

A Step-by-Step Guide to the Two-Proportion Z-Test in SAS

Authored by
Mohammed loot

November 14, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *A Step-by-Step Guide to the Two-Proportion Z-Test in SAS*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=1560>

In the advanced realm of [statistical inference](#), researchers constantly face the necessity of comparing characteristics across different populations or experimental groups. A particularly common and vital analytical challenge is determining whether the rates, or [population proportions](#), of a specific outcome genuinely differ between two independent groups. To address this need rigorously, the [two proportion z-test](#) stands as an indispensable analytic tool. This test provides a formal, quantitative framework for assessing whether an observed disparity in sample data is merely the result of expected random chance or if it truly signifies a fundamental, underlying structural distinction between the populations being studied. This comprehensive article will guide you through the essential theoretical foundations, practical assumptions, and step-by-step execution of this powerful statistical procedure using the industry-leading software suite, [SAS](#).

Theoretical Foundation and Key Assumptions

The **two proportion z-test** is a specialized methodology within [hypothesis testing](#) used exclusively to evaluate if a statistically significant difference exists between two unknown, independent population proportions, conventionally denoted as π_1 and π_2 . This test is foundational in comparative research designs, offering robust methods for applications such as evaluating the differential success rates of two distinct marketing campaigns, comparing the efficacy of a new medical treatment against a placebo, or benchmarking quality control metrics across separate manufacturing lines. The core objective is to compare the observed sample proportions (p_1 and p_2) and quantify the probability of obtaining such an observed difference, assuming that the true [population proportions](#) were, in reality, identical.

For the conclusions derived from the two proportion z-test to be statistically valid and reliable, several critical statistical prerequisites must be satisfied. First and foremost, the process of data collection requires that the samples drawn from both populations must be selected randomly, ensuring they are representative of their respective populations. Furthermore, these two samples must be absolutely **independent** of one another; the selection criteria or outcomes for subjects in the first group cannot influence those in the second group. Secondly, the nature of the data must be binary or strictly categorical, meaning the outcome being measured for the characteristic of interest must fall into one of two mutually exclusive categories, typically labeled as "success" or "failure."

The most crucial prerequisite involves the size of the samples (n_1 and n_2). The sample sizes must be sufficiently large to ensure that the sampling distribution of the difference between the sample proportions can be accurately approximated by a [normal distribution](#). This approximation is what allows us to effectively utilize the standard normal (Z) distribution for calculating the critical [p-value](#). A practical guideline often cited is that there should be a minimum of ten observed successes and ten observed failures within each of the two independent samples. Meeting this threshold ensures that the test provides statistically robust and accurate conclusions based on the collected sample

data.

Formulating the Null and Alternative Hypotheses

The execution of any statistical test, including the [two proportion z-test](#), mandates the formal establishment of a pair of competing statements: the [null hypothesis](#) and the [alternative hypothesis](#). These two statements represent exhaustive possibilities regarding the true relationship between the two population proportions (π_1 and π_2) and serve as the foundation for the statistical decision-making process.

The [null hypothesis](#), universally symbolized as **H₀**, represents the status quo or the statement of no effect or no difference. It always posits that there is no meaningful statistical difference between the two population proportions being compared. Under the assumption that H₀ is true, any observed disparity or variation noted between the calculated sample proportions is assumed to be purely attributable to the inherent randomness of the sampling process, which is expected whenever data is collected from a subset of the population.

H₀: $\pi_1 = \pi_2$ (Interpretation: The true population proportion of the measured characteristic in population 1 is statistically equivalent to that in population 2.)

Conversely, the [alternative hypothesis](#), denoted as **H₁** (or sometimes H_A), is the research hypothesis that directly challenges the null hypothesis. It suggests that a statistically significant difference does, in fact, exist between the population proportions. The specific mathematical formulation of H₁ is determined by the precise scope of the research question and whether the investigator is seeking evidence for a difference in any direction (a non-directional, or two-tailed test) or a difference where one proportion is specifically hypothesized to be larger or smaller than the other (a directional, or one-tailed test).

H₁ (Two-tailed): $\pi_1 \neq \pi_2$ (The two population proportions are unequal. This is the default choice when the primary interest is simply detecting any difference, regardless of whether π_1 is larger or smaller than π_2 .)

H₁ (Left-tailed): $\pi_1 < \pi_2$ (The population 1 proportion is hypothesized to be strictly less than the population 2 proportion. This directional test is used when the theory specifically predicts a reduction in the first proportion.)

H₁ (Right-tailed): $\pi_1 > \pi_2$ (The population 1 proportion is hypothesized to be strictly greater than the population 2 proportion. This directional test is used when the theory specifically predicts an increase in the first proportion.)

Calculating the Test Statistic and Pooled Proportion

The crucial step in formally evaluating these hypotheses is the calculation of the [test statistic](#). This

metric provides a standardized measure of how far our observed sample data deviates from the expected outcome specified by the [null hypothesis](#). For the two proportion z-test, this statistic takes the form of a Z-score, which is calculated by comparing the observed difference in sample proportions to the standard error of that difference. Because we are testing the hypothesis that the two population proportions are equal ($H_0: \pi_1 = \pi_2$), we must use a method that assumes this equality holds true, which involves calculating a [pooled proportion](#). The formula for the Z-score is structured as follows:

$$z = (p_1 - p_2) / \sqrt{p(1-p)(1/n_1 + 1/n_2)}$$

Understanding the roles of the variables is essential for accurate interpretation. The terms **p1** and **p2** represent the observed [sample proportions](#) for population 1 and population 2, respectively, derived by dividing the number of observed successes by the total [sample size](#) for each group. The variables **n1** and **n2** denote the respective total [sample sizes](#). Most critically, the variable **p** represents the [pooled proportion](#). This pooled value is a weighted average of the two sample proportions, calculated under the strict assumption that the [null hypothesis](#) is true. This pooling provides the most accurate, combined estimate of the common population proportion when we assume H_0 holds.

The formula used for calculating the [pooled proportion](#), **p**, is derived by summing the total number of "successes" across both samples and dividing this sum by the grand total number of observations (the combined sample size):

$$p = (p_1n_1 + p_2n_2) / (n_1 + n_2)$$

The resulting **z-score** quantifies the difference between the observed sample proportions in terms of standard errors. Since the expected difference under the null hypothesis is zero, a larger absolute Z-score signifies a greater deviation from this expectation, thereby providing stronger evidence against the validity of the [null hypothesis](#).

Statistical Decision Making: Interpreting the P-Value

Following the calculation of the [test statistic](#), the next critical step is to determine the associated [p-value](#). The [p-value](#) is a probability measure that quantifies the likelihood of observing sample data as extreme as, or more extreme than, the data actually collected, assuming that the [null hypothesis](#) (H_0) is entirely true. Consequently, a very small [p-value](#) indicates that the observed data configuration is highly improbable if there were truly no difference in the population proportions, suggesting that we should reject the notion of equality.

To make a definitive statistical conclusion regarding the [null hypothesis](#), the calculated [p-value](#) must be compared against a pre-selected [significance level](#), conventionally denoted by the Greek

letter α (alpha). The [significance level](#) sets the threshold for statistical significance and represents the maximum acceptable probability of committing a Type I error--the error of incorrectly rejecting the null hypothesis when it is, in fact, true. Standard, widely adopted choices for α are **0.10**, **0.05** (the most common industry standard), and **0.01**, depending on the required certainty of the conclusion.

The formal decision rule for this [hypothesis test](#) is both simple and definitive:

If the [p-value](#) is **less than or equal to** the chosen [significance level](#) ($p \leq \alpha$), we **reject the null hypothesis**. This outcome provides sufficient statistical evidence to confidently conclude that a significant, non-random difference exists between the two population proportions.

If the [p-value](#) is **greater than** the chosen [significance level](#) ($p > \alpha$), we **fail to reject the null hypothesis**. This result indicates that the observed difference is small enough that it could reasonably be attributed solely to expected random sampling variability, meaning we lack sufficient evidence to conclude that a true difference exists at the predetermined level of significance.

Practical Execution of the Z-Test in SAS

To transition the theoretical concepts into practical application, we will now walk through a detailed, step-by-step example of conducting a [two proportion z-test](#) using [SAS](#). Consider a scenario where an investigator is seeking to determine if there is a statistical disparity in the proportion of residents who express support for a proposed local law between two distinct geographical regions: County A and County B. For this example, we assume a [random sample](#) of 50 residents was surveyed from each county, and their responses were strictly categorized as either "Support" or "Reject."

The initial step in any [SAS](#) analysis is the careful structuring of the data. Since we are working with summarized frequency counts rather than raw, individual observations, we must input the data using the appropriate data structure. The following code snippet demonstrates the creation of a [SAS](#) dataset summarizing the frequency counts of residents based on their opinion (``status``) and their county of residence (``county``). The ``DATA`` step initializes the dataset ``my_data``, and the ``INPUT`` statement defines the variables.

```
/*create dataset - summarize counts*/
```

```
data my_data;
```

```
input county $ status $ count;
```

```
datalines;
```

```
A Support 34
```

```
A Reject 16
```

```
B Support 29
```

```
B Reject 21
```

```
;
```

```
run;
```

```
/*view dataset for verification*/
```

```
proc print data=my_data;
```

Following the data entry, the output generated by the `PROC PRINT` statement confirms that the dataset has been correctly structured, showing the distinct counts for "Support" and "Reject" clearly partitioned by County A and County B, thus preparing the data for statistical analysis.

Obs	county	status	count
1	A	Support	34
2	A	Reject	16
3	B	Support	29
4	B	Reject	21

Once the frequency data is prepared, the statistical procedure can be executed. The [two proportion z-test](#) is most efficiently performed in [SAS](#) using the versatile [PROC FREQ](#) procedure, combined with specialized options designed for risk difference calculations and hypothesis testing.

```
/*perform two proportion z-test using pooled variance*/
```

```
proc freq data=my_data;
```

```
weight count;
```

```
tables county * status / riskdiff(equal var = null);
```

```
run;
```

The key component here is the `WEIGHT count;` statement, which instructs [PROC FREQ](#) to use the values in the `count` variable as the frequency weight for each observation, ensuring correct calculation of total observations and sample proportions. The `TABLES county * status /` command requests the generation of a two-way contingency table. Crucially, the `RISKDIFF` option initiates calculations related to the difference in proportions. The sub-option `EQUAL VAR = NULL` specifically forces [PROC FREQ](#) to employ the standard error based on the assumption of equal variances (i.e., using the [pooled proportion](#)). This mathematically correct approach is essential when performing the [two proportion z-test](#) to test the [null hypothesis](#) of equal population proportions.

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of county by status		
	county	status	
		Reject	Support
A	16	34	50
	16.00	34.00	50.00
	32.00	68.00	
	43.24	53.97	
B	21	29	50
	21.00	29.00	50.00
	42.00	58.00	
	56.76	46.03	
Total	37	63	100
	37.00	63.00	100.00

Statistics for Table of county by status

Column 1 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.3200	0.0660	0.1907	0.4493	0.1952	0.4670
Row 2	0.4200	0.0698	0.2832	0.5568	0.2819	0.5679
Total	0.3700	0.0483	0.2754	0.4646	0.2756	0.4724
Difference	-0.1000	0.0960	-0.2882	0.0882		
Difference is (Row 1 - Row 2)						

Risk Difference Test	
H0: P1 - P2 = 0 Wald Method	
Risk Difference	-0.1000
ASE (H0)	0.0966
Z	-1.0356
One-sided Pr < Z	0.1502
Two-sided Pr > Z 	0.3004
Column 1 (status = Reject)	

Interpreting the SAS Output and Drawing Final Conclusions

Upon successful execution of the [SAS](#) code, the output generated by [PROC FREQ](#) will contain the pivotal **Risk Difference Test** table. This table furnishes the necessary metrics--specifically the calculated [test statistic](#) and the corresponding [p-value](#)--that allow us to make an informed statistical decision regarding the initial hypotheses.

Reviewing the **Risk Difference Test** table for the comparison of County A versus County B, we extract the following critical results:

The calculated **z-test statistic** is **-1.0356**. This score indicates that the observed difference between the support proportions of the two counties is approximately one standard error below the expected difference of zero (which is the value assumed under the [null hypothesis](#)).

The associated **two-sided p-value** is **0.3004**. This crucial probability tells us that if the two counties truly had the exact same population support rate, there would be a 30.04% chance of observing a difference in sample support rates as large as, or larger than, the one we found.

To conclude our [hypothesis test](#), we must compare the obtained [p-value](#) (0.3004) to our chosen [significance level](#) (α). Using the conventional standard threshold of $\alpha = 0.05$, we observe that 0.3004 is significantly greater than 0.05. Therefore, based on the statistical decision rule, we **fail to reject the null hypothesis**. This outcome signifies that the data provides insufficient statistical evidence to confidently conclude that the population proportion of law supporters differs significantly between County A and County B. The small disparity observed in the sample data is highly likely attributable to the expected, random variation inherent in the sampling process.

Extending Your Skills in Categorical Data Analysis

Successfully mastering the [two proportion z-test](#) in [SAS](#) is a foundational achievement in applied statistical analysis, providing a robust method for comparing two proportions. However, the broader field of statistical [hypothesis testing](#) encompasses a much wider range of advanced tools tailored to different data structures and research complexities.

To further enhance your analytical capabilities using [SAS](#), it is highly recommended to explore other fundamental [statistical tests](#). For analyzing continuous variables, understanding t-tests is essential for comparing means between two groups, while ANOVA (Analysis of Variance) allows for the comparison of means across multiple groups simultaneously. For categorical data involving two or more categories, the chi-square test is crucial for analyzing associations and independence. Building on the principles of the two proportion z-test, proficiency in these additional [statistical tests](#) will enable you to address a wider array of analytical challenges and derive deeper insights from diverse datasets.