

What a P Value Does and Does Not Mean

A p-value answers a narrow question. If the stated null model were true, how surprising would a result at least as extreme as the one you observed be, under the test procedure you selected?



What a P Value Does and Does Not Mean

Exported from statreason.com. Deterministic, browser-local content; not professional statistical or research advice.

The question a p-value actually answers

That opening sentence has several parts that matter. The p-value depends on the null model: a t-test, a chi-square test, and a correlation test do not ask exactly the same question. It also depends on the alternative direction, the selected test statistic, the sample size, and the assumptions that make the reference distribution useful. A p-value is not a universal score for how real, useful, or repeatable a finding is.

Worked example

Suppose two independent groups differ by 4 points on an outcome. The test returns $p = .03$ with alpha set to .05. The data provide evidence against the particular no-difference model used by the test at that threshold. The result does not tell you that there is a 3 percent chance that no difference exists. It does not tell you the difference is important in practice. It does not tell you the study would obtain the same result if repeated.

The order to read a result in

Read a result in an order that prevents common mistakes. Start with the estimated difference and its direction. Next read the confidence interval to see which values remain compatible with the data under the selected method. Then consider effect size or the raw unit difference in context. The p-value is useful after those pieces because it describes evidence against the stated null model, not the whole decision.

Why the same effect can give different p-values

A larger sample can make a small difference produce a small p-value. A small sample can leave a meaningful difference uncertain and produce a larger p-value. That is why the same observed difference can lead to different p-values in different studies. Precision matters along with magnitude.

What a result above alpha means

When a p-value is above alpha, say that the analysis did not provide sufficient evidence against the stated null model at the selected threshold. Do not say that no effect exists. The confidence interval may be narrow around a practically trivial value, or it may be wide enough to include several important values. Those are very different situations even when both have p-values above .05. See [What a Nonsignificant Result Means](#) for a fuller treatment of this specific case.

What the Assumption Audit can and cannot tell you

Before acting on any test result, review the calculator's Assumption Audit. It can show sample size, missing values, observed variance, tied ranks, expected counts, or a diagnostic plot. It cannot confirm that observations were independent, that an experiment supports a causal conclusion, or that an outcome measure is valid. Those are study facts that need human review, not something a p-value can settle.

Five things a p-value is never allowed to mean here

A p-value is not the probability the null hypothesis is true. It is not the probability your results happened by chance. It is not the probability you would get the same result again. It is not a measure of effect size or practical importance. And it is not evidence that an effect does not exist simply because it landed above your alpha threshold. Every engine on this site is written to avoid all five of these claims, and this page exists so you can check that discipline yourself.

Source

This page follows the interpretation conventions in the Cochrane Handbook's treatment of p-values and confidence intervals, and the same shared statistical reasoning contract every StatReason engine is built against.

Limitations

This page describes the standard frequentist p-value used throughout this site. It does not cover Bayesian posterior probabilities, which answer a genuinely different question and are not implemented on this site.

Next action: use the relevant confidence interval engine after a t-test, proportion test, correlation, ANOVA, or regression analysis. If you are still planning the study, use the matching [sample size and power engine](#). See also the [P-Value Calculator](#), the [Independent t-Test](#), and the [One-Way ANOVA Calculator](#).