

A/B Testing

Business & Management

Running an experiment that produces a trustworthy answer.

The discipline is in the design, not the analysis.

1 State one hypothesis

What change, on what metric, by how much.

2 Choose the primary metric

One. Guardrail metrics may be watched, not optimised.

3 Calculate the sample size

From the baseline rate and the effect you care about detecting.

4 Randomise properly

By user, not by session. Assign once and keep it.

5 Run the full duration

At least one full weekly cycle. Do not peek and stop.

6 Analyse once

Against the pre-declared metric.

COMMON ERRORS

Stopping when it looks significant

Testing many metrics and reporting the winner

Too small a sample to detect the effect

Ignoring novelty effects

Peeking and stopping when $p < 0.05$ raises the real false-positive rate far above 5%. Fix the duration in advance.