

Learning Fisher's Exact Test with Python: A Step-by-Step Guide

Authored by
Mohammed loot

November 8, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *Learning Fisher's Exact Test with Python: A Step-by-Step Guide*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=12702>

The Power of Fisher's Exact Test in Statistical Analysis

The [Fisher's Exact Test](#) stands as a cornerstone in analytical statistics, specifically designed for scrutinizing the association between two distinct [categorical variables](#). This powerful statistical procedure grants researchers the ability to determine with high precision whether a statistically significant relationship exists between the variables under investigation. What distinguishes Fisher's method from many common alternatives is its commitment to exact calculation: it computes the precise probability of observing the sampled data, or data exhibiting an even stronger relationship, assuming the marginal totals (row and column sums) are held constant. This exact computation removes the reliance on large sample approximations, offering unparalleled reliability when data is sparse.

Historically, the need for an exact test arose as a necessary alternative to the standard [Chi-squared test](#) for independence, particularly Pearson's formulation. While the Chi-squared test is widely used, it operates under crucial assumptions regarding the distribution of data within the [contingency table](#). Specifically, it assumes that the expected frequencies in each cell are sufficiently large. A general rule of thumb dictates that no more than 20% of expected cell counts should be less than 5, and no single expected cell count should be less than 1. When these fundamental assumptions are violated--a common occurrence in specialized studies--the results produced by the Chi-squared approximation become skewed and untrustworthy.

For the modern data scientist or statistician operating within the Python environment, implementing this rigorous test is exceptionally efficient. The comprehensive statistical tools available in the [SciPy library](#) make the process straightforward. This tutorial provides a meticulous walkthrough, covering everything from properly structuring your raw data to executing the necessary Python functions and, crucially, interpreting the resulting [p-value](#) to reach scientifically sound conclusions about the association between your variables. Mastering this technique is essential for maintaining statistical rigor, especially when sample constraints are unavoidable.

The Critical Need for Exact Testing in Small Samples

When researchers organize collected data into a 2x2 table--formally referred to as a [contingency table](#)--they are fundamentally assessing whether the classifications represented by the rows and columns are interdependent. The mathematical foundation marks the primary divergence between Fisher's methodology and the standard Chi-squared test. The latter relies on the assumption that the observed frequencies can be approximated using a continuous probability distribution, specifically the Chi-squared distribution. This approximation is robust only when the sample size is extensive enough to ensure a smooth, continuous distribution of expected counts across all categories.

However, many practical research settings preclude the possibility of gathering large datasets.

Fields such as niche clinical trials, experimental psychology, or specialized biological research often deal with severely limited sample sizes. If, for instance, a study enrolls only 25 participants, the expected frequency within specific, fine-grained categories is highly likely to fall significantly below the accepted threshold of 5. Relying on the [Chi-squared test](#) in these scenarios introduces substantial risk, potentially inflating the Type I error rate (the risk of a false positive) and leading to the incorrect rejection of the [null hypothesis](#).

In stark contrast, the [Fisher's Exact Test](#) sidesteps approximation entirely. It calculates the probability using the [hypergeometric distribution](#). This involves systematically enumerating every possible 2x2 table configuration that could arise given the fixed row and column totals (the marginals). By doing so, it determines the exact probability of observing the current table or any alternative table that represents an even stronger association. This exhaustive calculation guarantees the statistical inference is valid and precise, irrespective of how sparse or unevenly distributed the data is, making it the definitive choice for assessing independence in small, dichotomous datasets.

Establishing the Research Scenario in Python

To clearly demonstrate the practical utility of Fisher's Exact Test, we will examine a typical scenario found in social science investigation. Imagine a research team aiming to ascertain whether a college student's gender is statistically associated with their political party preference. This research involves two nominal, [categorical variables](#): Gender (categorized as Female or Male) and Political Preference (categorized as Democrat or Republican).

The team conducts a preliminary, random poll involving only 25 students across the campus. The resulting counts are meticulously organized into the 2x2 [contingency table](#) shown below. It is immediately apparent that although the overall sample size is 25, several individual cell counts are quite low. This observation is the immediate justification for choosing Fisher's Exact Test over the standard Chi-squared method, as the assumptions for the latter are likely compromised.

The following table summarizes the distribution of the 25 surveyed students based on their reported gender and political affiliation, which will serve as our input data matrix for the Python analysis:

	Democrat	Republican
Female	8	4
Male	4	9

Our primary statistical objective is to utilize these raw counts to definitively conclude whether the relationship observed possesses [statistical significance](#), or if the differences in preference patterns

are merely the result of inherent random variation typical of small-scale sampling.

Implementing Fisher's Exact Test with SciPy

Executing the [Fisher's Exact Test](#) within the Python ecosystem is streamlined and efficient, thanks to the powerful statistical capabilities embedded within the [SciPy library](#). The analytical workflow is divided into two essential tasks: correctly structuring the observed frequency data and invoking the specialized statistical function.

The initial step requires transforming the raw cell counts from the contingency table into a format Python can readily process, typically a nested list or a NumPy array that represents the 2x2 matrix. It is crucial to maintain the exact alignment established in the scenario: the first inner list corresponds to the counts for Females (Row 1), and the second inner list corresponds to the counts for Males (Row 2).

```
data = ,  
]
```

In this structured representation, the values 8 and 4 denote the counts for Females (Democrat and Republican, respectively), and 4 and 9 denote the counts for Males (Democrat and Republican, respectively). Once this matrix is accurately defined, the data is prepared for the core statistical calculation.

The analytical heavy lifting is performed by the `fisher_exact` function, which is nested within the `scipy.stats` module. This function requires the contingency table (our `data` matrix) as its primary input and allows for optional parameters to specify the intended directionality of the test. The syntax structure is clearly defined:

```
fisher_exact(table, alternative='two-sided')
```

The function parameters are precise: **table** must be the 2x2 array of observed frequencies; **alternative** specifies the alternative hypothesis (H1). The default setting, `'two-sided'`, tests for any form of association or difference. For hypotheses that predict a specific direction (e.g., one group having a higher rate than another), one can specify `'less'` or `'greater'` to execute a one-sided test. The following code block demonstrates the necessary import, execution, and resulting output:

```
import scipy.stats as stats
```

```
print(stats.fisher_exact(data))
```

```
(4.5, 0.1152)
```

The resulting output is always a tuple containing two critical statistical measures: the [odds ratio](#) (4.5) and the computed [p-value](#) (0.1152).

Interpreting Statistical Outcomes and Hypothesis Testing

Any rigorous statistical analysis begins by establishing a formal statistical hypothesis framework, consisting of the [null hypothesis](#) (H_0) and the alternative hypothesis (H_1). In the context of the [Fisher's Exact Test](#), these hypotheses focus squarely on the independence of the two categorical variables:

H_0 (Null Hypothesis): The student's gender and their political preference are statistically independent variables. In essence, there is no genuine association between them in the population.

H_1 (Alternative Hypothesis): The two variables are *not* independent. There exists a statistically significant association between gender and political preference.

The ultimate decision regarding the hypotheses hinges entirely upon the calculated [p-value](#), which in our scenario is **0.1152**. The p-value quantifies the probability of obtaining the specific observed data distribution (or an even more extreme one) under the condition that the null hypothesis is true. This value is then systematically compared against a predetermined significance level, often denoted as alpha (α), which is conventionally set at 0.05 (or 5%) across most scientific disciplines.

The process for making the final statistical inference is clear and unambiguous:

If the p-value is less than or equal to the significance level α (e.g., 0.05), the evidence is strong enough to reject H_0 , leading to the conclusion that the observed association is statistically significant.

If the p-value is greater than α , we must fail to reject H_0 , concluding that there is insufficient statistical evidence to assert a significant association between the variables.

Given that our calculated p-value of 0.1152 exceeds the standard threshold of 0.05, the obligatory conclusion is that we **fail to reject the null hypothesis**. This outcome indicates that the observed pattern of political preferences across genders could plausibly have occurred simply due to random sampling variability, even if gender and political preference are genuinely independent in the broader student population.

Conclusion: Rigor in Sparse Data Analysis

The analysis performed using the [SciPy](#) implementation of [Fisher's Exact Test](#) leads us to conclude that we lack sufficient evidence to claim a [significant association](#) between gender and

political party preference among the 25 surveyed college students. Based on this precise statistical measure, the two variables are treated as statistically independent within the limitations of this specific dataset.

It is fundamentally important to grasp that failing to reject the [null hypothesis](#) does not equate to proving independence. Rather, it suggests that either the association is genuinely weak, or, more likely given the methodology's purpose, the sample size ($n=25$) was inadequate to achieve the necessary statistical power required to detect a subtle association at the 0.05 significance level. Had the researchers aimed to increase their ability to detect a true difference, they would need to recruit a substantially larger sample. A larger sample size might subsequently satisfy the required expected cell count criteria, potentially allowing for the use of the more computationally expedient [Chi-squared test](#).

This tutorial successfully demonstrated the vital process of applying Fisher's Exact Test in Python. By meticulously calculating the exact probability, this method ensures that statistical inferences remain robust and reliable, providing the gold standard for analyzing independence when dealing with sparse data contained within 2x2 contingency tables, thereby ensuring the highest level of rigor in quantitative research.