# What is a Confidence Interval for the Difference in Proportions?

Authored by

**stats writer**

December 27, 2025

A confidence interval for the difference in proportions is a highly valuable statistical tool, defining a range of values that is statistically likely to contain the true difference between two population measures with a specified level of confidence. This method is fundamental for quantifying the inherent uncertainty present when estimating differences based on sampled data. The calculation involves determining the difference between the two sample proportions and then adding and subtracting the margin of error. The exact size of this margin is carefully dictated by three critical factors: the chosen sample size, the required confidence level, and the calculated standard error of the difference between the two proportions.

A **confidence interval (C.I.) for a difference in proportions** is a calculated range of values designed to estimate the true discrepancy between two underlying population proportions, ensuring this estimate is framed within a specific, measurable degree of confidence.

This comprehensive tutorial will guide you through the intricacies of this statistical method, detailing:

The essential motivation and rationale for constructing this specific confidence interval.
A breakdown of the precise mathematical formula required for its calculation.
A practical, step-by-step numerical example demonstrating the computation process.
Clear instructions on how to accurately interpret the resulting interval and draw meaningful conclusions.

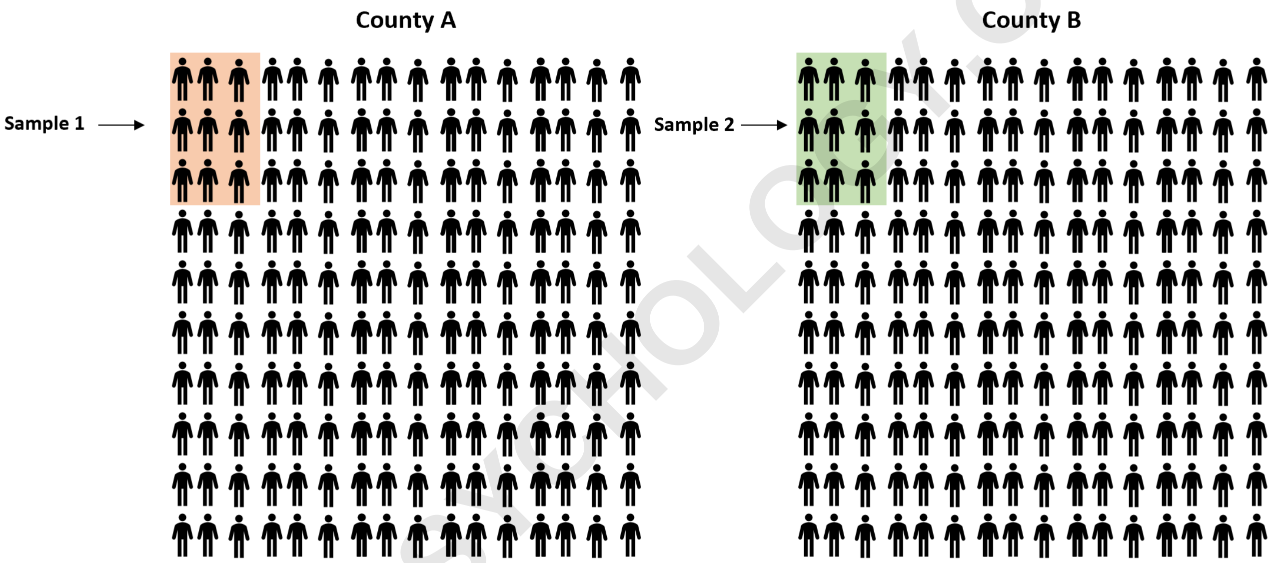
## Motivation for Constructing the Confidence Interval

In practical research and statistical analysis, practitioners are frequently tasked with the challenge of estimating the magnitude of the difference separating two specific population proportions. Whether comparing success rates, approval ratings, or incidence levels across two distinct groups, this estimation is crucial. To begin this estimation, researchers must first conduct a rigorous data collection process: they select an independent random sample from each target population and subsequently calculate the empirical proportion (the sample proportion) for each group. These two sample proportions are then directly compared to obtain a point estimate of the difference.

While the difference found between the two sample proportions serves as our best point estimate, it is improbable that this observed difference exactly matches the true, underlying difference between the unknown population parameters. This discrepancy arises because sampling inherently introduces variability and uncertainty. Consequently, relying solely on the point estimate (the difference between the sample proportions) fails to adequately capture this statistical uncertainty. This is precisely why statisticians construct a confidence interval for the difference between two proportions, providing a robust range of values that is highly likely to encompass the true difference between the two population proportions.

Consider a scenario where the goal is to assess the difference in the proportion of residents who express support for a new municipal law in County A compared to those in County B. Due to the vast number of residents potentially residing in each county, conducting a full census--surveying every single individual--would be prohibitively expensive, time-consuming, and resource-intensive.

Instead of attempting a census, the preferred methodological approach involves drawing a distinct random sample of residents from County A and an equivalent random sample from County B. The observed proportions favoring the law within these two samples (the sample proportions) are then used to generate a reliable estimate of the true difference between the two underlying population proportions.




Because the data collection relies on random samples, the calculated difference between the two sample proportions is subject to sampling error and is therefore not guaranteed to perfectly align with the true difference in the population parameters. To effectively account for and quantify this inherent uncertainty and potential sampling error, constructing a confidence interval becomes essential. This interval provides a statistically bounded range that, with a predefined level of assurance, contains the actual difference between the proportions of the two populations.

## The Formula for Calculating the Confidence Interval

The calculation of a confidence interval for the difference between two population proportions relies on a standardized, two-sample formula. This formula incorporates the observed sample proportions, the sample sizes, and a critical value derived from the chosen confidence level. Understanding each component is vital for accurate application and interpretation.

**Confidence interval = (p1-p2) +/- z*√(p1(1-p1)/n1 + p2(1-p2)/n2)**

The terms used in this formula represent critical statistical inputs necessary for determining both the center point and the width of the interval:

p1, p2: These denote the respective sample proportions observed in Sample 1 and Sample 2.
z: This represents the z-critical value, which is determined solely by the required level of confidence (e.g., 90%, 95%, or 99%).
n1, n2: These are the sample size for Sample 1 and Sample 2, respectively.

The appropriate z-critical value is intrinsically tied to the desired confidence level chosen for the analysis. Standard practice involves using common confidence levels such as 90%, 95%, or 99%. The following reference table illustrates the corresponding z-values for these widely accepted confidence level choices:

| Confidence Level | z-value |
|---|---|
| 0.90 | 1.645 |
| 0.95 | 1.96 |
| 0.99 | 2.58 |

It is important to observe the relationship between the confidence level and the resulting interval width. Higher confidence levels necessitate larger z-critical values. Since the z-value is a direct multiplier for the standard error in the calculation of the margin of error, an increase in confidence leads directly to a wider confidence interval. For instance, a 99% confidence interval will always be broader than a 90% confidence interval when calculated using the identical sample data, reflecting a greater assurance that the true population difference is captured.

## Step-by-Step Example Calculation

To demonstrate the practical application of the formula, let us revisit the scenario of estimating the difference in support for a specific law between residents of County A and County B. We will utilize hypothetical, yet realistic, sample data collected from each county to perform the calculation.

The following summarizes the key statistics derived from the independent random samples:

**Data for County A (Sample 1):**

n1 = 100 (The sample size of residents surveyed in County A)
p1 = 0.62 (This indicates that 62 out of 100 residents surveyed in County A support the proposed law, yielding the sample proportion).

**Data for County B (Sample 2):**

n2 = 100 (The sample size of residents surveyed in County B)
p2 = 0.46 (This indicates that 46 out of 100 residents surveyed in County B support the proposed law, yielding the second sample proportion).

Based on these inputs, we can calculate the point estimate for the difference in support: p1 - p2 = 0.62 - 0.46 = 0.16. Now, we proceed to calculate the confidence intervals for various common confidence levels.

**Calculating the 90% Confidence Interval:**

Using the z-critical value of 1.645:

(.62-.46) +/- 1.645*√(.62(1-.62)/100 + .46(1-.46)/100) =

**Calculating the 95% Confidence Interval:**

Using the z-critical value of 1.96:

(.62-.46) +/- 1.96*√(.62(1-.62)/100 + .46(1-.46)/100) =

**Calculating the 99% Confidence Interval:**

Using the z-critical value of 2.58:

(.62-.46) +/- 2.58*√(.62(1-.62)/100 + .46(1-.46)/100) =

***Note:*** *You can also find these confidence intervals by using the appropriate statistical software.*

## Interpreting the Difference in Proportions Confidence Interval

Once the confidence interval has been successfully calculated, the final and most crucial step is the correct interpretation of the resulting range. The interpretation must clearly connect the statistical result back to the real-world context of the two populations being compared. We will focus on interpreting the 95% confidence interval derived in the previous example: .

The formal interpretation of this specific 95% confidence interval is articulated as follows:

We are 95% confident that the true difference in the proportion of residents who favor the law between County A and County B lies somewhere between 0.0236 (2.36%) and 0.2964 (29.64%).

The presence or absence of the value zero within the interval is the key factor in determining statistical significance when comparing two proportions. If the entire interval is either positive (both lower and upper bounds are greater than zero) or entirely negative (both bounds are less than zero), this strongly suggests that a true, non-zero difference exists between the two population proportions.

In the case of the 95% C.I. , since both the lower and upper bounds are positive (i.e., the interval does not contain the value "0"), we can conclude with high statistical certainty (95% confidence) that there is a genuine and measurable difference in the proportion of residents who support the law between County A and County B. Specifically, County A has a higher proportion of supporters than County B. Conversely, if an interval (like the 99% C.I. of ) includes zero, it suggests that the true difference might be zero, meaning we lack sufficient evidence to claim a significant difference between the two populations at that confidence level.