

# Learning Fisher's Exact Test in SAS: A Step-by-Step Guide

Authored by  
**Mohammed loot**

November 1, 2025

## RECOMMENDED CITATION

Mohammed loot (2025). *Learning Fisher's Exact Test in SAS: A Step-by-Step Guide*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=7704>

## The Necessity of Fisher's Exact Test in Statistical Analysis

The [Fisher's Exact Test](#) stands as an indispensable tool in modern statistics, specifically designed for analyzing the relationship between two [categorical variables](#). Unlike approximation methods, this technique utilizes calculations based on exact probabilities to rigorously determine whether a statistically [significant association](#) exists between the variables of interest. Its non-parametric nature makes it exceptionally reliable, providing precise results without relying on assumptions about the underlying distribution of the data.

This test is typically deployed as a robust and necessary alternative to the standard [Chi-Squared Test](#), particularly when dealing with situations characterized by limited sample sizes. A critical rule of thumb dictates that Fisher's Exact Test must be employed whenever one or more of the expected cell counts in a [2x2 table](#) analysis falls below the threshold of 5. When expected counts are sparse, the Chi-Squared approximation becomes unreliable, potentially leading to inaccurate statistical inferences.

By meticulously calculating the exact probability of observing the given data (or more extreme data) under the assumption of independence, Fisher's test successfully bypasses the inherent approximations associated with asymptotic methods like the Chi-Squared approach. This ensures that researchers can draw reliable and trustworthy conclusions, even when analyzing datasets where sparsity or small cell frequency would otherwise compromise the validity of the analysis.

## Defining Hypotheses and the Decision Rule

Before any statistical test is executed, it is fundamental to clearly articulate the [null and alternative hypotheses](#) that frame the research question. These hypotheses provide the basis upon which we evaluate the evidence provided by our data:

**H0: (Null Hypothesis)** The two categorical variables are **independent**. This suggests that there is no statistically significant association or relationship between them in the population.

**H1: (Alternative Hypothesis)** The two categorical variables are **not independent**. This hypothesis posits that a significant association or relationship does indeed exist between the variables.

The ultimate decision regarding whether to reject the null hypothesis rests entirely on the calculated [P-value](#). The P-value represents the probability of obtaining the observed data (or results that are more extreme) assuming that the null hypothesis ( $H_0$ ) is definitively true. Consequently, a small P-value indicates that the observed data is highly unlikely if the variables were truly independent.

The decision rule is straightforward: If the calculated P-value resulting from Fisher's Exact Test is

less than the predefined [significance level](#) (conventionally denoted as  $\alpha$ , which is typically set at 0.05), then we possess sufficient statistical evidence to confidently reject  $H_0$ . Rejecting the null hypothesis allows us to conclude that the two categorical variables are significantly associated. Conversely, if the P-value is greater than  $\alpha$ , we fail to reject  $H_0$ , meaning the observed differences could plausibly be attributed to random sampling variability.

## Case Study: Gender and Political Preference Using SAS

To effectively demonstrate the implementation of Fisher's Exact Test within the statistical environment of [SAS](#) software, let us examine a typical research scenario. This example explores the potential existence of an association between a student's gender and their declared political party preference within a specific university setting. Understanding this relationship requires careful analysis of the collected frequency data.

For this demonstration, a modest random survey was conducted among 25 students on campus. Data were gathered regarding two key variables: the student's gender (categorized as Female or Male) and their primary political affiliation (categorized as Democrat or Republican). Given the inherent constraints of a small sample size, which leads directly to small expected cell counts, Fisher's Exact Test is the most appropriate and rigorous methodology for establishing whether a statistically significant relationship exists.

The results of this small, focused survey are systematically organized and presented in the 2x2 contingency table below, which summarizes the raw counts for each combination of gender and political affiliation:

|        | Democrat | Republican |
|--------|----------|------------|
| Female | 8        | 4          |
| Male   | 4        | 9          |

The subsequent steps will detail the precise procedure necessary to execute this analysis within the SAS environment, transforming this raw frequency table into actionable statistical evidence regarding the independence of these two variables.

### Step 1: Preparing and Structuring Data for SAS Analysis

The initial and most critical phase of any analysis performed in SAS involves accurately structuring the raw data into a usable format. Since the primary procedure for this test, [PROC FREQ](#), typically operates on transactional data (where each row represents a single observation), we must convert our summary counts into this structure.

We will define a new SAS dataset titled **my\_data**. This dataset contains two essential character variables: **Party** and **Gender**. The data lines following the **DATALINES** statement meticulously list the political preference and gender combination for all 25 participants, translating the counts from the contingency table into individual records.

```
/*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;
```

This efficiently executed code block ensures that the dataset is clean, properly formatted, and ready for advanced statistical processing, setting the stage for the execution of the exact

probability test in the following step.

## Step 2: Executing Fisher's Exact Test using PROC FREQ

In the SAS software environment, the dedicated procedure for analyzing cross-classified categorical data and calculating crucial non-parametric statistics, including Fisher's Exact Test, is `PROC FREQ`. This powerful procedure is used to generate contingency tables and output the necessary associated test statistics.

To perform the required analysis, we utilize the `TABLES` statement. Within this statement, the two variables under investigation--`Party` and `Gender`--are listed and separated by an asterisk (\*), which instructs SAS to create a cross-tabulation. Crucially, we append the `/ FISHER` option to the statement. This specific option explicitly mandates that SAS calculate the exact P-value utilizing Fisher's methodology, ensuring the accuracy required for small sample analysis.

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

Upon execution, this code generates a comprehensive output that includes the raw contingency table, expected cell counts, standard statistical measures such as the Chi-Squared results (which should be disregarded in favor of the exact test), and finally, the specialized output block containing the precise results of Fisher's Exact Test.

## The FREQ Procedure

Frequency  
Percent  
Row Pct  
Col Pct

| Table of Party by Gender |        |       |        |
|--------------------------|--------|-------|--------|
| Party                    | Gender |       |        |
|                          | Female | Male  | Total  |
| Dem                      | 8      | 4     | 12     |
|                          | 32.00  | 16.00 | 48.00  |
|                          | 66.67  | 33.33 |        |
|                          | 66.67  | 30.77 |        |
| Rep                      | 4      | 9     | 13     |
|                          | 16.00  | 36.00 | 52.00  |
|                          | 30.77  | 69.23 |        |
|                          | 33.33  | 69.23 |        |
| Total                    | 12     | 13    | 25     |
|                          | 48.00  | 52.00 | 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 |

## Interpreting the Results from SAS Output

The objective of interpreting the SAS output is to determine whether we possess sufficient evidence to reject the null hypothesis of independence between gender and political preference. Since we are testing for any general association (a difference in either direction), we focus on the results of the two-sided test.

In the final section of the SAS output, clearly designated "Fisher's Exact Test," we must locate and

identify the **Two-sided P-value**. For this specific analysis, the calculated two-sided P-value is reported as **0.1152**. This value represents the probability of observing our specific sample distribution, or one even more extreme, if gender and party preference were truly independent.

To finalize the statistical decision, we compare the calculated P-value (0.1152) against our predetermined significance level,  $\alpha = 0.05$ . Because 0.1152 is substantially greater than 0.05, we conclude that the data do not provide sufficient statistical evidence to warrant the rejection of the null hypothesis ( $H_0$ ).

Therefore, the definitive statistical conclusion derived from this sample data is that we cannot assert the existence of a significant association between a student's gender and their political party preference. The observed variation in counts within the 2x2 table is likely attributable to random sampling variability rather than reflecting a true, underlying relationship in the broader student population.

## Further Reading and Resources

For readers seeking to deepen their understanding of exact probability tests, non-parametric statistics, and the comprehensive applications of SAS procedures, the resources listed below offer invaluable supplementary information and practical tutorials on Fisher's Exact Test and related methodologies.