Two-Tailed Hypothesis Tests: 3 Example Problems

Authored by

Mohammed looti

October 29, 2025

In the field of statistics, we employ **hypothesis testing** to rigorously evaluate whether a specific claim regarding a **population parameter** holds true. This formalized procedure provides a structured method for making decisions about populations based on sample data.

Whenever a statistical hypothesis test is conducted, two contrasting statements are always formulated: the **Null Hypothesis** and the **Alternative Hypothesis**. These foundational statements define the scope of the investigation.

These hypotheses typically follow these general formats, where the population parameter (often denoted as µ for the mean) is compared against a hypothesized value:

H0 (Null Hypothesis): Population parameter = ≤, or ≥ some specified value. This represents the status quo or the condition of no effect.

HA (Alternative Hypothesis): Population parameter <, >, or ? some specified value. This represents the claim the researcher seeks to find evidence for.

Hypothesis tests are categorized into two main types based on the directionality specified in the Alternative Hypothesis:

One-tailed test: The alternative hypothesis is directional, containing either the **<** (less than) or **>** (greater than) sign. This focuses the rejection region entirely on one side of the distribution.
Two-tailed test: The alternative hypothesis is non-directional, containing the **?** (not equal) sign. This splits the rejection region into two tails of the distribution.

Understanding Hypothesis Testing and its Types

The selection between a one-tailed and a two-tailed test is one of the most critical preliminary steps in statistical analysis. This decision must be made before collecting data and is entirely dependent on the research question and the existing theoretical knowledge. A **two-tailed test** is appropriate whenever we are interested in detecting a difference or an effect, regardless of whether that effect is positive or negative relative to the hypothesized population mean.

In a **two-tailed test**, the alternative hypothesis strictly specifies that the population parameter is simply "not equal" (**?**) to the null value. This framework is used when the researcher has no prior theoretical basis or strong empirical evidence to predict the direction of the change. It is fundamentally a test of discrepancy; we are testing whether the observed sample statistic is far enough away from the null value in either direction to warrant rejecting the status quo.

Because the rejection region is split between the two extremes of the sampling distribution, the two-tailed test is considered a more conservative approach than a one-tailed test. To reject the null hypothesis, the calculated test statistic must fall into either the upper or the lower critical region, meaning the observed data must show a substantial deviation from the expected mean, either positively or negatively.

The Mechanics of the Two-Tailed Test

The core mechanism of a two-tailed test involves comparing the observed sample data against the null hypothesis using a test statistic (such as a t-statistic or z-statistic) and assessing the probability of observing such data if the null hypothesis were true. This probability is known as the **p-value**. In the two-tailed scenario, the calculated p-value represents the probability mass distributed across both tails of the sampling distribution.

The decision criterion relies on a pre-determined significance level, usually denoted as alpha (α), which is commonly set at 0.05. For a two-tailed test with $\alpha = 0.05$, the critical region is split, meaning 0.025 (or 2.5%) of the distribution area is allocated to the upper tail, and 0.025% is allocated to the lower tail. If the calculated p-value is less than the total alpha (e.g., p-value < 0.05), we reject the null hypothesis, concluding there is significant evidence of a difference.

Conversely, if the p-value is greater than alpha, we fail to reject the null hypothesis, indicating that the observed difference between the sample mean and the hypothesized mean is likely due to random sampling variability rather than a genuine effect. Understanding these mechanics is crucial for accurately interpreting the results of the following practical examples, which illustrate how non-directional hypotheses are tested in real-world scenarios.

Example 1: Assessing Widget Weight Deviations (Factory Widgets)

Consider a factory where the long-established standard dictates that the average weight of a particular widget is 20 grams. A factory engineer introduces a new manufacturing technique, but he is uncertain whether this new method will increase or decrease the average weight. His primary concern is simply whether the average weight is now different from the established 20 grams.

To properly test this non-directional belief, the engineer must perform a two-tailed hypothesis test. The statistical hypotheses are formally defined as follows:

H0 (Null Hypothesis): μ = 20 grams (The new method has no effect on the mean weight.)
HA (Alternative Hypothesis): μ ? 20 grams (The new method results in a mean weight different from 20 grams.)

This setup confirms it is a **two-tailed hypothesis test** because the alternative hypothesis uses the non-directional "? " sign. The engineer is looking for any significant deviation from the mean, regardless of whether that deviation represents an increase or a decrease in widget weight.

The engineer collects a sample using the new method and gathers the following crucial sample statistics:

n = **20** widgets (Sample size)
x = **19.8** grams (Sample mean weight)
s = **3.1** grams (Sample standard deviation)

Using these sample statistics and assuming an alpha level (α) of 0.05, the data is plugged into the appropriate t-test formula (since the population standard deviation is unknown). This calculation yields the following statistical results:

t-test statistic: **-0.288525**
two-tailed **p-value**: **0.776**

Since the resulting p-value (0.776) is substantially greater than the significance level of 0.05, the engineer must fail to reject the **Null Hypothesis**. The small difference observed in the sample mean (19.8 grams) compared to the hypothesized mean (20 grams) is not statistically significant. Consequently, the engineer lacks sufficient evidence to conclude that the true mean weight of widgets produced by the new method is statistically different from 20 grams. The new method appears to have had no measurable effect on the average weight.

Example 2: Analyzing Novel Fertilizer Impact (Plant Growth)

A second scenario involves a botanist studying plant growth. A standard fertilizer has historically resulted in an average growth of 10 inches over a certain period for a specific plant species. The botanist introduces a novel fertilizer, believing it will influence growth, but is unsure if it will stimulate more growth or inhibit it. Therefore, the research question is simply: does the new fertilizer cause growth different from 10 inches?

This scenario mandates a two-tailed test, structured with the following competing hypotheses:

H0 (Null Hypothesis): μ = 10 inches (The new fertilizer yields the same average growth as the standard.)
HA (Alternative Hypothesis): μ ? 10 inches (The new fertilizer yields an average growth different from 10 inches.)

The presence of the non-directional "? " sign in the Alternative Hypothesis confirms this setup as a **two-tailed hypothesis test**. The botanist is prepared to accept an effect if the mean growth is significantly higher or significantly lower than the established 10 inches.

To test her claim, the botanist applies the new fertilizer to a simple random sample of 15 plants and records the growth data:

n = **15** plants
x = **11.4** inches
s = **2.5** inches

Using a standard significance level of α = 0.05, these values are used to calculate the **t-test statistic** for the sample mean difference:

t-test statistic: **2.1689**
two-tailed p-value: **0.0478**

In this case, the resulting p-value (0.0478) is less than the predetermined **significance level** of 0.05. This result falls into the rejection region of the distribution, leading the botanist to reject the **Null Hypothesis**.

She has found sufficient statistical evidence to conclude that the new fertilizer causes an average growth that is significantly different from 10 inches. Although the sample mean was 11.4 inches, the two-tailed test only confirms a difference exists, leaving the implication that the new fertilizer appears effective, whether that effect is positive or negative (though in this case, the sample mean suggests a positive effect).

Example 3: Evaluating a New Study Technique (Exam Scores)

A university professor hypothesizes that introducing a new studying technique will alter the mean score on a difficult standardized exam. Currently, the historical mean score is 82. The professor is interested in whether the technique influences the score in any way--either improving or hurting student performance--but has no strong basis to predict the direction of change.

To conduct a fair assessment of this intervention, the professor sets up a two-tailed test focusing on non-directional change:

H0: μ = 82 (The new technique does not change the mean score.)
HA: μ ? 82 (The new technique results in a mean score different from 82.)

As in the previous examples, this is confirmed as a **two-tailed hypothesis test** because the alternative hypothesis uses the "? " sign. The scope of the investigation covers potential effects on both the higher and lower ends of the scoring distribution.

The professor implements the technique with a sample of students and collects the following descriptive statistics:

n = **25** students
x = **85** (Sample mean score)

s = **4.1** (Sample standard deviation)

Assuming an alpha level of 0.05, the professor calculates the necessary statistics to determine if the sample mean of 85 is significantly different from the population mean of 82:

t-test statistic: **3.6586**
two-tailed p-value: **0.0012**

Because the resulting p-value (0.0012) is much less than the standard **significance level** of 0.05, the professor has compelling evidence to reject the **Null Hypothesis**.

The statistical conclusion is that the new studying method produces exam scores with an average score that is statistically different than 82. Given the high sample mean (85) and the extremely low p-value, the professor can confidently report that the new technique has a significant, positive influence on student performance.

Conclusion and Further Study

Two-tailed hypothesis tests are indispensable tools in statistical inference when researchers are investigating whether an intervention or change has occurred, without a preconceived direction for that change. By requiring the statistical evidence to be strong enough to fall into either the extreme positive or extreme negative regions of the distribution, the two-tailed approach provides a robust and unbiased method for testing the existence of an effect. The examples above demonstrate that while the framework remains the same, the ultimate conclusion--rejecting or failing to reject the null hypothesis--depends entirely on the relationship between the calculated p-value and the chosen alpha level.

The following resources offer opportunities for additional exploration into the fundamental concepts of hypothesis testing and statistical decision-making:

Additional Resources

The following tutorials provide additional information about hypothesis testing: