

A/B Tests

Split Or Bucket Testing

Contents



- I. Introduction
- II. Usage of A/B Tests
- III. Set-Up
- IV. Outcome & Evaluation
- V. Example

Introduction

- also known as split or bucket testing
- is a method in the category of „testing“
- conducted mainly digitally
- enables quantitative measures
- It works by comparing two alternate variants - A and B - of a website, app or digital content against each other to identify the better performing one.

Design Phases



Usage

A/B-Testing is useful in the following cases:

- to evaluate your concept in operation and real terms
- to understand the effects of different concepts on your users/ customers
- in the testing phase of a design or development process
- if a website, product or prototype exists in at least two variants

Set-Up

Preparation	Problem	Research	Hypotheses	Test	Analysis
<p>Two concepts/ designs ready to be tested</p> <p>Determine sample size, variance, significance level and "stopping rules"</p>	<p>Define a problem that should be improved</p> <p>E.g. a concrete element with low response rates</p>	<p>Try to understand the problem as good as possible to find about its cause</p> <p>E.g. benchmark analysis</p>	<p>Form concrete hypotheses about the problem</p> <p>They serve as a guideline for your test</p>	<p>The test groups need to have the same size, they either see A or B</p> <p>Computer mediated- by using a testing tool</p>	<p>Examine the test results with regard to the hypotheses</p> <p>Usually, the winner of the test will be implemented</p>

Instructions

Outcome and Evaluation



The outcome of the A/B testing method is a clear result of the performance of A and B → A “winner” of both alternatives.

This can be very useful in marketing and customer experience, e.g. to improve a website’s quality, extend the users’ stay on it, reduce bounce rate or increase conversion rate.

Strengths:

- Feasible with reasonable effort if conducted with the right tools.
- Unambiguous, objective result with clear implications → empirical foundation for decisions.
- Focus on test group (user/customer base), not individual's perception.
- Can be extended to A/B/C etc. Strong in combination with heat / click maps.

Weaknesses:

- Danger of interpreting data wrong → statistical preparation crucial.
- Does not work optimally for pages with low traffic or non-digitally.
- No explanations or follow-ups possible, no insights on the “why”.

Instructions

Outcome and Evaluation



You are marketing manager of a fair exhibitor. To gain more traffic on fairs, you send out mail newsletters to all customers, in which they can sign up for a free coffee (→ **A**) at your booth. You are unhappy about the response rate of the recent fairs and want to improve your invite.

Problem: response rate to the mail is not high, customers don't take the effort to sign up

Research: you find out that competitors are more successful if they offer other types of beverages

Hypothesis: a cold beverage as free drink (→ **B**) will increase the response rate

- You have developed an alternate version to your current invite and calculated sample size + statistical parameters
- Set up in an adequate tool and run the test
- You test with a sample size of 600, significance level of <5% splitting your contacts randomly
- Conversion rate of A: 10%, B: 35%
- **Version B is performing significantly better and will therefore replace A!**



Contact

usability-siegen.de

fablab-siegen.de