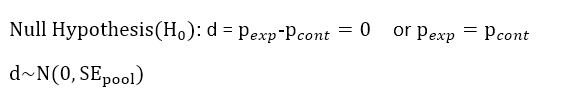
In this test, we use the pooled proportion. It is calculated as:



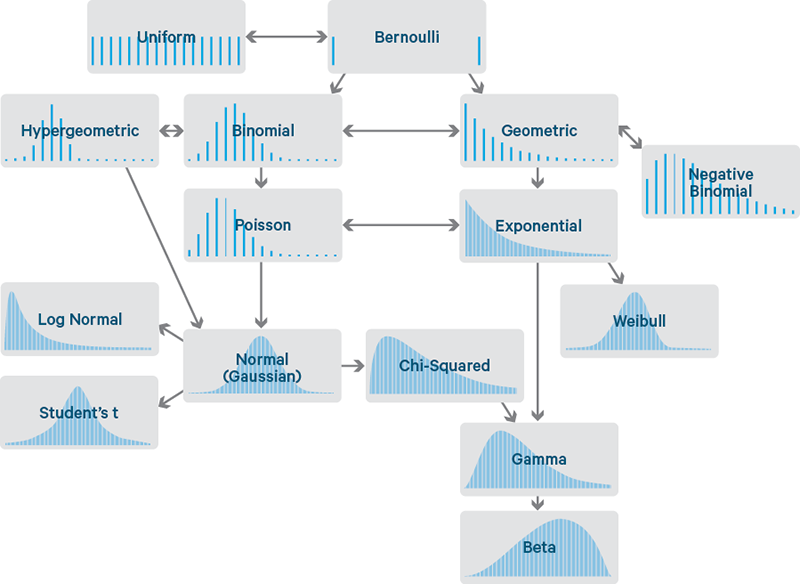
Standard pool error calculation:

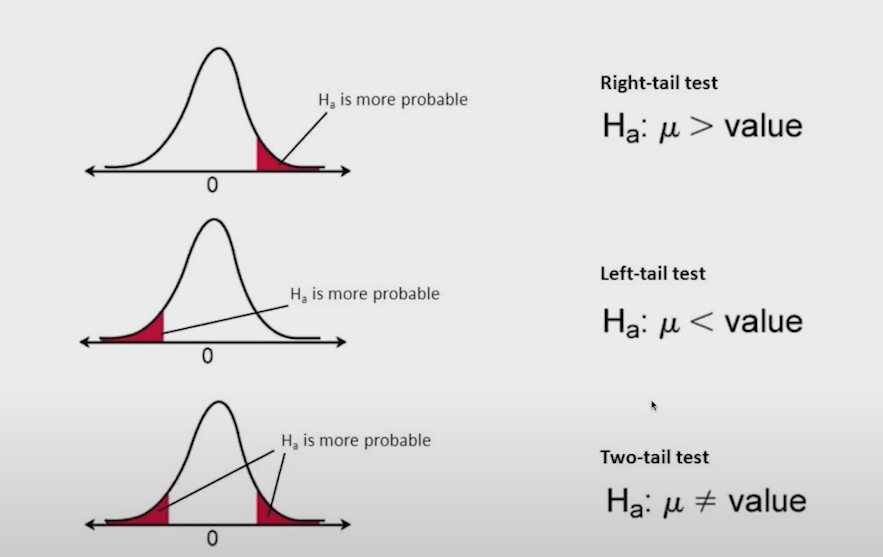


To accept a null hypothesis hypothesis the mean difference between two independent samples should be 0.

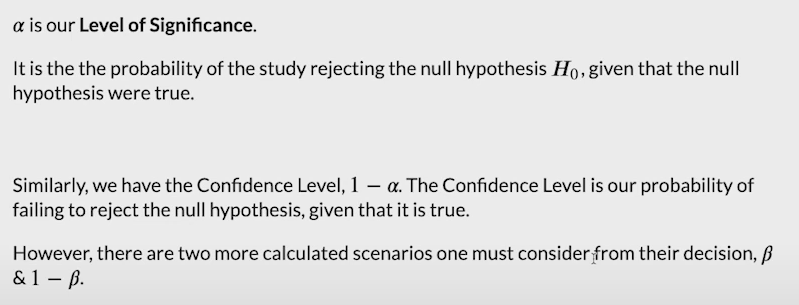


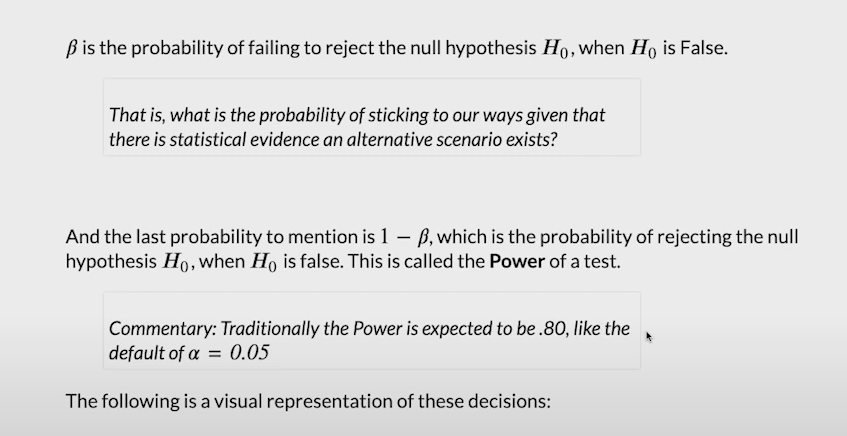
If the distribution for the differences is not equal to 0, we accept the alternative hypothesis.





Key terms:





https://towardsdatascience.com/data-science-you-need-to-know-a-b-testing-f2f12aff619a

 First need to form our question as a

1. **hypothesis**, we then need to work out our
2. **randomization strategy**, **sample size** and finally our
3. **method of measurement**.

## Determine Sample Size Required

Next we want to figure out how many data points will be required to run an experiment. There are four test parameters that need to be set to enable the calculation of a suitable sample size:

1. **Baseline rate** — an estimate of the metric being analyzed before making any changes
2. **Practical significance level**— the minimum change to the baseline rate that is useful to the business, for example an increase in the conversion rate of 0.001% may not be worth the effort required to make the change whereas a 2% change will be
3. **Confidence level**— also called significance level is the probability that the null hypothesis (experiment and control are the same) is rejected when it shouldn’t be
4. **Sensitivity**— the probability that the null hypothesis is not rejected when it should be

# **How To Calculate Sample Size For Clinical Trials in 5 Steps**

<https://www.youtube.com/watch?v=qjK7ASBbW74&list=PLMs0YrKTW-JqDLhKZOB-Y4Xx25LvQWESd>

https://ryanwingate.com/assets/projects/A-B\_Test\_Result\_Analysis.html

https://stackoverflow.com/questions/19746350/how-does-one-change-color-in-markdown-cells-ipython-jupyter-notebook

https://github.com/statsmodels/statsmodels/issues/3931

<https://stackoverflow.com/questions/13851535/how-to-delete-rows-from-a-pandas-dataframe-based-on-a-conditional-expression>

Confidence Interval and Confidence level :

|  |  |  |
| --- | --- | --- |
| **Significance Level** | **Confidence level** | **Confidence Interval** |
| * In a hypothesis test, the significance level, alpha, is the probability of making the wrong decision when the [null hypothesis](https://www.statisticshowto.com/probability-and-statistics/null-hypothesis/) is true. | * The probability that if a poll/test/survey were repeated over and over again, the results obtained would be the same. A confidence level =  1 - alpha. | * <http://www.stat.yale.edu/Courses/1997-98/101/confint.htm> * A range of results from a poll, experiment, or survey that would be expected to contain the population parameter of interest. For example, an average response. Confidence intervals are constructed using significance levels / confidence levels. * <https://www.statisticshowto.com/probability-and-statistics/confidence-interval/#WhatisCI> * https://www.statisticshowto.com/probability-and-statistics/confidence-interval/ |

## What is a Margin of Error?

* A **margin of error** tells you **how many percentage points your results will differ**from the real population value. For example, a 95% [confidence interval](https://www.statisticshowto.com/probability-and-statistics/confidence-interval/) with a 4 percent margin of error means that your [statistic](https://www.statisticshowto.com/statistic/)will be within 4 percentage points of the real population value 95% of the time.
* The **margin of error**is the [range](https://www.calculushowto.com/types-of-functions/domain-and-range-of-a-function/)of values below and above the [sample statistic](https://www.statisticshowto.com/sample-statistic-definition-examples/) in a [confidence interval](https://www.statisticshowto.com/probability-and-statistics/confidence-interval/).

**The margin of error can be calculated in two ways, depending on whether you have**[**parameters**](https://www.statisticshowto.com/what-is-a-parameter-statisticshowto/)**from a population or**[**statistics**](https://www.statisticshowto.com/statistic/)**from a sample**:

1. Margin of error = Critical value x [Standard deviation](https://www.statisticshowto.com/relative-standard-deviation/) for the population.
2. Margin of error = Critical value x [Standard error](https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/what-is-the-standard-error-of-a-sample/) of the sample.

## How to Calculate Margin of Error: Steps

Step 1: **Find the critical value**. The critical value is either a [**t-score**](https://www.statisticshowto.com/probability-and-statistics/t-distribution/t-score-formula/) or a **z-score**. If you aren’t sure, see: [T-score vs z-score](https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/t-score-vs-z-score/). In general, for small [sample sizes](https://www.statisticshowto.com/probability-and-statistics/find-sample-size/) (under 30) or when you don’t know the population [standard deviation](https://www.statisticshowto.com/probability-and-statistics/standard-deviation/), use a [t-score](https://www.statisticshowto.com/probability-and-statistics/t-distribution/t-score-formula/). Otherwise, use a [z-score](https://www.statisticshowto.com/probability-and-statistics/z-score/).  
[Click here for a minute video that shows you how to find a critical value.](https://www.youtube.com/watch?v=RAnFyF_6zHk)

Step 2: **Find the**[**Standard Deviation**](https://www.statisticshowto.com/probability-and-statistics/standard-deviation/)**or the**[**Standard Error**](https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/what-is-the-standard-error-of-a-sample/)**.** These are essentially the same thing, only you must know your population [parameters](https://www.statisticshowto.com/what-is-a-parameter-statisticshowto/)in order to calculate standard deviation. Otherwise, calculate the standard error (see: [What is the Standard Error?](https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/what-is-the-standard-error-of-a-sample/)).  
[Click here](https://www.youtube.com/watch?v=aBXJnvQ6KFk)for a short video on how to calculate the standard error.

Step 3: **Multiply the**[**critical value**](https://www.statisticshowto.com/probability-and-statistics/t-distribution/t-score-formula/)from Step 1**by the**[**standard deviation**](https://www.statisticshowto.com/probability-and-statistics/standard-deviation/) or [standard error](https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/what-is-the-standard-error-of-a-sample/) from Step 2. For example, if your CV is 1.95 and your SE is 0.019, then:  
1.95 \* 0.019 = 0.03705

[moe](https://www.statisticshowto.com/wp-content/uploads/2013/08/moe.png)  
  
  
Where:

* [phat](https://www.statisticshowto.com/wp-content/uploads/2009/10/phat.bmp)= sample proportion (“P-hat”),
* n = sample size,
* z = z-score.

## What is the standard error?

<https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/what-is-the-standard-error-of-a-sample/>

<https://www.statisticshowto.com/probability-and-statistics/standard-deviation/>

## Resources :

How To Calculate A/B Testing Sample Size

https://vwo.com/blog/how-to-calculate-ab-test-sample-size/

<https://www.invespcro.com/blog/calculating-sample-size-for-an-ab-test/>

<https://data36.com/statistical-significance-in-ab-testing/>

<https://www.jwilber.me/permutationtest/>

<https://www.evanmiller.org/how-not-to-run-an-ab-test.html>

<https://towardsdatascience.com/required-sample-size-for-a-b-testing-6f6608dd330a>

<https://mcgaw.io/blog/how-to-build-an-a-b-testing-framework-our-vice-framework-revealed/#gs.0n640a>

https://www.analyticsvidhya.com/blog/2020/10/ab-testing-data-science/