Learning the Chi-Square Goodness of Fit Test: A Step-by-Step Guide Using the TI-84 Calculator

Authored by

Mohammed looti

November 8, 2025

The **Chi-Square Goodness of Fit Test** is a foundational statistical procedure designed to determine if the frequency distribution observed in a sample deviates significantly from a hypothesized or theoretical distribution. This essential tool allows analysts to rigorously test whether a categorical variable aligns with a specific probability pattern, or if the variance between what is observed and what is expected is merely attributable to random sampling fluctuation.

Proficiency in executing and interpreting this test is indispensable for accurate data analysis in fields ranging from market research to biology. This detailed, step-by-step tutorial is specifically structured to guide users through performing the Chi-Square Goodness of Fit Test with precision and efficiency using a TI-84 calculator, focusing on clarity in input and robust interpretation of the resulting statistical metrics.

The Purpose and Framework of the Chi-Square Goodness of Fit Test

The primary objective of the Chi-Square Goodness of Fit Test is to validate an assumed distribution model against actual data collected from a population. Unlike parametric tests that typically compare measures of central tendency (like means), this non-parametric test operates by contrasting the observed counts in various categories against the counts predicted if the theoretical distribution were perfectly accurate. The resulting test statistic, often denoted as X^2, serves as a quantitative measure of the overall discrepancy between the empirical data and the theoretical model.

Central to conducting any hypothesis test is the formal definition of the two competing hypotheses. The analysis must always be framed around the null hypothesis (H?) and the alternative hypothesis (H?). The null hypothesis invariably asserts that there is no significant difference between the two distributions--meaning the observed data fits the expected distribution. Conversely, the alternative hypothesis states that the observed data does not fit the expected distribution, implying that the underlying theoretical model is an inaccurate representation of the true population distribution.

Consider a scenario where we hypothesize that customer foot traffic is evenly distributed across four different retail locations. The null hypothesis (H?) assumes that the true population distribution of customers is uniform across all four locations. The alternative hypothesis (H?) suggests that the distribution is not uniform, indicating that some locations attract a disproportionately higher or lower volume of customers than others. The statistical goal of the Goodness of Fit Test is to determine if the available sample evidence is strong enough to justify rejecting H? in favor of the alternative explanation, providing conclusive proof that the theoretical distribution is flawed.

Essential Prerequisites and Assumptions for Reliable Results

To ensure that the results generated by the TI-84 calculator are statistically valid, several fundamental assumptions related to the data collection and structure must be rigorously satisfied. Failing to adhere to these prerequisites can lead to distorted P-values, resulting in inaccurate conclusions and fundamentally undermining the integrity of the hypothesis test.

Firstly, the data must consist of raw frequencies or counts pertaining to a single categorical variable, and these observations must be obtained through a process of random sampling from the population of interest. This requirement is paramount because it ensures the independence of observations; the count recorded for one category must not influence, or be influenced by, the count recorded for any other category. For example, if we are analyzing sales by region, the sales volume in Region A must be independent of the sales volume in Region B.

Secondly, and perhaps the most critical mathematical assumption for the Chi-Square Goodness of Fit Test, is the requirement regarding expected frequencies. It is conventionally mandated that the expected count (E) for every single category must be greater than or equal to five ($E \geq 5$). If any expected frequency falls below this threshold, the sampling distribution of the test statistic may not accurately approximate the theoretical Chi-Square distribution, potentially necessitating corrective measures such as combining categories or utilizing alternative statistical methods that do not rely on large sample approximations.

Finally, the categories used in the analysis must be both mutually exclusive and exhaustive. Mutually exclusive means that every single observation can only belong to one category, preventing double counting. Exhaustive means that every possible observation relevant to the study must be accounted for within the defined categories. Meeting these assumptions ensures that the calculated P-value accurately reflects the probability of observing such data if the null hypothesis were true, thus lending confidence and trustworthiness to the final statistical conclusion.

Practical Example: Analyzing Customer Distribution

To illustrate the computational procedure, we will use a straightforward business scenario. Imagine a retail shop owner who confidently asserts that customer traffic is uniformly distributed across the five standard weekdays (Monday through Friday). To verify this assertion, a research team records the number of customers arriving over a representative week, yielding the following observed data:

Monday: 50 customers
Tuesday: 60 customers
Wednesday: 40 customers
Thursday: 47 customers
Friday: 53 customers

The shop owner's assertion--that customer distribution is equal across all weekdays--serves as our null hypothesis (H?). Our task is to employ the **Chi-Square Goodness of Fit Test** to determine if the observed counts are statistically consistent with this theoretical uniform distribution or if the variations are significant enough to warrant rejecting the owner's claim.

Before using the TI-84 calculator, we must first determine the expected frequency for each category. The total number of observed customers tracked across the week is calculated by summing the daily counts: 50 + 60 + 40 + 47 + 53 = 250 total customers. Since the hypothesized distribution is equal (uniform) across the five weekdays, the expected count for any single day is the total number of customers divided by the number of categories (k=5). Thus, the expected count (E) for every weekday is 250 / 5 = 50. Since all expected counts are 50, which is significantly greater than the requirement of 5, the critical assumption for the test is met, and we can proceed with the technological analysis.

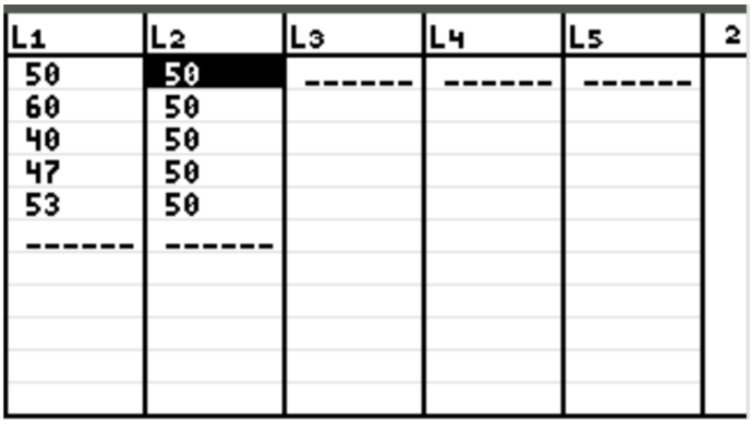
Step 1: Preparing and Inputting Data on the TI-84

The initial and most critical stage of using the TI-84 calculator involves accurately inputting both the observed frequencies and the calculated expected frequencies into designated lists. Proper list organization is essential, as the calculator relies on these stored values to compute the test statistic and P-value correctly.

To begin the data entry process, press the Stat button, and then select the EDIT menu option. This action opens the list editor, where data is typically managed in lists L1, L2, L3, and so forth. We will meticulously enter the observed frequencies (the actual customer counts gathered by the researcher) into list L1 and the calculated expected frequencies into list L2.

For our specific example, the observed counts (50, 60, 40, 47, 53) must be entered sequentially into L1. Correspondingly, the expected counts (50, 50, 50, 50, 50) must be entered into L2. It is absolutely vital to ensure that the lists are perfectly aligned row-by-row; the entry in L1 representing Monday's observed traffic must be directly paired with the entry in L2 representing Monday's expected traffic. Misalignment will result in a meaningless test statistic.

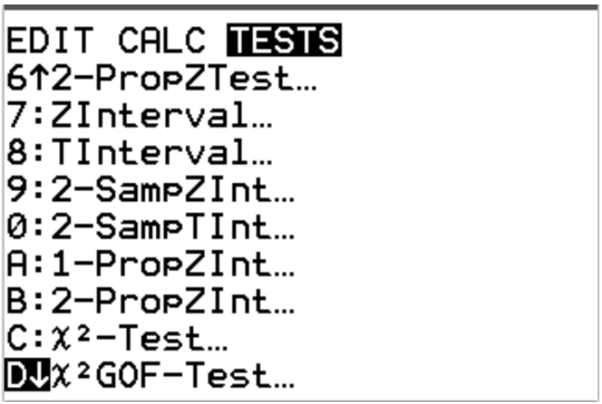
Upon completion of the data entry, the calculator screen should clearly display the raw values in the structured list format, confirming that both the observed and expected values have been correctly stored for the subsequent calculations. As noted earlier, failure to maintain alignment between these two lists will guarantee an erroneous result when executing the **Chi-Square Goodness of Fit Test**.



Step 2: Executing the X²GOF-Test

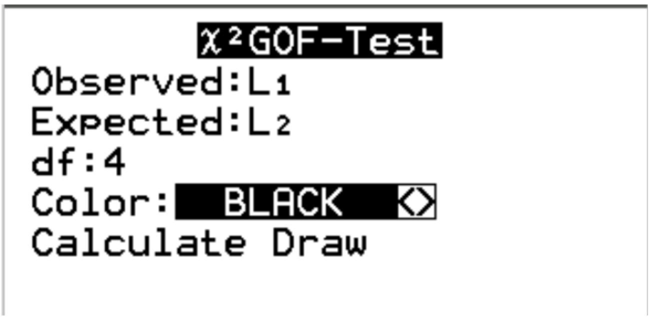
With the observed data (L1) and expected data (L2) correctly stored, the next step is to instruct the TI-84 calculator to perform the statistical calculation. All hypothesis tests are conveniently located within the STAT menu, making the execution procedure highly standardized.

Press the Stat button once more, and then scroll horizontally using the arrow keys to the **TESTS** menu. Navigate down through the comprehensive list of available hypothesis tests until the **X²GOF-Test** (Chi-Square Goodness of Fit Test) option is visible. Select this option and press Enter.



The calculator interface will then prompt the user for three essential parameters: the list containing the **Observed** frequencies (L1), the list containing the **Expected** frequencies (L2), and the required degrees of freedom (df). The TI-84 typically defaults to L1 and L2, which matches our established input structure. Crucially, the calculation for degrees of freedom in the Goodness of Fit Test is determined by the number of categories (k) minus one (df = k - 1). In our specific customer traffic example, we have 5 categories (the five weekdays), resulting in df = 5 - 1 = 4. Enter the value '4' for df.

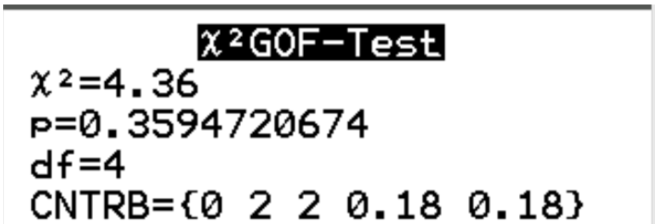
Once all parameters are correctly specified, scroll down to highlight the **Calculate** option and press Enter. The TI-84 calculator will instantaneously process the lists, compute the component differences, sum them to obtain the X^2 statistic, and display the final output screen detailing the complete test results. This automated process significantly simplifies the analysis, removing the need for tedious manual computation.



Interpreting the Calculated Results and Making a Decision

The final output screen produced by the TI-84 calculator contains all the necessary figures for making a statistical decision regarding the null hypothesis. Understanding these core output values--the test statistic (X^2) and the corresponding P-value--is the most critical phase of the hypothesis testing procedure.

For the analysis of the customer traffic distribution, the TI-84 output displays the following results:



The calculated X^2 test statistic is **4.36**, and the corresponding P-value is approximately **0.3595**. The X^2 statistic itself quantifies the magnitude of the difference between what was observed and what was expected; a higher X^2 value indicates a poorer fit between the sample data and the theoretical model. However, the definitive choice to reject or fail to reject the null hypothesis hinges almost entirely on the P-value.

The standard decision rule requires comparing the P-value to a pre-determined threshold known as the significance level (alpha, often set at 0.05). If the P-value is smaller than alpha ($p < 0.05$), we reject the null hypothesis, concluding that the observed distribution is statistically and significantly different from the expected distribution. Conversely, if the P-value is greater than or equal to alpha ($p \geq 0.05$), we fail to reject the null hypothesis, indicating that the observed differences are not large enough to dispute the theoretical claim.

In our example, 0.3595 is substantially greater than the conventional alpha of 0.05. Consequently, we must **fail to reject the null hypothesis**. This crucial statistical outcome implies that even though the observed counts (50, 60, 40, etc.) were not precisely equal to the expected 50 customers per day, the observed deviations are well within the range of what could reasonably occur due to random sampling variability. We conclude that there is insufficient statistical evidence to dispute the shop owner's claim of a uniform customer distribution across the weekdays.

Summary and Final Conclusion

The **Chi-Square Goodness of Fit Test** stands as an invaluable diagnostic tool used for validating assumptions regarding population distributions based on sample data. By skillfully utilizing the automated computational power of the TI-84 calculator, the complex process of comparing observed frequencies against expected theoretical frequencies becomes both rapid and highly reliable, provided that the underlying statistical assumptions are meticulously upheld.

Our analysis of the shop owner's customer traffic data clearly demonstrated that the observed patterns of daily arrivals are statistically consistent with a uniform distribution across the work week. The relatively high P-value of 0.3595 provided strong evidence that the minor fluctuations from the expected 50 customers per day are not statistically significant when measured against the 5% level of significance.

For the shop owner, the practical implication is that based on the evidence collected during the sampled week, there is no statistical justification to implement operational changes--such as staffing adjustments or targeted marketing--to account for widely varying customer traffic between Monday and Friday. The distribution of customers can confidently be treated as essentially uniform. This tutorial provides a robust, repeatable framework for successfully conducting the Chi-Square Goodness of Fit Test. While the TI-84 calculator handles the computation, the critical intellectual components--defining the hypotheses, verifying assumptions, calculating the degrees of freedom, and interpreting the P-value--must always remain the focused responsibility of the analyst.