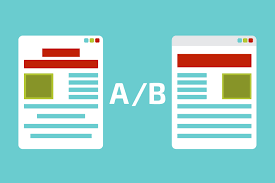
**What is A/B testing?**

A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.

In A/B testing, A refers to ‘**control**’ or the original testing variable. Whereas B refers to ‘**variation**’ or a new version of the original testing variable.

A/B testing is one of the most prominent and widely used statistical tools.

**How does A/B Testing Work?**



Let’s say there is an e-commerce company XYZ. It wants to make some changes in its newsletter format to increase the traffic on its website. It takes the original newsletter and marks it A and makes some changes in the language of A and calls it B. Both newsletters are otherwise the same in color, headlines, and format.

**Testing process:**

**Objective**

Our objective here is to check which newsletter brings higher traffic on the website i.e the conversion rate.

To perform A/B testing we have to make some hypothesis:

**1. Make a Hypothesis:**

“**By making changes in the language of the newsletter, we can get more traffic on the website**”.

In hypothesis testing, we have to make two hypotheses i.e **Null hypothesis** and the **alternative hypothesis**. Let’s have a look at both.

1. **Null hypothesis or H0:**

The null hypothesis is the one that states that sample observations result purely from chance. From an A/B test perspective, the null hypothesis states that there is **no difference between the control and variant groups**. It states the default position to be tested or the situation as it is now, i.e. the status quo. Here our H0 is ” there is **no difference in the conversion rate in customers receiving newsletter A and B”.**

1. **Alternative Hypothesis or H0:**

The alternative hypothesis challenges the null hypothesis and is basically a hypothesis that the researcher believes to be true. The alternative hypothesis is what you might hope that your A/B test will prove to be true.

**2.Create Control Group and Test Group**

**The Control group**, and **the Test (variant) group**.

The Control Group is the one that will receive newsletter A and the Test Group is the one that will receive newsletter B.

**3.Conduct the A/B Test and Collect the Data**

One way to perform the test is to calculate **daily conversion rates** for both the treatment and the control groups. Since the conversion rate in a group on a certain day represents a single data point, the sample size is actually the number of days. Thus, we will be testing the difference between the mean of daily conversion rates in each group across the testing period.

When we run our experiment for one month, we noticed that the mean conversion rate for the Control group is 16% whereas that for the test Group is 19%.

Deciding whether to adopt the changes from the experiment would depend on not only the statistical significance but also on other factors such as the cost of implementation, user feedback, and other business considerations.

**Statistical significance** of our test:

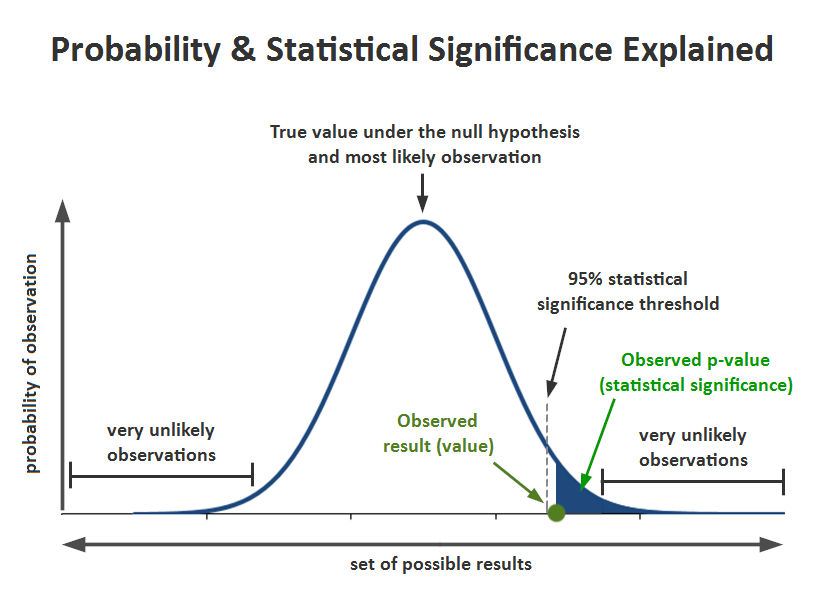
There are two types of errors that may occur in our hypothesis testing:

1. **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
2. **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

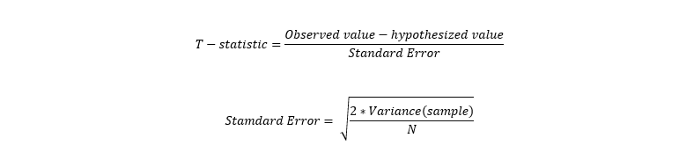
To avoid these errors we must calculate the statistical significance of our test.

That means the difference between your control version and the test version is not due to some error or random chance. To prove the statistical significance of our experiment we can use a two-sample T-test.

The two–sample t–test is one of the most commonly used hypothesis tests. It is applied to compare whether the average difference between the two groups.



1. **Significance level (alpha):** The significance level, also denoted as alpha or α, is the probability of rejecting the null hypothesis when it is true. Generally, we use the significance value of 0.05
2. **P-Value:** It is the probability that the difference between the two values is just because of random chance. P-value is evidence against the null hypothesis. The smaller the p-value stronger the chances to reject the H0. For the significance level of 0.05, if the p-value is lesser than it hence we can reject the null hypothesis
3. **Confidence interval:** The confidence interval is an observed range in which a given percentage of test outcomes fall. We manually select our desired confidence level at the beginning of our test. Generally, we take a 95% confidence interval



**What Mistakes Should we Avoid While Conducting A/B Testing?**

There are a few key mistakes I’ve seen data science professionals making. Let me clarify them for you here:

* **Invalid hypothesis**: The whole experiment depends on one thing i.e the hypothesis. What should be changed? Why should it be changed, what the expected outcome is, and so on? If you start with the wrong hypothesis, the probability of the test succeeding, decreases

**Testing too Many Elements Together:** Industry experts caution against running too many tests at the same time. Testing too many elements together makes it difficult to pinpoint which element influenced the success or failure. Thus, prioritization of tests is indispensable for successful A/B testing

**Ignoring Statistical Significance:**It doesn’t matter what you feel about the test. Irrespective of everything, whether the test succeeds or fails, allow it to run through its entire course so that it reaches its statistical significance

**Not considering the external factor:**Tests should be run in comparable periods to produce meaningful results. For example, it is unfair to compare website traffic on the days when it gets the highest traffic to the days when it witnesses the lowest traffic because of external factors such as sale or holidays

**When Should We Use A/B Testing**?

A/B testing works best when testing incremental changes, such as UX changes, new features, ranking, and page load times