A/B Testing Basics

An A/B test is a controlled experiment that compares two versions of one thing, such as a web page, by showing each version to a different random group of people. Version A is the current one, called the control, and version B is the change, called the variant. The method lets a team learn from real behaviour, not from opinions. We follow Rami, a product manager at an online bookshop.

1. Start with a Hypothesis

Every good test begins with a clear question, written as a hypothesis, such as that changing the button text will increase purchases. We choose one main success metric in advance, for example the share of visitors who buy. We also choose guardrail metrics, such as page speed or refund rates, that must not get worse. Testing one change at a time makes the result easy to understand.

Example: A baker has a clear question: will more customers choose the cookie made with a new recipe? The baker decides to measure one thing, the number of cookies sold.

At work: Rami's hypothesis is that the button text "Buy now" will lead to more purchases than "Add to basket". His main metric is the share of visitors who buy, and his guardrail metric is the refund rate.

2. Split Fairly

Next, we assign people to the groups at random, so that both groups are alike in everything except the change. Each person should see the same version every time, and both versions should run at the same time, so that holidays or news do not favour one group. Random assignment is what lets us say that a difference was probably caused by the change.

Example: A café offers a sample of a new drink to randomly chosen tables, while the other tables get nothing, on the same day. Because the tables were chosen at random, a difference in orders is more likely to come from the sample.

At work: Rami's system assigns each visitor to A or B by chance, and the visitor sees the same version on every visit. Both versions run at the same time, so a weekend sale cannot favour one of them.

3. Plan the Sample Size

Before starting, we decide how many people we need and how long to run the test. A small sample can only detect large differences, and a larger sample can detect smaller ones, so a

This document is a property of Asaseeyat Academy - © Copy Right Protected.

sample size calculator uses the expected change to give an answer. It is wise to run the test for full weeks, so that every day of the week is included. We should not stop early just because the result looks good, because early results move around a lot.

Example: A coin that lands on heads three times in a row is not proof of a trick coin, because a few flips tell us very little. We need many flips before we can trust a pattern, just as we need many visitors in a test.

At work: A sample size calculator tells Rami that he needs about 8,000 visitors per group to detect a rise in purchases from 5 percent to 6 percent. He chooses 20,000 per group to be safe and runs the test for two full weeks. On day three the result looks great, but he waits until the end as planned.

4. Read the Results

When the test ends, we compare the groups on the main metric and check whether the difference is bigger than chance would normally produce. A common tool is the p-value: a value below a threshold chosen in advance, often five percent, suggests that chance alone is an unlikely explanation. We also look at the size of the difference, and at whether it is large enough to matter for the business. Finally, we check the guardrail metrics and record what we learned, even when the variant loses.

Example: A school tries two revision methods. A difference of one mark among six students could easily be chance, but the same difference among six hundred students would be more convincing.

At work: In Rami's test, 5 percent of group A bought and 6 percent of group B bought. The p-value is far below five percent, the rise is large enough to matter and refunds did not increase. He launches "Buy now" and records what he learned.

Conclusion

In summary, an A/B test compares a control and a variant on random groups to learn what really works. Four steps put it to work: start with a hypothesis, split fairly, plan the sample size and read the results. Used with discipline, it lets evidence replace opinion.

Sources

• Kohavi, R., Tang, D. and Xu, Y. (2020). Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing. Cambridge University Press – https://experimentguide.com/
• NIST/SEMATECH e-Handbook of Statistical Methods (hypothesis tests and p-values) – https://www.itl.nist.gov/div898/handbook/

This document is a property of Asaseeyat Academy - © Copy Right Protected.