How to Interpret a P-Value Greater Than 0.05

Authored by
stats writer

February 26, 2026

Interpret a P-Value Greater Than 0.05 (With Examples)

When conducting quantitative research, the **p-value** serves as a fundamental metric used to determine the strength of the evidence against a specific claim. Specifically, it is utilized to evaluate whether a **hypothesis test** regarding a population parameter yields results that are sufficiently unusual to warrant the rejection of a baseline assumption. By calculating the **probability** of obtaining test results at least as extreme as the results actually observed, researchers can make informed decisions about the validity of their experimental findings.

Whenever we perform a **hypothesis test**, we always define a **null hypothesis** and an **alternative hypothesis** to create a clear framework for our analysis. These two statements are mutually exclusive and collectively exhaustive, ensuring that the outcome of the test leads to a definitive statistical conclusion. The objective of the test is rarely to prove the **null hypothesis** is true, but rather to determine if the data provides enough evidence to suggest that the alternative is more plausible given the observed patterns.

Null Hypothesis (H0): This assumption posits that the sample data occurs purely from chance and that no real effect or relationship exists between the variables. It represents the status quo or the default position that there is no change or no difference in the population being studied.
Alternative Hypothesis (HA): This proposition suggests that the sample data is influenced by some non-random cause and that a significant effect or relationship does indeed exist. It is the statement that the researcher typically hopes to support through the collection and analysis of empirical evidence.

The Foundations of Statistical Significance and Alpha Levels

When performing a **hypothesis test**, we must specify the **significance level** to use as a threshold for decision-making. This value, often denoted by the Greek letter alpha (α), represents the maximum risk we are willing to take of committing a **Type I error**, which occurs when we incorrectly reject a true **null hypothesis**. By setting this threshold before data collection begins, researchers maintain the integrity of the **statistical significance** criteria and avoid the temptation to manipulate results post-hoc.

Common choices for a **significance level** include standard values that reflect different levels of stringency depending on the field of study. In many social sciences and general research contexts, an alpha of .05 is the conventional standard, implying a 5% risk of a false positive. In more critical fields like medicine or engineering, a more conservative alpha of .01 might be used, while exploratory studies might occasionally use a more relaxed alpha of .10 to identify potential trends for further investigation.

α = .01 (High stringency)
α = .05 (Standard stringency)
α = .10 (Low stringency)

If the **p-value** of the **hypothesis test** is less than the specified **significance level**, then we can reject the **null hypothesis** and conclude that we have sufficient evidence to say that the **alternative hypothesis** is true. This outcome suggests that the observed data is highly unlikely to have occurred under the assumption of the null, leading us to believe that a meaningful relationship or effect has been detected within the **sample size** provided.

Conversely, if the **p-value** is not less than the specified **significance level**, then we fail to reject the **null hypothesis** and conclude that we do not have sufficient evidence to say that the **alternative hypothesis** is true. It is critical to note that "failing to reject" is not the same as "accepting" the null; it simply means the evidence is not strong enough to move away from the initial assumption. The following examples explain how to interpret a **p-value** greater than .05 in practice across various professional domains.

The Meaning of a P-Value Greater Than 0.05

When a **p-value** is greater than 0.05, it means that there is a high **probability** that the observed data is due to chance rather than a significant relationship or effect. This indicates that the variations seen in the data are consistent with the random fluctuations one would expect if the **null hypothesis** were true. Consequently, the results are categorized as not reaching **statistical significance**, meaning they cannot be confidently attributed to the independent variables being tested in the experiment.

This result typically indicates that there is no significant difference between the compared groups or conditions within the constraints of the study. For instance, if a researcher is comparing the performance of two different teaching methods and finds a **p-value** of 0.12, they must conclude that any observed difference in test scores could easily be the result of **sampling error**. Without a **p-value** below the 0.05 threshold, the evidence is simply too weak to justify a change in theory or practice.

Consider a scenario where a study measures the effectiveness of a new medication on patients with a certain illness. If the resulting **p-value** is greater than 0.05, it suggests that the medication does not have a significant impact on the patients' condition compared to a **placebo** or existing treatment. Another example could be a survey comparing the opinions of two political parties, where a **p-value** greater than 0.05 would indicate that there is no significant difference in the opinions of the two parties regarding a specific policy issue.

Example 1: Interpret P-Value Greater Than 0.05 (Biology)

Suppose a biologist believes that a certain fertilizer will cause plants to grow more during a one-year period than they normally do, which is currently an average of 20 inches. This belief stems from the chemical composition of the fertilizer, which is designed to enhance nitrogen uptake in the root systems. To rigorously test this hypothesis, she applies the fertilizer to each of the plants in her laboratory for a controlled three-month observation period, ensuring that all other environmental factors like light and water remain constant.

She then performs a **hypothesis test** using the following hypotheses to structure her statistical analysis:

The null hypothesis (H0): μ = 20 inches (the fertilizer will have no effect on the mean plant growth, and any deviation is due to random chance).

The alternative hypothesis (HA): μ > 20 inches (the fertilizer will cause mean plant growth to increase significantly beyond the historical average).

Upon conducting a **hypothesis test** for a mean using a **significance level** of α = .05, the biologist receives a **p-value** of **0.2338**. This value is derived from the test statistic and represents the likelihood of seeing these growth results if the fertilizer actually did nothing at all to the plants.

Since the **p-value** of **0.2338** is significantly greater than the **significance level** of **0.05**, the biologist fails to reject the **null hypothesis**. This means she cannot conclude that the fertilizer is effective. The data suggests that the observed growth, even if slightly above 20 inches, is not statistically different from the normal growth rate, likely requiring a larger **sample size** or a different formulation to prove efficacy.

Example 2: Interpret P-Value Greater Than 0.05 (Manufacturing)

A mechanical engineer believes that a new production process will reduce the number of faulty widgets produced at a certain factory, which is currently 3 faulty widgets per batch on average. Improving quality control is essential for reducing waste and increasing the factory's **profit margins**. To test this theory, he implements the new process on a trial basis to produce a new series of widget batches and carefully records the defect rates for each.

He then performs a **hypothesis test** using the following hypotheses to evaluate the process change:

The null hypothesis (H0): μ = 3 (the new process will have no effect on the mean number of faulty widgets per batch).

The alternative hypothesis (HA): $\mu < 3$ (the new process will cause a significant reduction in the mean number of faulty widgets per batch).

The engineer performs a **hypothesis test** for a mean using a **significance level** of $\alpha = .05$ and receives a calculated **p-value** of **0.134**. This result indicates that if the new process were identical in quality to the old one, there would still be a 13.4% chance of seeing the defect reduction observed in the sample data.

Since the **p-value** of **0.134** is greater than the **significance level** of **0.05**, the engineer fails to reject the **null hypothesis**. Thus, he concludes that there is not sufficient evidence to say that the new process leads to a reduction in the mean number of faulty widgets produced in each batch. He might decide to refine the process further or investigate other variables that could be contributing to the defect rate.

Nuances and Misconceptions of Non-Significant Results

A common misconception in research is that a **p-value** greater than 0.05 proves that the **null hypothesis** is true. In reality, a non-significant result only indicates that the study did not find enough evidence to disprove the null. This could be due to a variety of factors, such as a **sample size** that was too small to detect a subtle but real effect, or excessive noise in the data that obscured the signal of the relationship being studied.

Statisticians often refer to this as a lack of **statistical power**. If a test has low power, it is more likely to result in a **Type II error**, which is the failure to reject a false **null hypothesis**. Therefore, when a researcher encounters a **p-value** of 0.06 or 0.08, they may describe it as "trending toward significance," acknowledging that while the threshold was not met, the data may still warrant a follow-up study with more rigorous controls or a larger participant pool.

Furthermore, it is important to distinguish between **statistical significance** and practical significance. A result can be statistically significant ($p < 0.05$) but have such a small **effect size** that it is useless in a real-world application. Conversely, a study with a **p-value** slightly above 0.05 might still point toward a practically important effect that deserves further investigation in a more focused experimental design.

Summary of Best Practices for Reporting P-Values

When documenting the results of a **hypothesis test**, it is best practice to report the exact **p-value** rather than simply stating that it was greater than or less than 0.05. Providing the exact value allows other researchers to understand the margin by which the **null hypothesis** was or was not rejected. This transparency is crucial for the **reproducibility** of scientific findings and allows for meta-analyses where multiple studies are combined to find a more definitive conclusion.

In addition to the **p-value**, researchers should always include descriptive statistics such as the mean, **standard deviation**, and **confidence intervals**. These metrics provide a more complete picture of the data distribution and the precision of the estimates. A **confidence interval** that crosses the value of "no effect" (such as zero for a difference between means) is the mathematical equivalent of a **p-value** greater than 0.05.

Ultimately, the interpretation of a **p-value** is just one piece of the puzzle in data analysis. Effective decision-making requires a holistic view of the experimental design, the quality of the data, the **sample size**, and the theoretical context of the research. By understanding that a **p-value** greater than 0.05 simply denotes a failure to reach a specific evidentiary threshold, analysts can avoid overstating their findings and maintain the rigor required for scientific progress.

The following tutorials provide additional information about p-values and their application in various statistical tests: