

How to Easily Perform a Chi-Square Goodness of Fit Test in SAS

Authored by
stats writer

December 1, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Easily Perform a Chi-Square Goodness of Fit Test in SAS*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=103459>

The Chi-Square Goodness of Fit Test is a fundamental statistical tool used to determine if the observed distribution of a categorical variable significantly differs from a specified theoretical or hypothesized distribution. When working within the SAS Statistical Analysis System, this test is executed efficiently using the powerful PROC FREQ procedure. This process requires meticulously preparing the data set to include the observed counts for each category. Once the data is ready, we utilize the TABLES statement coupled with the CHISQ option to compel SAS to calculate the necessary Chi-Square test statistic and the corresponding p-value. The final step involves rigorously interpreting these results to determine whether the observed data provides sufficient evidence to reject the initial hypothesis regarding the population distribution.

Understanding how to implement this specific test within the SAS environment is essential for researchers dealing with frequency data. The procedure is surprisingly streamlined, relying heavily on the appropriate setup of the input data and the correct specification of procedural options. The core requirement for the test is comparing the frequencies we actually recorded (observed counts) against the frequencies we would theoretically expect if the null hypothesis were true (expected counts). By following the structured steps outlined below, you can confidently apply the Chi-Square Goodness of Fit Test and draw statistically sound conclusions regarding the fit of your data to a specified model.

A Chi-Square Goodness of Fit Test is fundamentally used to determine whether or not a distribution of a categorical variable follows a specified, hypothesized distribution. This powerful statistical technique is widely applied across various fields, including market research, biology, and quality control, whenever the uniformity or adherence of observed outcomes to an expected model needs verification.

The following comprehensive example explains how to execute, interpret, and report the results of a Chi-Square Goodness of Fit Test using the SAS software package.

Example: Testing for Uniform Customer Distribution in SAS

Consider a scenario involving a local shop owner who strongly asserts that the customer traffic flowing into his establishment is evenly distributed across all five weekdays (Monday through Friday). This assertion implies that, under normal operating conditions, an equal proportion of total weekly customers should visit the shop on each specific day. To test the validity of this claim-- which serves as our null hypothesis--a dedicated researcher meticulously records the total number of customers who enter the shop throughout a representative single week.

The goal of the study is to use the collected data to statistically assess if the observed frequencies deviate significantly from the theoretically expected uniform distribution. If the shop owner's claim is accurate, the expected count for each day would be the total number of customers divided by five. The researcher collected the following observed customer counts for the week:

Monday: 50 customers

Tuesday: 60 customers

Wednesday: 40 customers

Thursday: 47 customers

Friday: 53 customers

We will utilize the subsequent steps to formally perform the Chi-Square Goodness of Fit Test in SAS. This statistical procedure will provide empirical evidence to determine definitively if the collected data is consistent with the shop owner's claim of uniform daily traffic, or if we must conclude that the distribution of customer visits varies significantly across the weekdays.

Defining Hypotheses for the Goodness of Fit Test

Before diving into the coding, it is essential to formally establish the statistical hypotheses that the Chi-Square Goodness of Fit Test is designed to evaluate. These hypotheses guide the interpretation of the final statistical output and determine the course of action regarding the shop owner's claim. In this specific scenario, we are testing for a uniform distribution, meaning the probability of a customer arriving on any given weekday (P_Mon, P_Tue, etc.) is exactly 1/5 or 20%.

The test employs the following set of null and alternative hypotheses:

H0 (Null Hypothesis): The distribution of customer arrivals across the five weekdays is uniform. This asserts that the observed frequencies do not differ significantly from the expected frequencies, based on the assumption that all weekdays are equally likely.

HA (Alternative Hypothesis): The distribution of customer arrivals across the five weekdays is not uniform. This suggests that the observed data deviates substantially from the assumed expected distribution, indicating that some days attract significantly more or fewer customers than others.

Our objective is to calculate the test statistic and corresponding p-value, which will allow us to assess the strength of the evidence against the null hypothesis. Typically, we compare the p-value against a predetermined significance level, often set at 0.05 (or 5%), to make a conclusive decision.

Step 1: Creating the Dataset Structure in SAS

The first operational step in SAS involves accurately inputting the collected observed frequencies into a structured dataset. Since our data is already aggregated--meaning we have the total count for each category rather than a list of individual customer records--we must ensure the dataset is set up to reflect this frequency weighting. We will create a dataset named **my_data**, which contains two essential variables: the categorical variable identifying the day (``Day``) and the variable holding

the observed frequency (`Customers`).

The following SAS code block demonstrates the construction of this dataset using the `DATA` and `DATALINES` statements. This syntax is fundamental for initializing any analysis in SAS, allowing us to define the structure of our data before processing. The character variable `Day` is denoted by the dollar sign (\$) in the `INPUT` statement, while `Customers` is defined as a numeric count variable.

```
/*create dataset containing observed customer counts*/
```

```
data my_data;  
input Day $ Customers;  
datalines;  
Mon 50  
Tue 60  
Wed 40  
Thur 47  
Fri 53  
;  
run;
```

```
/*print dataset for verification*/
```

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

Executing the `PROC PRINT` step immediately confirms that the data has been loaded correctly into the SAS environment, reflecting the exact counts observed by the researcher.

Obs	Day	Customers
1	Mon	50
2	Tue	60
3	Wed	40
4	Thur	47
5	Fri	53

The resulting table confirms that 250 total customers were observed throughout the five-day period ($50 + 60 + 40 + 47 + 53 = 250$). This total count is crucial, as the Chi-Square test internally calculates the expected count by dividing this total (250) equally among the five categories ($250 / 5 = 50$ expected customers per day), thereby testing the uniformity assumption.

Step 2: Executing the Chi-Square Goodness of Fit Test using PROC FREQ

With the dataset successfully created, the next critical phase involves invoking the PROC FREQ procedure, which is the standard mechanism in SAS for analyzing frequency distributions and performing categorical data tests. Two key components must be specified within this procedure to ensure the Goodness of Fit Test is executed correctly.

Firstly, we must use the `WEIGHT` statement, pointing it to the `Customers` variable. This command instructs SAS that the values in the `Customers` column represent the frequencies (or weights) of the observations, rather than the observations themselves. This is absolutely essential for using aggregated data with PROC FREQ for this specific test. Secondly, the `TABLES` statement specifies the categorical variable being analyzed (`Day`), and the crucial `CHISQ` option is appended to request the computation of the Chi-Square statistic. By default, when no specific expected proportions are given, PROC FREQ assumes the expected frequencies are equal across all categories, perfectly aligning with our test for uniform distribution.

```
/*perform Chi-Square Goodness of Fit test, assuming equal probability (uniform fit)*/  
proc freq data=my_data;  
tables Day / chisq;  
weight Customers;  
run;
```

The execution of this code generates extensive output, including frequency tables and, most importantly, the statistical summary required for our hypothesis test. The output is structured to first display the observed and expected counts and then present the summary statistics for the Chi-Square test itself.

The FREQ Procedure

Day	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Fri	53	21.20	53	21.20
Mon	50	20.00	103	41.20
Thur	47	18.80	150	60.00
Tue	60	24.00	210	84.00
Wed	40	16.00	250	100.00

Chi-Square Test for Equal Proportions	
Chi-Square	4.3600
DF	4
Pr > ChiSq	0.3595

Interpreting the Statistical Output and Test Results

The output generated by `PROC FREQ` contains the core numerical evidence required to reach a conclusion regarding the shop owner's claim. We must focus specifically on the table labeled "Chi-Square Test for Specified Proportions," although since we did not specify proportions, SAS automatically uses the uniform distribution as the expected proportion.

From the key statistical metrics provided in the output, we extract the following crucial values:

The Chi-Square test statistic (χ^2): **4.36**. This value quantifies the total discrepancy between the observed customer counts and the expected counts (50 for each day). A larger statistic indicates a greater deviation from the hypothesized uniform distribution.

The degrees of freedom (df): **4**. This is calculated as the number of categories (5 weekdays) minus one ($5 - 1 = 4$).

The corresponding asymptotic p-value: **0.3595**. The p-value represents the probability of observing data as extreme as, or more extreme than, the data collected, assuming the null hypothesis (uniform distribution) is true.

These three statistics are the foundation for the final decision. The Chi-Square test assesses whether the sum of squared differences between observed and expected counts, relative to the expected counts, is large enough to be statistically significant.

Drawing Conclusions Based on the P-Value

The final stage of the hypothesis test involves comparing the calculated p-value to our predetermined significance level (alpha), which is conventionally set at 0.05. The rule for decision making is straightforward: if the p-value is less than alpha (0.05), we reject the null hypothesis (H_0). Conversely, if the p-value is greater than or equal to alpha, we fail to reject the null hypothesis.

In this specific example, the calculated p-value is 0.3595. Comparing this to the threshold:

$$0.3595 > 0.05$$

Since the p-value (0.3595) is substantially greater than the significance level of 0.05, we formally fail to reject the null hypothesis. This finding indicates that the differences observed between the actual customer counts (e.g., 60 on Tuesday, 40 on Wednesday) and the expected count of 50 for each day are likely due to random sampling variation, and are not statistically significant deviations from a uniform distribution.

Final Statistical Conclusion

The failure to reject the null hypothesis leads to the final interpretation: we do not possess sufficient statistical evidence, at the 5% significance level, to conclude that the true distribution of customer arrivals differs significantly from the uniform distribution claimed by the shop owner. While the observed counts are not perfectly uniform, the statistical test confirms that these small variations are within the range expected under the assumption of equal traffic across all five weekdays.

Therefore, we conclude that the data is consistent with the shop owner's assertion that an equal number of customers visit the shop each weekday. This demonstrates the power of the Chi-Square Goodness of Fit Test in validating theoretical distributions against real-world frequency data using the robust analytical capabilities of SAS software.

Further Exploration of Chi-Square Testing

For researchers seeking to deepen their knowledge of categorical data analysis, mastering the nuances of the Chi-Square family of tests is highly recommended. The goodness of fit test is just one application; the Chi-Square test for independence, for instance, is used to determine if there is an association between two different categorical variables.

Understanding how to handle non-uniform expected distributions (where the expected proportion for each category must be explicitly defined in PROC FREQ using the `EXACT` or `P=` options)

provides an even greater level of analytical flexibility within the SAS environment. The ability to precisely define the hypothesized model is crucial for advanced statistical modeling.

The following resources provide additional information and alternative examples illustrating various applications of the Chi-Square Goodness of Fit test and related methodologies:

ARABPSYCHOLOGY.COM