A/B testing fundamentals

You have a hunch. Maybe the red button will get more clicks. Maybe the longer headline will tank your conversion rate. **A/B testing fundamentals** are the set of rules and mental models that turn those hunches into measurable, reliable decisions. It is not about guessing. It is about creating a controlled experiment where you show version A to half your traffic and version B to the other half, then measure which one performs better on a single, pre-defined metric. Without these basics, you are just flipping coins with a spreadsheet.

The single-variable trap: why most people invalidate their own tests

You change the headline, the button color, and the image all at once. The test shows a 12% lift. But what caused it? You have no idea. That is the cardinal sin. The entire point of controlled experimentation is isolation. Change one variable. Only one. If you want to test the headline, keep the button blue. Keep the image the same. Keep the font, the spacing, the offer. If you change multiple things, you have a multivariate test, which requires a much larger sample size and a completely different statistical framework. Most people run a multivariate test and call it an A/B test. It is not. It is noise.

Statistical significance is not a suggestion

You ran the test for three days. Version B is winning by 8%. You ship it. Three weeks later, conversions drop. What happened? You hit the "publish" button before reaching statistical significance. This is the most common failure pattern in the industry. Statistical significance (usually calculated at a 95% confidence level) tells you that the observed difference is likely not due to random chance. If you stop the test early, you are literally making decisions on random noise. Use a calculator like the one from [Evan Miller] to determine the required sample size before you start. Do not peek at the results every day. Do not stop early because you "see a trend." Trends lie.

> Rule of thumb: If you are tempted to check the results before the sample size is reached, you have not waited long enough. Stop checking. Let the test run.

Sample size, duration, and the weekday effect

A common mistake is running a test for exactly one week. But what if your traffic patterns are different on weekends? What if you launched the test on a Tuesday and the control group got a spike from a newsletter that went out on Wednesday? The minimum duration for most B2B SaaS tests is two full business cycles, usually 14 days. This accounts for weekly seasonality. If you have a low-traffic site, you might need 30 days or more. Do not use a sample size calculator that only asks for baseline conversion rate and minimum detectable effect. You also need to account for the fact that your traffic is not perfectly uniform. Use a tool like [Optimizely's sample size calculator] and be honest about your minimum detectable effect. If you want to detect a 5% lift, you need a lot of traffic. If you only have 1000 visitors a month, you can only reliably detect huge swings. Accept that.

Three mistakes that kill your test before it starts

- **Peeking at results:** Every time you check the data and make a decision before the sample size is reached, you inflate your false positive rate. It is not a harmless look. It is p-hacking.
- **Testing too many variations:** Running 5 variations of a landing page with 2000 visitors each is not a test. It is a lottery. Stick to A/B. If you must run A/B/C, triple your sample size.
- **Ignoring the novelty effect:** A new design often gets a temporary boost because it is different. Users click the shiny new thing. After two weeks, the novelty wears off and the effect disappears. Run the test long enough to burn through the novelty.

What to test when you have no idea what to test

If you are staring at a blank page, start with the highest friction points in your funnel. Look at your analytics. Where do people drop off? The checkout page? The signup form? The pricing page? That is where you test. Do not test the color of the "Buy Now" button on a page that only 2% of visitors ever reach. Test the page that 80% of your traffic sees and bounces from. A classic starting point is the headline. Headlines carry disproportionate weight. Change the headline from a feature statement ("Enterprise-grade security") to a benefit statement ("Your data stays yours. Always.") and measure the difference. Another high-impact test is the call-to-action text. "Get started" versus "Try free for 14 days" can produce wildly different results. Test one thing at a time.

Myth vs reality: three common A/B testing lies

Myth 1: "You need 10,000 visitors per variation to get reliable results."
Reality: It depends on your baseline conversion rate and the size of the effect you want to detect. A site with a 10% conversion rate can reach significance faster than a site with a 1% conversion rate. Use a calculator, not a guess.

Myth 2: "If the test is not significant, the changes are worthless."

Reality: A non-significant result is still a result. It tells you the change is not large enough to be reliably measured with your current traffic. That is useful information. It means you either need more traffic, a bigger change, or the change does not matter.

Myth 3: "A/B testing always gives you a clear winner."
Reality: Many tests end inconclusively. The two versions are functionally equivalent. That is fine. It means you can pick either version and move on. Do not force a winner where none exists.

Decision tree: should you even run this test?

Before you fire up your testing tool, ask yourself these questions in order. If the answer is "no" at any point, stop. Do not run the test.

Question 1: Do you have a clear, measurable hypothesis? (e.g., "Changing the CTA from 'Learn More' to 'Start My Free Trial' will increase click-through rate by at least 10%.")
If no: Go back and write a hypothesis. Do not test without one.
If yes: Proceed.

Question 2: Can you isolate a single variable?
If no: Simplify the test. Remove the extra variables.
If yes: Proceed.

Question 3: Do you have enough traffic to detect the smallest effect you care about within two weeks?
If no: Either accept that you can only detect large effects, or get more traffic before testing.
If yes: Run the test. Do not peek.

The one thing you must do after the test ends

You have a winner. You ship it. Good. Now document the test. Write down the hypothesis, the sample size, the duration, the result, and the confidence level. This is not busywork. It is the only way to build an institutional memory of what works and what does not. Without documentation, you will run the same test six months later because nobody remembers the outcome. That is wasted time. Store it in a shared drive, a wiki, or a Notion page. Make it searchable. And if the test failed, document that too. A failed test that is documented is infinitely more valuable than a successful test that is forgotten.

> ### Technical Verification Node
>
> read more here

Report ID: 9A4202E4 | Signature: 4b3f0899cb4c80248224af1e6c7e5412