

# How to Perform Fisher's Exact Test: A Step-by-Step Guide with Example

Authored by  
**stats writer**

March 13, 2026

## RECOMMENDED CITATION

stats writer (2026). *How to Perform Fisher's Exact Test: A Step-by-Step Guide with Example*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=135477>

## Understanding the Fundamentals of Fisher's Exact Test

In the expansive field of statistics, **Fisher's Exact Test** stands as a powerful tool for researchers seeking to analyze the relationship between two categorical variables. This method is specifically categorized as an "exact" test because the significance of the deviation from a null hypothesis can be calculated exactly, rather than relying on an approximation that becomes accurate only as the sample size grows large. Named after the prolific statistician **Sir Ronald Fisher**, the test is renowned for its utility in small-scale experiments where the assumptions required for larger asymptotic tests are not met. By focusing on the exact distribution of the data, it provides a rigorous basis for inferential statistics when data points are scarce.

The primary objective of **Fisher's Exact Test** is to determine if there is a non-random association between two variables that can be classified into mutually exclusive categories. In many practical scenarios, researchers encounter datasets where the observed frequencies are too low to satisfy the requirements of more common tests. While other methods might provide an estimate of the p-value, Fisher's method computes the precise probability of the observed data configuration under the assumption of independence. This makes it an indispensable asset in fields such as medical research, social sciences, and quality control, where gathering massive amounts of data is often either impossible or prohibitively expensive.

To implement this test effectively, one must understand that it operates on the principle of combinations and permutations. It assesses every possible way the observed marginal totals could be distributed across a contingency table. By summing the probabilities of tables that are as extreme as, or more extreme than, the one actually observed, the test yields a statistical significance value. This high level of precision ensures that researchers do not inadvertently commit a Type I error, which involves falsely claiming a relationship exists when it is merely a result of random chance. Consequently, it serves as a gold standard for small-sample categorical data analysis.

## Criteria for Selecting Fisher's Exact Test Over Chi-Square

A frequent question in statistical analysis is when one should opt for **Fisher's Exact Test** instead of the more widely known Pearson's Chi-Square Test. The Chi-Square test relies on the assumption that the sampling distribution of the test statistic approximates a Chi-Square distribution. However, this approximation fails when the expected frequencies in any cell of a contingency table are low--typically defined as less than five. In such instances, the Chi-Square test may provide misleading results, either overestimating or underestimating the true significance of the data. **Fisher's Exact Test** bypasses this limitation entirely by calculating the exact probability rather than an approximation.

The "Rule of Five" is a common heuristic used by statisticians: if more than 20% of the cells in your

contingency table have expected counts less than five, or if any single cell has an expected count of zero, the Chi-Square test is generally considered unreliable. In these specific conditions, **Fisher's Exact Test** is the preferred alternative. While it was historically limited to 2x2 tables due to the computational intensity required for larger matrices, modern computing power has expanded its reach. Nonetheless, it remains most frequently applied to 2x2 tables where the variables are dichotomous, such as "Success/Failure" or "Male/Female."

Furthermore, **Fisher's Exact Test** is particularly robust because it does not assume a specific underlying distribution for the population. It is a non-parametric test that treats the marginal totals of the contingency table as fixed. This fixed-marginal assumption is a cornerstone of the test's mathematical framework. By keeping the row and column totals constant, the test explores all possible internal cell arrangements that could produce those specific totals. This rigorous approach ensures that the resulting p-value is a true reflection of the data's internal structure, providing clarity in high-stakes research environments.

## Defining the Research Hypotheses

Before performing any statistical calculation, it is essential to establish a clear set of hypotheses. For **Fisher's Exact Test**, the process begins with the **Null Hypothesis (H0)**. This hypothesis posits that there is no association between the two categorical variables being studied. In other words, the variables are independent, and any observed difference in proportions is due to random sampling variation. Establishing a strong null hypothesis is the first step toward determining whether the data provides enough evidence to suggest a meaningful pattern exists in the broader population.

Conversely, the **Alternative Hypothesis (H1)** suggests that there is a significant association between the variables. This implies that the variables are dependent on one another, meaning the distribution of one variable differs based on the category of the second variable. The alternative hypothesis can be either one-tailed or two-tailed. A one-tailed test is used when a researcher predicts a specific direction of the relationship, while a two-tailed test is used to detect any significant association regardless of direction.

In the context of **Fisher's Exact Test**, the hypotheses are formally stated as follows:

**H0 (Null Hypothesis):** The two categorical variables are independent; there is no association between them in the population.

**H1 (Alternative Hypothesis):** The two categorical variables are dependent; a significant association exists between them.

Once these hypotheses are defined, the researcher selects a significance level, commonly denoted as alpha ( $\alpha$ ). The most frequent choice is 0.05. If the calculated p-value is less than or

equal to this alpha level, the null hypothesis is rejected in favor of the alternative. If the p-value is greater, the researcher fails to reject the null hypothesis, concluding that there is insufficient evidence to support a relationship between the variables. This structured approach provides an objective framework for making scientific claims.

## Structural Overview of the 2x2 Contingency Table

The mathematical foundation of **Fisher's Exact Test** relies on the organization of data into a 2x2 contingency table. This table serves as a visual and mathematical representation of the frequencies for each combination of the two variables. Each cell in the table represents a specific intersection of categories. For example, in a study of treatment efficacy, one variable might be "Treatment Group" (Treated vs. Control) and the other might be "Outcome" (Recovered vs. Not Recovered). The internal cells, typically labeled **a**, **b**, **c**, and **d**, contain the raw counts for these intersections.

To perform the test, we must also consider the marginal totals, which are the sums of the rows and columns. These totals are vital because **Fisher's Exact Test** treats them as fixed constants. The sum of the first row ( $a+b$ ) and the second row ( $c+d$ ) must equal the total sample size ( $n$ ). Similarly, the sum of the first column ( $a+c$ ) and the second column ( $b+d$ ) must also equal  $n$ . This constraint is what allows for the calculation of an "exact" probability, as it limits the number of possible table configurations that can exist with those specific marginal constraints.

The standard notation for a 2x2 table is illustrated below:

Grouping	Outcome A	Outcome B	Row Total
Category 1	a	b	$a+b$
Category 2	c	d	$c+d$
Column Total	$a+c$	$b+d$	$a+b+c+d = n$

By arranging data in this format, the researcher can easily visualize the distribution and prepare for the **Factorial** calculations required by the Fisher formula. The relationship between these cells is the core of the test: if the ratio of  $a$  to  $b$  is significantly different from the ratio of  $c$  to  $d$ , it suggests that the category of the group influences the outcome, pointing toward a rejection of the null hypothesis. This clear structural representation is the starting point for all subsequent mathematical derivations.

## The Mathematical Formula and Factorial Logic

The calculation of the p-value in **Fisher's Exact Test** involves the use of factorials, denoted by the exclamation point (!). A factorial is the product of an integer and all the integers below it down to

one. Because the test deals with combinations of outcomes, these factorials are necessary to determine the total number of ways the data could have been arranged. The formula specifically calculates the probability of obtaining the exact set of frequencies observed in the table, given the fixed marginal totals.

The mathematical formula for the **one-tailed p-value** in a 2x2 table is expressed as:

$$p = \frac{1}{\dots}$$

In this equation, the numerator is the product of the factorials of all the marginal totals, while the denominator is the product of the factorials of each individual cell count multiplied by the factorial of the total sample size. This calculation yields the probability of the specific table configuration occurring by chance. To find the total p-value, one must also calculate the probabilities of all other possible tables that are even more extreme (those that show an even stronger association) and sum them together. This ensures that the test accounts for all evidence against the null hypothesis.

One of the challenges of **Fisher's Exact Test** is that factorials grow extremely large very quickly. For instance, while 5! is only 120, 10! is over 3.6 million, and 25! is a number with 25 digits. Historically, this made the test difficult to perform by hand for even moderate sample sizes. However, modern statistical software handles these large-scale calculations with ease using logarithms or specialized algorithms. Despite the complexity, the logic remains the same: by comparing the observed probability to the sum of all possible outcomes, we obtain a precise measure of how "rare" or "significant" our observed data truly is.

## Connection to the Hypergeometric Distribution

From a theoretical perspective, **Fisher's Exact Test** is deeply rooted in the hypergeometric distribution. This is a discrete probability distribution that describes the probability of **k** successes in **n** draws, without replacement, from a finite population of size **N** that contains exactly **K** successes. In the context of a 2x2 table, the "population" is the total sample size (**n**), and the "successes" are the counts in one of the categories. Because the marginal totals are fixed, the test essentially models a process of sampling without replacement.

When we calculate the p-value using Fisher's formula, we are essentially finding the probability of a specific outcome in a hypergeometric experiment. The parameters of this distribution correspond directly to the values in our contingency table. Specifically, the population size is **n**, the number of successes in the population is **a+b**, the sample size drawn is **a+c**, and the number of successes in that sample is **a**. This mathematical alignment allows statisticians to use the Cumulative Distribution Function (CDF) of the hypergeometric distribution to determine significance.

Understanding this connection is crucial for advanced statistical modeling. It explains why the test

is "exact"--it is based on a known, discrete distribution rather than a continuous curve that only approximates the data. This foundation makes **Fisher's Exact Test** uniquely reliable. While the Chi-Square test is an "asymptotic" test (meaning it gets better as the sample size approaches infinity), Fisher's test is accurate regardless of how small the sample size is, provided the assumption of fixed marginals holds true for the experimental design.

## Distinguishing One-Tailed and Two-Tailed Results

When reporting results from **Fisher's Exact Test**, it is vital to distinguish between one-tailed and two-tailed p-values. A one-tailed p-value measures the probability in only one direction of the distribution. It is used when the researcher has a prior hypothesis that one group will specifically have a higher or lower proportion than the other. However, in most scientific research, the two-tailed test is the standard requirement because it is more conservative and accounts for the possibility of an association in either direction.

Calculating the two-tailed p-value for **Fisher's Exact Test** is not as simple as doubling the one-tailed value. Because the hypergeometric distribution can be asymmetrical, doubling the one-tailed probability might lead to an incorrect result. Instead, the two-tailed p-value is typically calculated by summing the probabilities of all possible contingency tables that have a probability less than or equal to the probability of the observed table. This approach ensures that all "extreme" outcomes, regardless of which group they favor, are included in the final significance calculation.

Due to these complexities, most practitioners rely on statistical software to derive the two-tailed value. The process involves identifying all permutations of the 2x2 table that maintain the fixed marginal totals, calculating the probability for each using the factorial formula, and then aggregating the values that meet the extremity criteria. This rigorous method provides a comprehensive view of the evidence, ensuring that any conclusion about the association between variables is backed by the full weight of the observed data.

## Fisher's Exact Test: A Practical Example

To illustrate the application of **Fisher's Exact Test**, consider a study investigating the relationship between gender and political party preference. Suppose a researcher wants to know if there is a significant association between being Male or Female and preferring the Democrat or Republican party. Given the small scope of the study, a simple random sample of 25 voters is collected. The results are organized into the following contingency table:

Gender	Democrat	Republican	Total
Male	4	9	13
Female	8	4	12

<b>Total</b>	12	13	25
--------------	----	----	----

In this scenario, our sample size is 25, which is relatively small. Upon inspection, we see that the expected counts in some cells might be low, making **Fisher's Exact Test** the most appropriate choice. We begin by defining our hypotheses: the **Null Hypothesis (H0)** states that gender and political preference are independent, while the **Alternative Hypothesis (H1)** states that they are associated. We set our significance level at 0.05 to determine if the results are statistically meaningful.

Next, we calculate the two-tailed p-value. Using the factorial formula for the observed table and adding the probabilities of more extreme tables, we arrive at the following result:

	Group 1	Group 2
Category 1	<input type="text" value="4"/>	<input type="text" value="9"/>
Category 2	<input type="text" value="8"/>	<input type="text" value="4"/>

One-tailed p value: **0.081178**

Two-tailed p value: **0.115239**

The calculated two-tailed p-value is **0.115239**. This number represents the probability of observing such a distribution (or one even more extreme) if gender and political party preference were completely unrelated. Since 0.115239 is greater than our alpha of 0.05, we do not have enough evidence to reject the null hypothesis. We conclude that, based on this sample, there is no statistically significant association between gender and political party preference.

## Interpreting Results and Drawing Conclusions

Interpreting the output of **Fisher's Exact Test** requires a careful understanding of the p-value in the context of the research question. A common misconception is that a non-significant p-value (one above 0.05) proves that no relationship exists. In reality, it simply means that the data at hand does not provide enough evidence to rule out chance as an explanation for the observed

differences. In our previous example, while we saw that more females preferred the Democrat party (8 vs 4) and more males preferred the Republican party (9 vs 4), the sample size of 25 was too small for this difference to be considered statistically significant.

When the p-value is below the threshold, researchers can confidently state that an association exists. However, it is important to remember that **Fisher's Exact Test** identifies the *existence* of an association but does not inherently measure the *strength* of that association. To understand the magnitude of the effect, researchers often use additional metrics such as the Odds Ratio or Relative Risk. These measures complement the p-value by providing a more complete picture of the practical significance of the findings.

In summary, **Fisher's Exact Test** is a foundational tool for rigorous data analysis. It provides a precise mathematical answer to questions about categorical relationships, especially when dealing with limited data. By adhering to the exact distribution of the data rather than relying on large-sample approximations, it ensures that scientific conclusions are built on a solid logical and mathematical foundation. Whether you are working in a laboratory, a political campaign, or a business environment, mastering this test allows for more accurate and reliable decision-making.

## Computational Resources and Further Learning

While the manual calculation of **Fisher's Exact Test** is educational, modern researchers almost exclusively use statistical software to perform these analyses. Programs like **SPSS**, **R**, and **Python** have built-in functions that can compute p-values for 2x2 and even larger contingency tables in milliseconds. These tools are essential for handling larger datasets or complex experimental designs where manual factorial calculation would be impossible.

For those interested in the technical implementation of this test across different platforms, various tutorials and documentation are available. Utilizing **Python** libraries such as *SciPy* or **SPSS** procedures allows for seamless integration of **Fisher's Exact Test** into broader data science workflows. These tools not only provide the p-value but often include the Odds Ratio and confidence intervals, which are critical for a comprehensive statistical report.

To deepen your understanding of how to perform these tests using specific software, you may explore the following resources:

[How to Perform Fisher's Exact Test in SPSS](#)

[How to Perform Fisher's Exact Test in Python](#)

By leveraging these computational tools, you can apply **Fisher's Exact Test** to real-world data with confidence and precision. Understanding both the mathematical theory and the practical application ensures that your statistical analyses are both accurate and meaningful.