

An Explanation of P-Values and Statistical Significance

Authored by
Mohammed looti

November 9, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *An Explanation of P-Values and Statistical Significance*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=14276>

In the realm of statistics, the concept of **p-values** forms the cornerstone of inferential analysis. These values are routinely employed across virtually all forms of quantitative research, including [t-tests](#), [chi-square tests](#), regression analysis, and [ANOVAs](#).

Despite their ubiquitous presence, **p-values** are frequently misinterpreted, leading researchers and analysts to draw flawed conclusions regarding the results of their studies or experiments. Understanding the true meaning of the p-value is critical for sound [statistical inference](#).

This detailed guide aims to demystify the p-value, explaining its precise definition, its role within [hypothesis testing](#), and providing clear, actionable interpretations to ensure accuracy in data analysis.

The Foundation: Understanding Statistical Hypothesis Testing

To fully grasp the significance of a p-value, we must first establish a firm understanding of **hypothesis testing**, which is the formal statistical framework within which p-values operate. A hypothesis test is a structured procedure used to determine whether there is enough evidence in a sample of data to infer that a certain condition or effect exists in the larger population.

Consider a scenario where a company develops a new procedure or drug. We hypothesize that this new method offers a demonstrable benefit over the current standard. To test this claim rigorously, we construct two opposing statements about the population: the null hypothesis and the alternative hypothesis.

These opposing hypotheses are structured as follows:

Null hypothesis (H_0) - This is the statement of no effect, no difference, or no change. It posits that the new method is statistically identical to the old method, or that the observed difference is purely due to random chance.

Alternative hypothesis (H_a) - This is the statement we are trying to find evidence for. It posits that an effect or difference truly exists between the methods being compared. If we reject the null hypothesis, we accept the alternative hypothesis.

The entire purpose of calculating a **p-value** is to quantify the strength of the evidence against the [null hypothesis](#), based on the observed sample data.

The Precise Definition of the P-Value

The **p-value** is fundamentally a probability. It quantifies how compatible our observed data are with the scenario defined by the null hypothesis. The textbook definition, which requires careful reading, is as follows:

A **p-value** is the probability of observing a sample statistic that is at least as extreme as, or more extreme than, the statistic actually observed, assuming that the null hypothesis is true.

In practical terms, the p-value assumes a world where the null hypothesis reigns supreme (i.e., there is no real effect). It then asks: If this world is true, how likely would it be for us to randomly collect a sample that looks like the one we just obtained? A very small p-value suggests that obtaining such an extreme result is rare under the null hypothesis, making the null hypothesis highly implausible given the [sample data](#).

Conversely, a large p-value suggests that the observed data are quite common and expected if the null hypothesis were true, meaning there is insufficient evidence to challenge the status quo defined by H_0 .

P-Values and the Decision Rule: The Significance Level

The p-value itself does not automatically dictate the outcome; it must be compared against a predefined threshold known as the **significance level**, or [alpha \(\$\alpha\$ \)](#). This threshold represents the maximum acceptable risk of rejecting a true null hypothesis.

Researchers must select this significance level before conducting the test. Common choices for α include 0.05 (5%), 0.01 (1%), or 0.10 (10%). The standard threshold used in most scientific disciplines is 0.05.

The statistical decision rule based on the p-value is straightforward:

If the **p-value** is *less than* the chosen significance level (α), we have sufficient evidence to **reject the null hypothesis**. The result is deemed **statistically significant**.

If the **p-value** is *equal to or greater than* the significance level (α), we **fail to reject the null hypothesis**. The result is not statistically significant at that level.

For example, suppose a factory claims its tires have a mean weight of 200 pounds. An auditor runs a hypothesis test to see if the true mean is different from 200 pounds and finds a p-value of 0.04. If the auditor set the significance level at $\alpha = 0.05$, since $0.04 < 0.05$, the auditor would reject the null hypothesis. The observed sample data are inconsistent with the factory's claim, suggesting the true mean weight is likely not 200 pounds.

Crucial Misconceptions: What a P-Value Is *Not*

One of the most profound and frequent errors in statistical reporting is equating the p-value with the probability of making a mistake. Specifically, analysts often mistakenly believe that the p-value is equivalent to the probability of rejecting a true null hypothesis, which is formally known as a [Type I error](#). This interpretation is incorrect.

There are two primary conceptual reasons why the p-value cannot represent the error rate or the probability that the null hypothesis is true:

1. The P-Value's Conditional Nature: The p-value calculation is entirely predicated on the assumption that the null hypothesis is 100% true. It calculates the probability of the data ($P(\text{Data} | H_0 \text{ is True})$), not the probability of the hypothesis being true ($P(H_0 \text{ is True} | \text{Data})$). Since the calculation starts from the premise that the null is true and that any observed difference is simply due to [random chance](#), the p-value cannot logically tell us the probability that this premise is false.

2. Ambiguity of Low Probability: Although a low p-value indicates that your sample data are unlikely if the null is true, it does not distinguish between two possible scenarios:

The null hypothesis is genuinely false, and the alternative hypothesis is true.

The null hypothesis is true, but you happened to obtain an extremely rare, odd sample due to random sampling variability.

Returning to the tire example (p-value = 0.04), here is the critical distinction in interpretation:

Correct Interpretation: Assuming the factory truly produces tires with a mean weight of 200 pounds, a difference as large as, or larger than, the one observed in your audit would occur in only 4% of repeated audits solely due to random sampling error.

Incorrect Interpretation: If you reject the null hypothesis, there is only a 4% chance that you are making a mistake (a Type I error). (Note: The probability of a Type I error is fixed by your α , usually 5%, before the test is run.)

Practical Illustrations of P-Value Interpretation

The following examples apply the formal definition of the p-value to real-world hypothesis testing scenarios, illustrating the correct phrasing necessary for accurate communication of statistical results.

Example 1: Testing Customer Satisfaction Claims

A major phone company asserts that 90% of its massive customer base is satisfied with their service. To verify this claim, an independent researcher selects a representative random sample of 200 customers. The survey results show that only 85% of the sampled customers reported satisfaction. A hypothesis test is conducted against the company's claim ($H_0: p = 0.90$), yielding a p-value of 0.018. If the significance level is set at $\alpha = 0.05$, this result is statistically significant.

Correct interpretation of p-value (0.018): Assuming that 90% of the company's customers are

truly satisfied with their service (i.e., the null hypothesis is true), a researcher would obtain a sample proportion as low as 85%--or an even more extreme result--in only 1.8% of repeated random samples due to expected [random sampling error](#). Because this probability (1.8%) is very low (less than 5%), we reject the company's claim.

Example 2: Comparing Battery Lifespans

A technology firm introduces a new phone battery, claiming it lasts significantly longer than the old model--specifically, at least 10 minutes longer. A researcher conducts a two-sample t-test comparing the average lifespan of the new batteries (Sample Mean: 120 minutes; SD: 12 minutes) against the old batteries (Sample Mean: 115 minutes; SD: 15 minutes), using 80 units of each type. The test aims to see if the new battery is truly superior. The test results in a p-value of 0.011.

Correct interpretation of p-value (0.011): Assuming that the new battery performs the same as, or worse than, the old battery (which is the null hypothesis of no benefit), the observed difference in mean lifespan (5 minutes) or a more extreme difference in favor of the new battery would occur in only 1.1% of similar comparative studies due to inherent random variation. Since this probability is exceptionally low (and assuming $\alpha = 0.05$), we conclude that the new battery does indeed offer a statistically significant improvement.