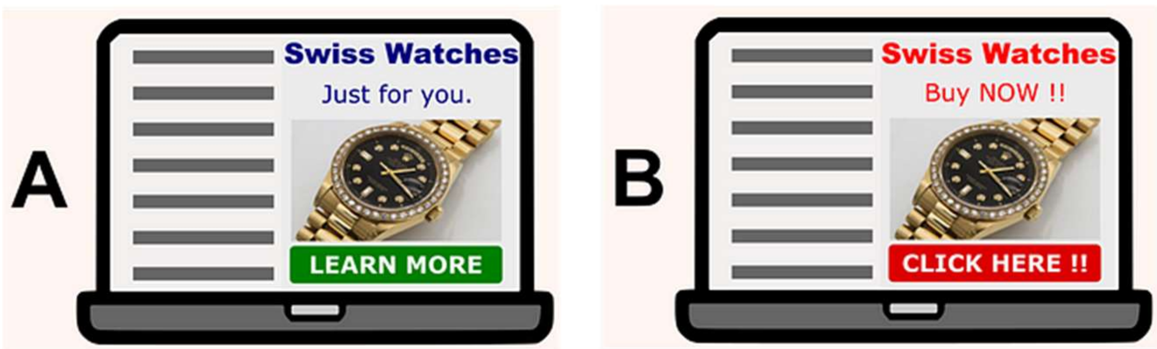
— IRONHACK

A/B Testing and Optimization Strategies

AB TESTING & OPTIMIZATION FUNDAMENTALS

Introduction to A/B Testing

A/B testing, or **split testing**, is a method of comparing two versions of a webpage or app against each other to determine which one performs better. **In classic statistics, we call it hypothesis testing. Optimization** is the ongoing process of using A/B testing results and other data to continuously improve the user experience and achieve business objectives.

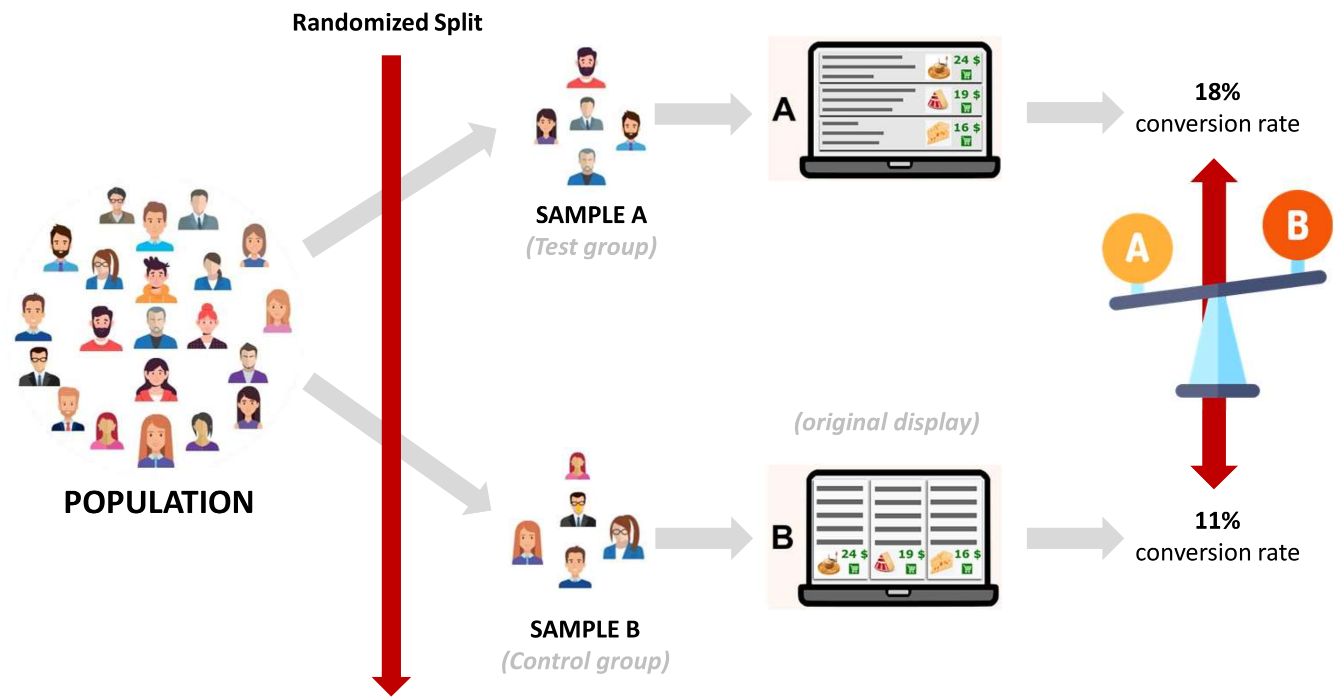


What KPIs do we usually look at here?

- Discrete or binomial metrics (0 or 1 values only possible): *Click-through rate (if a user is shown and advertisement, do they click on it?), conversion rate (if a user is shown an advertisement, do they convert into customers?) and Bounce rate (if a user visits a website, is the following visited page on the same website?)*
- Continuous metrics or non-binomial metrics: *Average revenue per user (how much revenue does a user generate in a month?), Average session duration (for how long does a user stay on a website in a session?) and Average Order value (what is the total value of the order of a user?)*

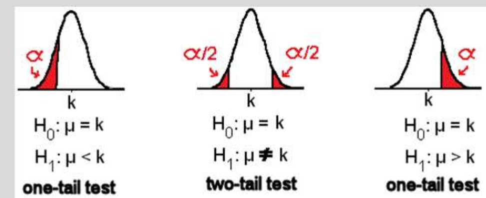
AB TESTING & OPTIMIZATION FUNDAMENTALS

Introduction to A/B Testing



Minimum Detectable Effect (MDE) is the smallest change or difference you want to be able to spot with your test. It helps decide how many people or items you need to include in your test to confidently say there's a real change and it's not just chance.

With the data we collected from the activity of users of our website, we can compare the efficacy of the two designs A and B. Simply comparing mean values wouldn't be very meaningful, as we would fail to assess the **statistical significance** of our observations. It is indeed fundamental to determine how likely it is that the observed discrepancy between the two samples originates from chance.



Our **null hypothesis H0** is that the two designs A and B have the same efficacy, i.e. that they produce an equivalent click-through rate, or average revenue per user, etc. The statistical significance is then measured by the **p-value**, i.e. the probability of observing a discrepancy between our samples at least as strong as the one that we actually observed.

$$p_{val} = p(\text{data at least as extreme as actual observation} \mid H_0)$$

The p-value is a measure used in statistics to help you determine whether your results are due to chance. If you have a p-value of 5%, it means that if there were no real effect or difference (like no difference between two groups you're comparing), **you would still see the result you got (or something more extreme) about 1 out of 20 times just by random chance.** It's like flipping a coin and getting heads many times in a row – a low p-value suggests it might be a weighted coin rather than a fair one.

AB TESTING & OPTIMIZATION FUNDAMENTALS

Introduction to A/B Testing

Choosing a statistical test for A/B testing is like picking the right key for a lock, ensuring we correctly interpret our data.

•For categories (like survey responses):

- Use **Pearson's chi-squared test** when you have a lot of data.
- Use **Fisher's exact** test for smaller datasets.

•For numbers (like test scores):

- If you know how spread out the data is (variance), consider a **Z-test**.
- If you're unsure about the spread, but it's roughly equal between groups, use a **Student's t-test**.
- If the spread is unequal, pick **Welch's t-test**.

•If the data is skewed or unusual:

- The **Mann-Whitney U test** works well when data doesn't follow a bell curve.

This guide helps ensure we choose a test that fits our data's story, giving us confidence in our A/B testing results.