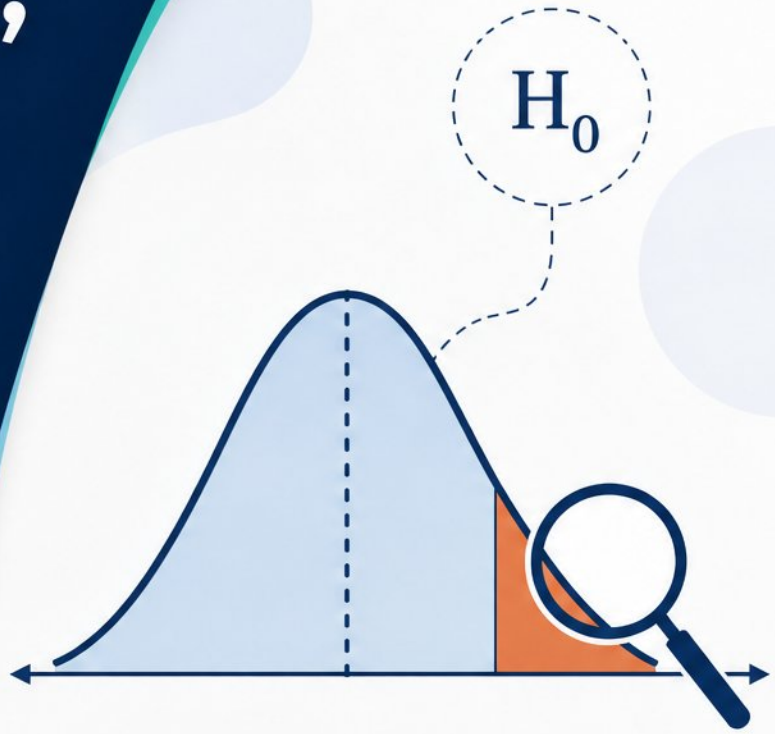
P Value in Statistics: Concepts, Purpose, and Examples

- Learn what a p-value is, why it is used, and how to interpret it correctly
- Agenda covers hypothesis testing foundations, common misinterpretations, and applied examples
- Built for data analysts, students, and business analytics teams with mixed backgrounds

Anchor: probability of seeing a result at least as extreme as yours, assuming the null hypothesis is true



Why P-Values Exist: The Problem of Sampling Variability

- Sample statistics imperfectly estimate population parameters
- Random sampling error creates variation across samples
- Same study can show different results by chance alone
- P-values ask: how surprised should we be if nothing real happened?

Null Hypothesis Significance Testing: The Framework

- Null hypothesis (H_0): the "nothing happening" explanation, such as no difference or no effect
- Alternative hypothesis (H_1): the claim of a real relationship or difference in the population
- NHST follows a fixed logic: assume H_0, compute how extreme the data are, then decide whether to reject H_0
- Statistical significance is about evidence against H_0, not about practical importance

Test Statistics and the Significance Level Alpha

- Test statistic converts data into one measurable signal: t, z, chi-square
- P-value compares the test statistic against a reference distribution
- Alpha is a pre-specified threshold, conventionally 0.05
- Alpha defines the false-positive tolerance for the test
- p = 0.049 and p = 0.051 are nearly identical evidence

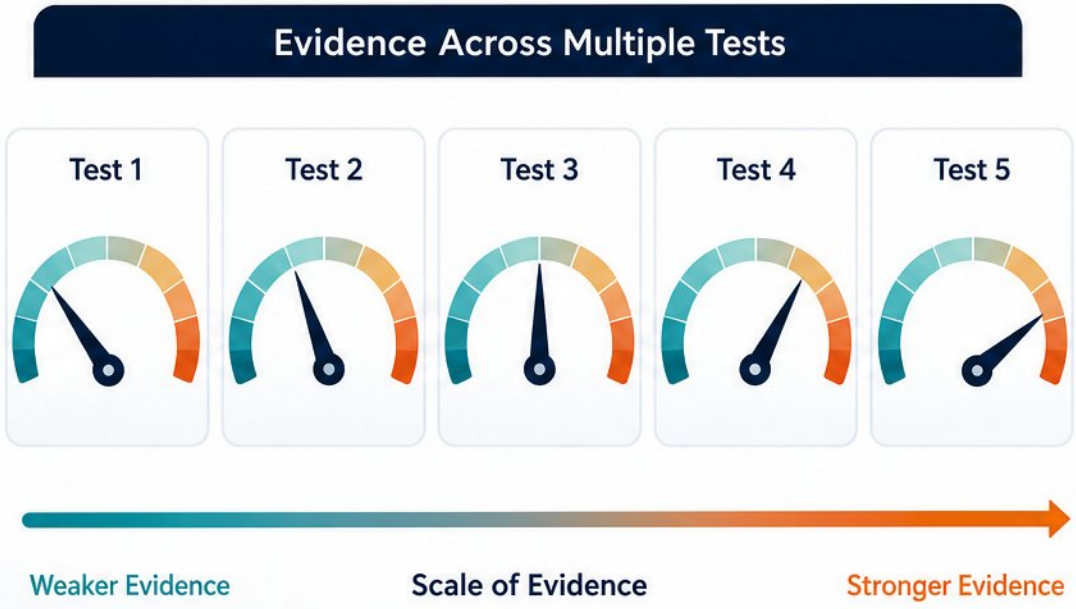
Defining the P-Value with Precision

- Probability of data at least this extreme, if H0 is true
- A statement about the data, not about the hypothesis
- Small p-value means surprising data; does not prove H0 false
- Separate from effect size; tiny effects can have tiny p-values

WEAKER EVIDENCE

STRONGER EVIDENCE

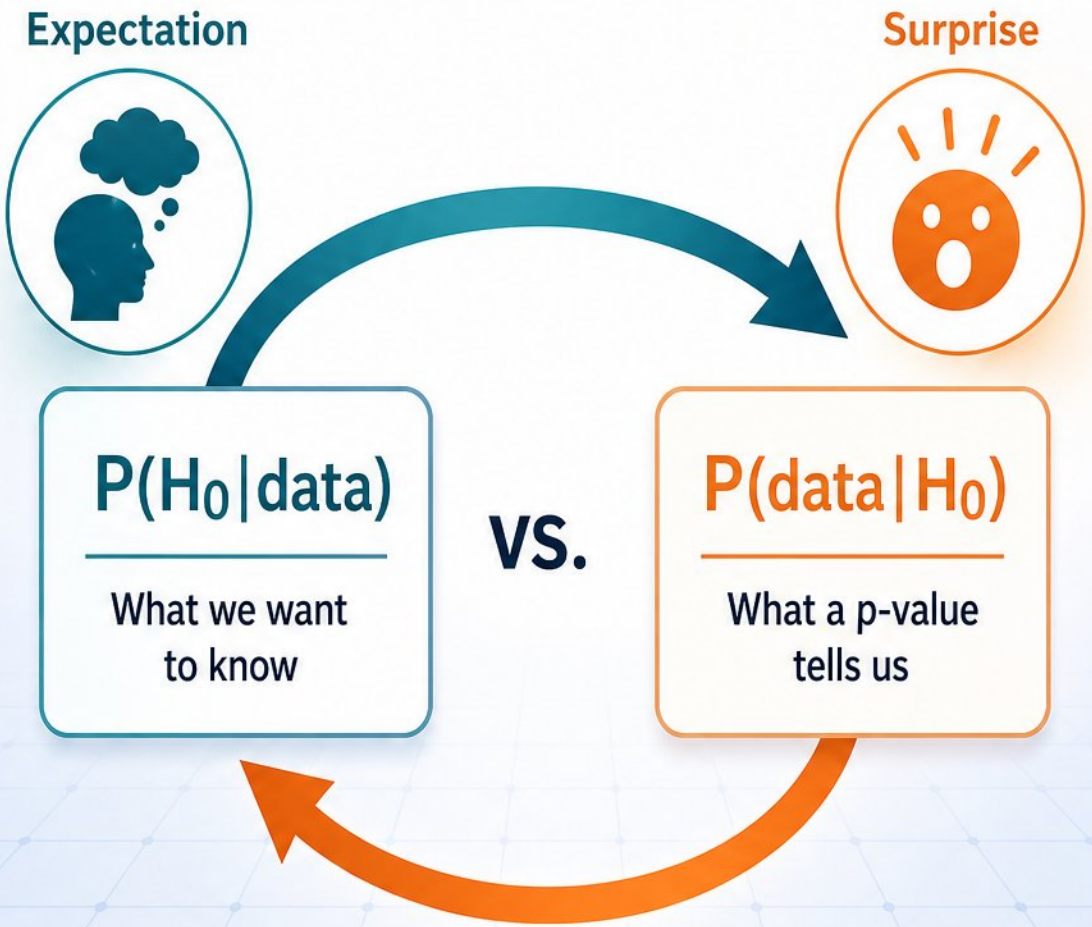
Why Analysts Use P-Values



- Standardized, comparable evidence measure across tests and teams
- Continuous measure of surprise, not a binary yes/no label
- Compares results against pre-specified thresholds like alpha
- One input among several, not the sole basis for decisions
- Supports pre-registered analysis and consistent reporting habits

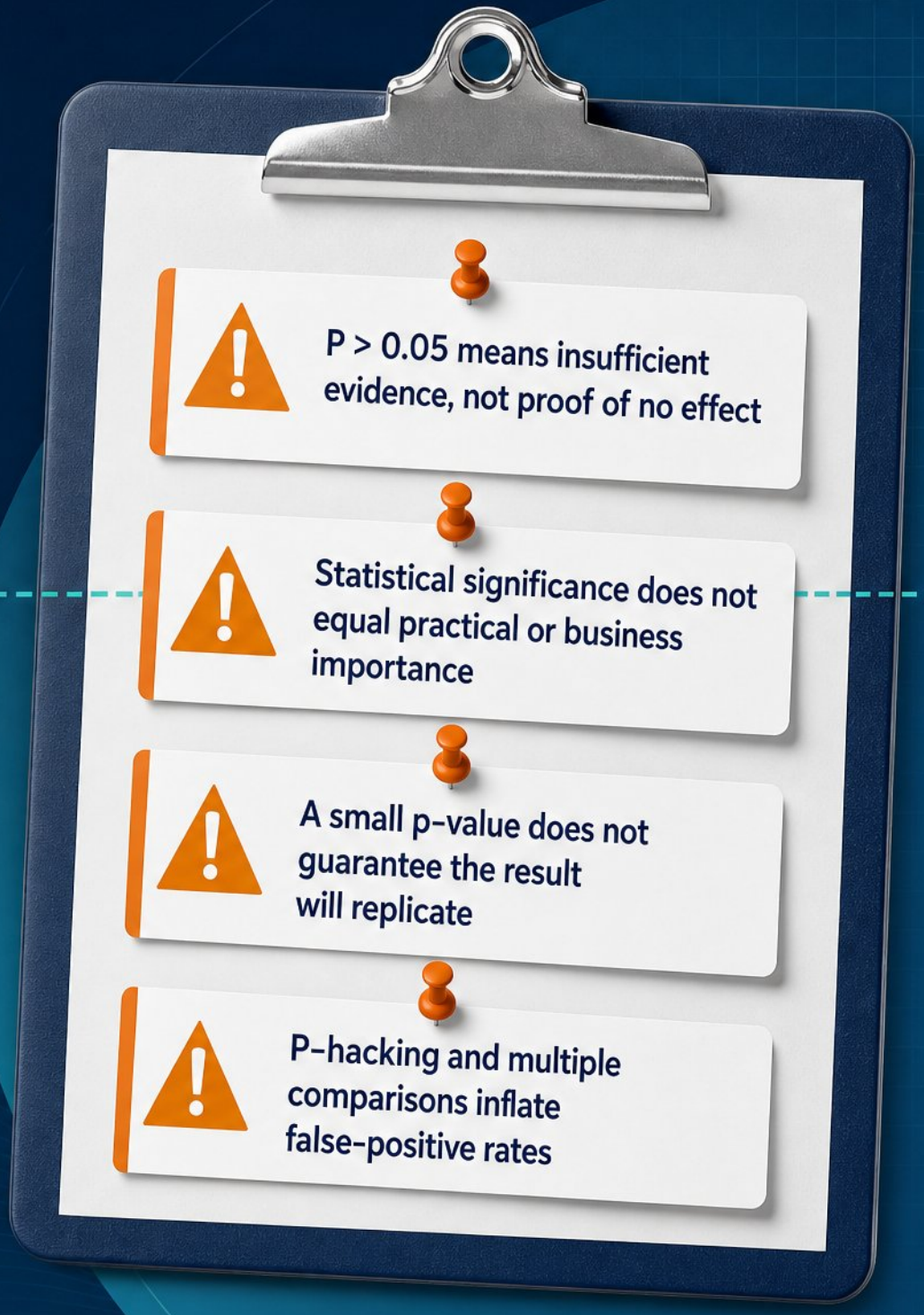
What a P-Value Is Not: The Inverse Probability Fallacy

- A *p*-value is not the probability that the null hypothesis is true.
- $P(\text{data} \mid H_0)$ is not the same as $P(H_0 \mid \text{data})$.
- A small p-value does not prove or make the alternative hypothesis likely.
- Probability of a hypothesis requires a prior probability that *p*-values do not supply.



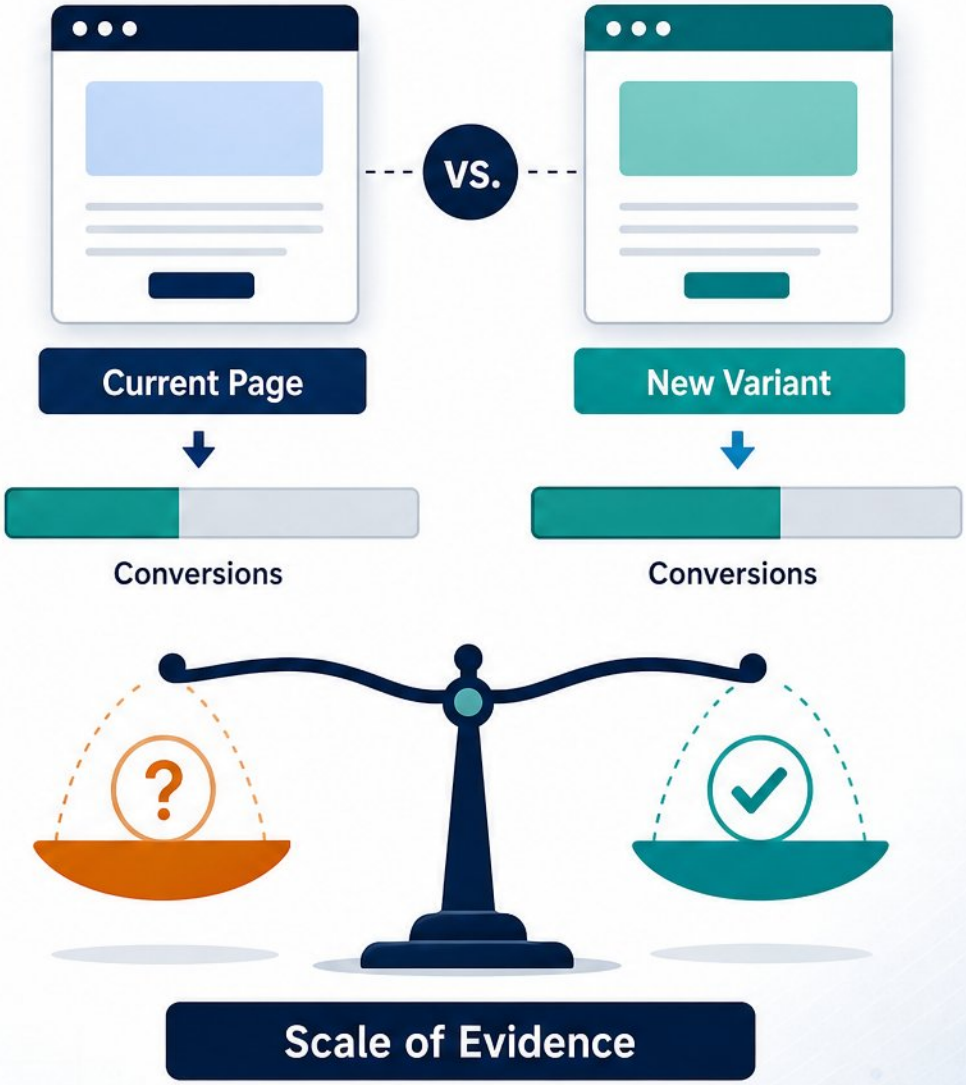
Common Misinterpretations in Practice

- P > 0.05 means insufficient evidence, not proof of no effect
- Statistical significance does not equal practical or business importance
- A small p-value does not guarantee the result will replicate
- P-hacking and multiple comparisons inflate false-positive rates



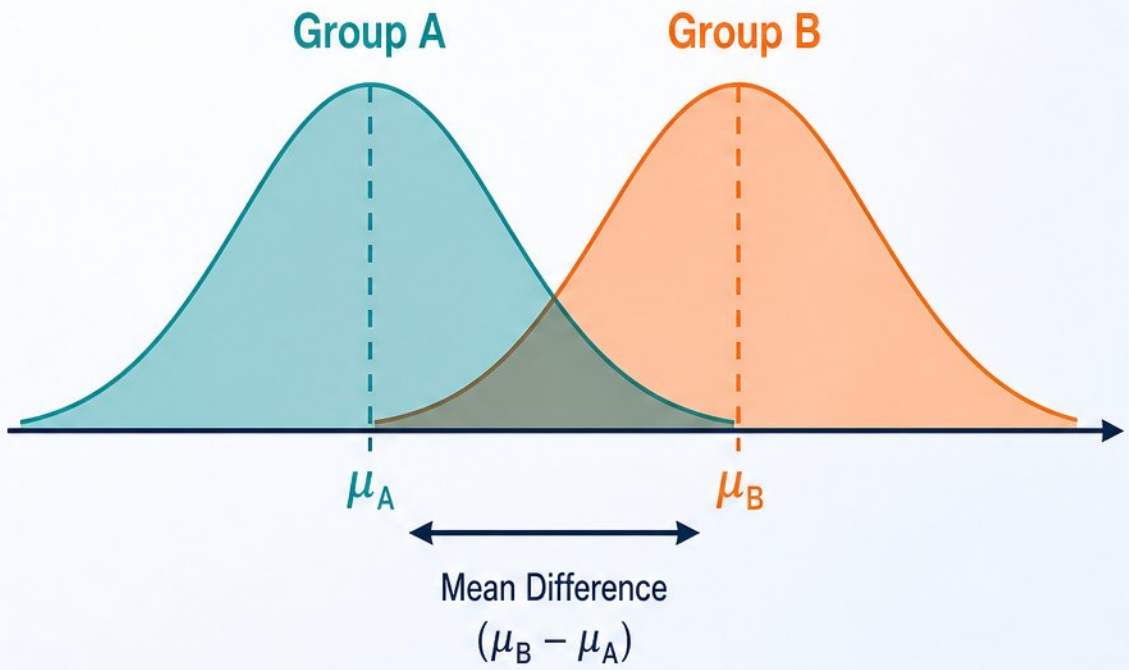
Worked Example: A/B Test Conversion Rate

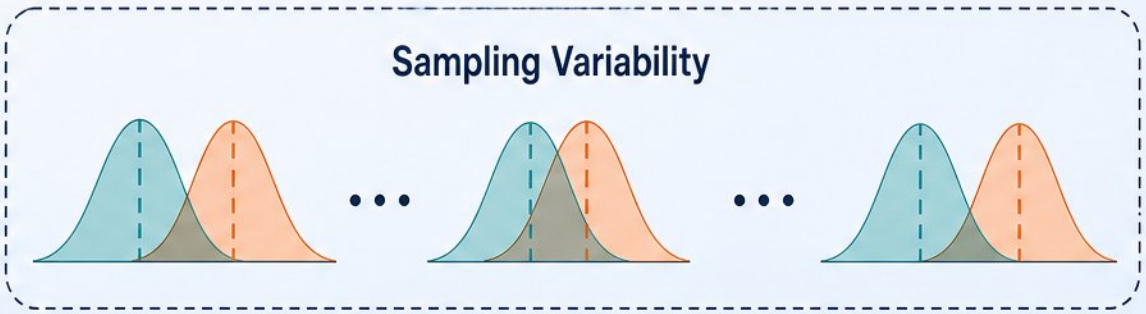
- Compare webpage conversion rates: current page vs. new variant
- Run a two-proportion test to get a p-value
- If p is small, observed lift is surprising under 'no difference' (H0)
- p-value alone never proves the variant is worth shipping
- Decision needs: effect size, confidence interval, practical impact



Worked Example: Comparing Group Means

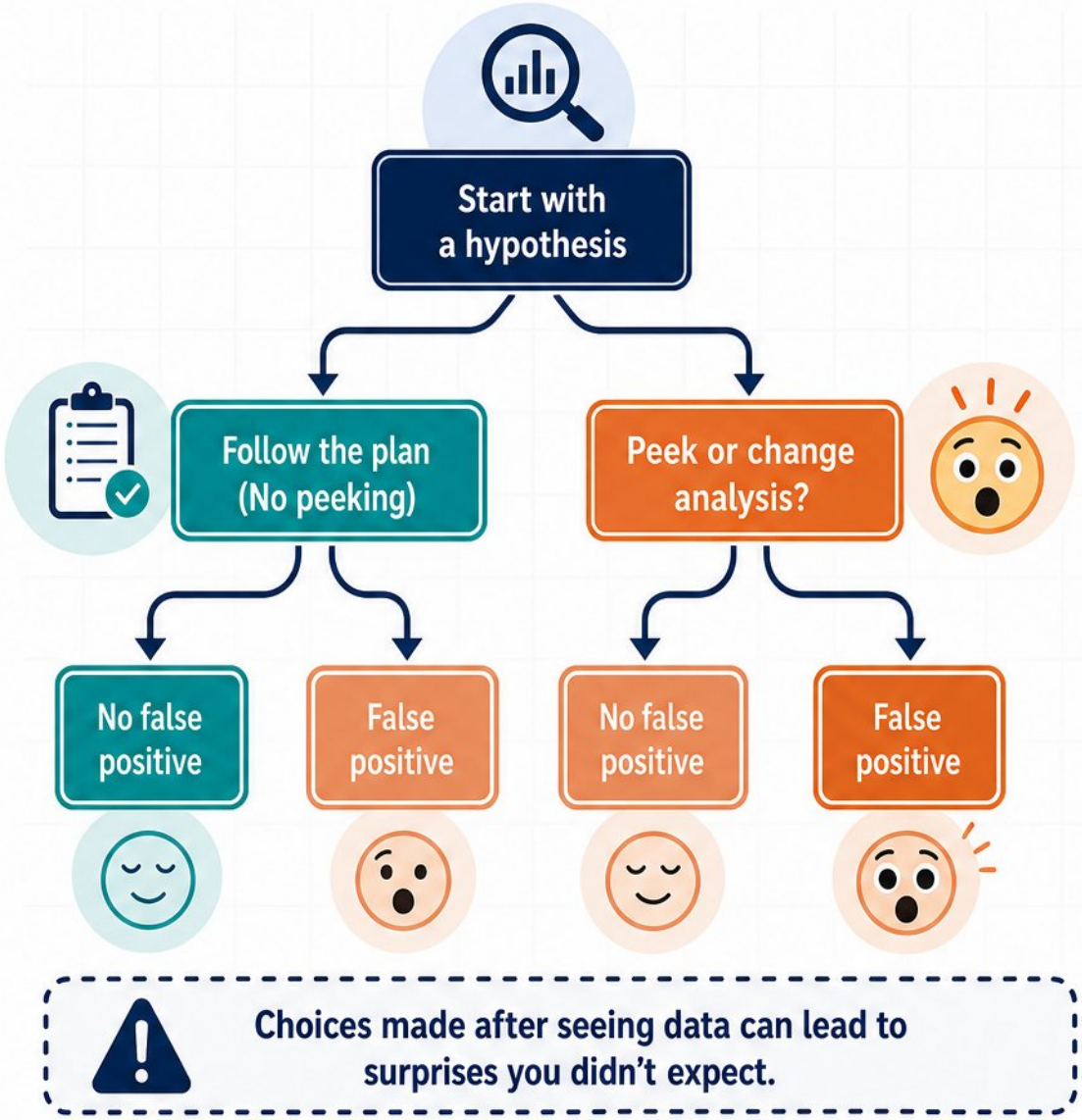
- **Scenario:** compare average sales or completion time between two groups.
- **Read the t-test output:** t, df, p-value, and mean difference.
- **Check assumptions:** independence, approximate normality, and variance.
- **Report the mean difference** and confidence interval, not just the p-value.





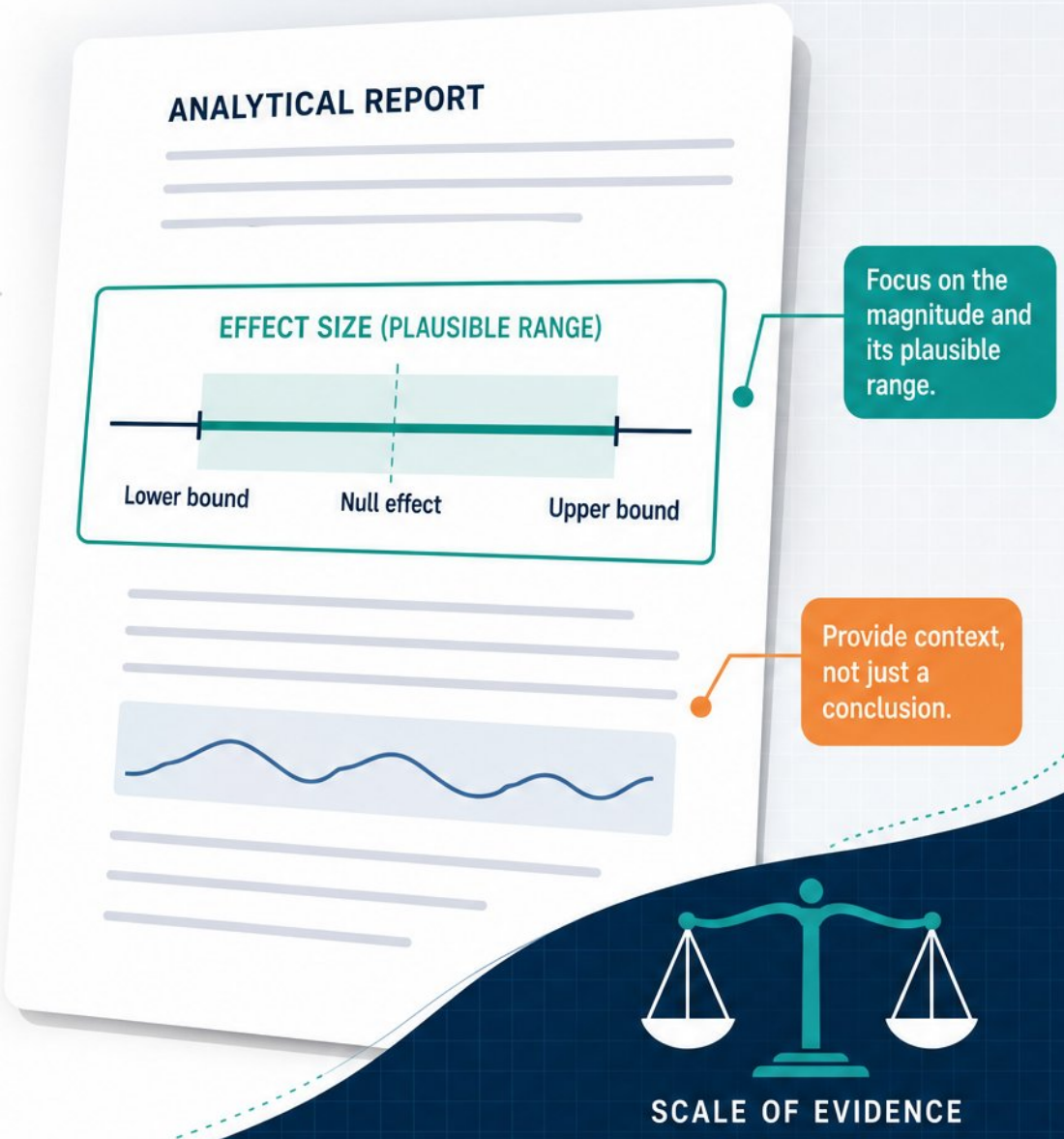
P-Hacking, Peeking, and Multiple Comparisons

- **P-hacking:** changing analysis after seeing data to force $p < 0.05$
- **Peeking** and stopping early inflates the false-positive rate
- **Multiple comparisons** raise the chance of at least one false positive
- **20 tests at alpha 0.05 give a** 64% chance of one false positive
- **Bonferroni** divides alpha; **Benjamini-Hochberg** controls the false discovery rate



Practical Reporting Strategies for Analysts

- Lead with effect size and business impact, not p-values
- Always report confidence intervals to show the plausible range
- Use calibrated language: 'consistent with,' not 'proves'
- Pre-specify hypotheses, metrics, and success criteria before data collection



A Decision Checklist for Reading Any P-Value

- Verify test assumptions and sidedness before interpreting
- Read effect size and confidence interval, not only p
- Check for peeking and multiple comparisons
- Ask if the effect is practically meaningful
- Pre-specify analysis plans to avoid p-hacking

Summary and Next Steps

Correct Interpretations

A small p-value means the data would be surprising if the null were true.
Use p-values with context and other evidence.

Incorrect Interpretations

A small p-value does not prove the hypothesis.
A large p-value does not prove no effect.
Avoid the inverse probability fallacy.

- P-value measures surprise under the null, not hypothesis truth probability
- Avoid inverse probability fallacy and treating non-significance as no effect
- Combine p-values with effect sizes, confidence intervals, and business context
- Explore confidence intervals, power, and Bayesian alternatives

PersonWise.

PersonWise • AI digital-human course platform

Scan the code or click anywhere to watch this course live, with voice Q&A

personwise.ai/t/d0c3e9f4/public