

How to Easily Perform Fisher's Exact Test in SAS

Authored by
stats writer

December 1, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Easily Perform Fisher's Exact Test in SAS*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=103484>

The Fisher's Exact Test, implemented efficiently within the **SAS statistical software environment**, is a crucial nonparametric test designed for the rigorous analysis of categorical data. Specifically, it is employed when researchers need to assess the independence between two nominal variables presented in a two-way contingency table, typically of the 2x2 structure. The core objective is to determine whether the observed association or differences between the categories are statistically significant or merely due to random chance inherent in the sampling process.

In SAS, this powerful statistical procedure is executed using the standard **PROC FREQ** procedure. By applying the specific `EXACT` option, SAS calculates the exact probability of observing the data distribution, or a distribution even more extreme, assuming the variables are truly independent. Unlike approximation methods, the output of this procedure yields the precise p-value, which is essential for making robust decisions regarding the presence of a statistically significant association between the two categorical variables under investigation. This precise calculation makes Fisher's Exact Test indispensable in small-sample research.

Understanding the Role of Fisher's Exact Test

The primary utility of the Fisher's Exact Test lies in its ability to accurately evaluate the association between two categorical variables, particularly when dealing with small sample sizes. A significant association implies that the distribution of one variable is dependent on the level or category of the second variable. This test provides a definitive method for establishing whether or not there is a meaningful and statistically significant relationship between the factors being studied, such as gender and political preference, or exposure to a risk factor and disease outcome.

It is commonly utilized as a precise alternative to the traditional Chi-square test of independence. This preference arises because the Chi-square test relies on the assumption that the expected frequency count in every cell of the contingency table is greater than a specified value (usually 5). When one or more of these expected cell counts fall below this threshold--a frequent occurrence in medical studies, pilot projects, or specialized surveys with limited participants--the Chi-square approximation becomes unreliable, potentially inflating the risk of a Type I error (false positive). The Fisher's Exact Test avoids this reliance on approximation by calculating exact probabilities based on the hypergeometric distribution, guaranteeing a reliable result regardless of cell count size.

This test is particularly valuable in fields like epidemiology, genetics, and quality control, where researchers often work with restricted data sets or must analyze rare events. By providing an exact solution, Fisher's test upholds the integrity of the statistical inference even under constraints that would invalidate less robust methods.

Formulating the Hypotheses

Before conducting the analysis, it is mandatory to define the formal null and alternative hypotheses. These statements formalize the research question and dictate how the resulting p-value should be interpreted. For Fisher's Exact Test, which examines the association between two nominal variables, the hypotheses are framed around the concept of statistical independence:

H0: (Null Hypothesis) The two variables are statistically **independent**. This suggests that the categories of one variable occur irrespective of the categories of the second variable; there is no association.

H1: (Alternative Hypothesis) The two variables are **not independent** (i.e., they are associated or dependent). This is the hypothesis the researcher seeks to support if the evidence against H0 is strong enough.

The statistical decision hinges entirely on the calculated p-value. This value represents the probability of observing the current data (or more extreme data) assuming the null hypothesis (H0) is true. If this probability is extremely low--specifically, smaller than the pre-determined significance level (α , commonly set at 0.05)--we possess sufficient evidence to reject the null hypothesis and conclude that a statistically significant association exists.

Illustrative Example: Association Between Gender and Political Preference

To demonstrate the practical application of Fisher's Exact Test within SAS, we will examine a case study in which researchers investigate whether a student's **gender** is associated with their preferred **political party affiliation** at a specific college.

A small, targeted sample of 25 students is surveyed regarding their political preference (Democrat or Republican). Because the total sample size is small ($N=25$), and the cell counts are likely to be low, using the Fisher's Exact Test is the appropriate methodology to ensure accurate results. The raw survey responses are aggregated into the 2x2 contingency table presented below, showing the observed frequencies for each gender/party combination:

	Democrat	Republican
Female	8	4
Male	4	9

From the margins, we observe 12 female students (total) and 13 male students (total). We also see 12 Democrats and 13 Republicans. The analytical goal is to determine if the specific distribution shown (e.g., the higher proportion of Republican males compared to Republican females) is statistically significant enough to reject the idea that gender and political preference are

independent variables. The following steps outline the exact procedure in SAS.

Step 1: Preparing and Entering Data in SAS

The foundational step in running any analysis in SAS is correctly structuring the raw data. Since Fisher's Exact Test requires frequency counts based on the interaction of two categorical factors, we must create a dataset where each row corresponds to an individual observation. We define the dataset as **my_data** and use the `DATALINES` statement to input the results sequentially. We designate two character variables using the dollar sign (`$`): `Party` and `Gender`.

The code below meticulously creates this dataset, replicating the exact counts displayed in the 2x2 table. This setup is mandatory for the subsequent execution of the `PROC FREQ` procedure.

```
/*create data to hold survey results*/
```

```
data my_data;
```

```
input Party $ Gender $;
```

```
datalines;
```

```
Rep Female
```

```
Rep Female
```

```
Rep Female
```

```
Rep Female
```

```
Rep Male
```

```
Rep Male
```

```
Rep Male
```

```
Rep Male
```

```
Rep Male
```

```
Rep Male
```

```
Rep Male
```

```
Rep Male
```

```
Rep Male
```

```
Dem Female
```

```
Dem Female
```

```
Dem Female
```

```
Dem Female
```

```
Dem Female
```

```
Dem Female
```

```
Dem Female
```

```
Dem Female
```

```
Dem Male
```

```
Dem Male
```

```
Dem Male
```

```
Dem Male
```

```
;
```

```
run;
```

Ensuring that `Party` and `Gender` are defined as character variables (indicated by `$`) is crucial, as this instructs SAS to treat them as nominal categories, which is the required data type for the Fisher's Exact Test.

Step 2: Executing the Fisher's Exact Test Using PROC FREQ

With the data correctly loaded, we can now proceed to the statistical analysis using **PROC FREQ**. This procedure is specifically designed in SAS for producing frequency tables and performing tests of association for categorical data.

To instruct SAS to perform the Fisher's Exact Test, we must use the `TABLES` statement to specify the cross-tabulation (`Party*Gender`) and append the key option: `/ FISHER`. This option is what distinguishes the request from a standard Chi-square analysis, forcing SAS to calculate the exact probability based on the theoretical distribution of the observed data.

```
/*perform Fisher's Exact test*/  
proc freq data=my_data;  
tables Party*Gender / fisher;  
run;
```

The resulting output will include the detailed contingency table, followed by various test statistics. The `FISHER` option specifically requests the precise calculation required for the exact test, providing highly reliable results, which is essential given the low cell counts in this scenario.

Interpreting the SAS Output for Statistical Significance

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Party by Gender			
	Party	Gender		Total
		Female	Male	
Dem	8 32.00 66.67 66.67	4 16.00 33.33 30.77	12 48.00	
Rep	4 16.00 30.77 33.33	9 36.00 69.23 69.23	13 52.00	
Total	12 48.00	13 52.00	25 100.00	

Statistics for Table of Party by Gender

Statistic	DF	Value	Prob
Chi-Square	1	3.2216	0.0727
Likelihood Ratio Chi-Square	1	3.2927	0.0696
Continuity Adj. Chi-Square	1	1.9439	0.1632
Mantel-Haenszel Chi-Square	1	3.0927	0.0786
Phi Coefficient		0.3590	
Contingency Coefficient		0.3379	
Cramer's V		0.3590	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	8
Left-sided Pr <= F	0.9869
Right-sided Pr >= F	0.0812
Table Probability (P)	0.0681
Two-sided Pr <= P	0.1152

The SAS output provides the necessary statistics to evaluate the test. We recall that the null hypothesis states that gender and political party preference are independent. Since the research question is whether there is any association at all (a non-directional question), we are performing a two-sided test.

Therefore, the focus of the interpretation must be the **two-sided p-value** listed under the "Fisher's Exact Test" section in the output table. For this example, the calculated two-sided p-value is reported as **0.1152**.

To reach a statistical decision, we compare this p-value against the conventional significance level ($\alpha = 0.05$). The critical rule for hypothesis testing is: If $p \leq \alpha$, we reject the null hypothesis. Since $0.1152 > 0.05$, we **fail to reject the null hypothesis**. The probability of observing this data distribution (or one more extreme) if the variables were independent is 11.52%, which is too high to dismiss the null hypothesis.

Conclusion and Summary of Findings

Based on the results of the Fisher's Exact Test conducted in SAS, we conclude that there is insufficient evidence to claim a statistically significant association between a student's gender and their political party preference within this college sample. In simpler terms, the observed distribution of party affiliations across genders, while showing slight numerical differences, is not unusual enough to definitively prove dependence between the two variables. The differences observed could easily be the result of natural random variation in the sampled population.

This analysis underscores the power and necessity of using exact methods for categorical data analysis, especially when faced with data scarcity or non-uniform distributions. If a Chi-square test had been incorrectly applied, the resulting approximate p-value might have led to an inaccurate inference, potentially jeopardizing the validity of the research conclusion.

Further Resources for Categorical Data Analysis

For those looking to deepen their understanding of categorical data analysis and hypothesis testing within SAS, the following tutorials provide additional information concerning the nuances of Fisher's Exact Test, its theoretical background, and related statistical methods for evaluating nominal variables: