Understanding the G-Test of Goodness of Fit: Definition and Practical Example

Authored by

Mohammed looti

November 5, 2025

In the expansive field of **statistics**, one of the most fundamental tasks is rigorously determining whether observed experimental or sampled data aligns with established theoretical expectations. The **G-test of Goodness of Fit** stands out as an exceptionally powerful and versatile statistical instrument specifically engineered for this assessment. It is primarily used to evaluate if the frequency distribution of a **categorical variable** successfully adheres to a particular predetermined or **hypothesized distribution**.

This test is technically classified as a likelihood-ratio test, and it serves as a robust and often preferred alternative to the more traditional Chi-squared test, particularly in contemporary statistical modeling environments. The G-test offers notable advantages when researchers are handling exceptionally large datasets, or in situations where potential data outliers might introduce unnecessary skew or bias into the results derived from alternative statistical methods. Its operational mechanism involves comparing the ratio of the actual observed frequencies against the expected theoretical frequencies, leveraging the mathematical power of the natural logarithm to compute its core statistic.

To establish a solid foundation for utilizing the G-test, the analysis must begin by clearly defining the competing claims regarding the underlying population distribution. These claims are universally formalized using the conventional statistical structure of the **null and alternative hypotheses**:

H0 (Null Hypothesis): The distribution of the observed variable precisely follows the specific hypothesized distribution (i.e., there is no significant difference).
HA (Alternative Hypothesis): The distribution of the observed variable *does not* follow the hypothesized distribution (i.e., a statistically significant difference exists).

G-test vs. Chi-Squared Test: Theoretical Foundations and Practical Benefits

Although the G-test and the Chi-squared test frequently produce numerically proximate results, especially when applied to sufficiently large sample sizes, their fundamental theoretical underpinnings are distinctly different. The Chi-squared test quantifies discrepancies based on the squared differences between the observed and expected counts, a method rooted in classical statistics. Conversely, the **G-test** is conceptually grounded in information theory and the principle of maximum likelihood estimation, providing a more modern statistical perspective.

The foremost operational advantage of the **G-test** stems from its remarkable additive properties. In the context of complex experimental designs, such as those that involve multi-way contingency tables or intricate nested models, the derived G statistics can be cleanly summed and partitioned. This crucial characteristic allows researchers to break down the sources of variance and non-fit more effectively than is possible with standard Chi-squared statistics. This additivity makes the G-test the preferred methodological choice in advanced scientific fields, including molecular biology and ecology, when analyzing highly complex frequency data structures.

Consequently, the G-test is strongly recommended in virtually all analytical scenarios where the theoretical basis of the statistical inference relies heavily on likelihood ratios. It is imperative to remember, however, that when dealing with very small sample sizes--specifically, when the expected counts (E) drop below 5 in any individual cell--neither the G-test nor the Chi-squared test is entirely reliable without applying appropriate corrective measures. Depending on the specific data structure, statisticians often turn to methods such as Yates' correction or Fisher's exact test to maintain the validity of the statistical conclusion.

Calculating the G Statistic: The Core Formula

The statistical procedure necessary for executing the G-test fundamentally relies on calculating the G statistic itself. This resulting value serves as the primary quantitative measure of the discrepancy existing between the observed empirical data and the data expected under the assumption of the Null Hypothesis (H0). The formula for this test statistic is designed such that its distribution approximates a Chi-squared distribution, enabling the subsequent calculation of the critical value and p-value. The defining equation is:

G = 2 * Σ

To successfully apply this formula, two essential components must be accurately derived directly from the sample data. These components are systematically summed (represented by the symbol Σ) across every single category or cell within the distribution being analyzed:

O (Observed Count): This is the actual frequency or count recorded empirically within a specific cell or category of the dataset.
E (Expected Count): This represents the theoretical frequency or count expected in that same cell, calculated strictly under the assumption that the Null Hypothesis holds perfectly true.

The process of multiplying the observed count (O) by the natural logarithm (ln) of the ratio between observed and expected counts (O/E) is central to the test's theoretical power. This operation rigorously quantifies the "information loss" or the divergence incurred when the simplified expected model is used as a substitute for the complexity of the actual observed data distribution. The final, aggregated G value represents the critical statistic used for all subsequent statistical inference and hypothesis testing.

Essential Assumptions and Degrees of Freedom

For any statistical test, the reliability of the conclusions drawn depends entirely on meeting certain underlying assumptions, and the G-test is no exception. The most critical assumptions ensure that the calculated G statistic accurately conforms to the theoretical Chi-squared distribution, which is the necessary reference used for determining the p-value.

Independence of Observations: It is mandatory that every single data point recorded within the sample must be statistically independent of all others. For instance, if conducting a behavioral survey, the response provided by one participant must not exert any influence or bias on the response recorded for any other participant.
Random Sampling: The data must be collected exclusively through a process of rigorous random sampling drawn from the entire population of interest. This critical step minimizes systematic bias and ensures that the sample is truly representative of the population parameters being tested.
Adequate Sample Size: While the G-test is notably robust for large datasets, the approximation to the Chi-squared distribution remains valid only if the **expected frequency (E)** in every single cell is ideally greater than 5. Failure to meet this requirement often necessitates the use of exact tests or pooling categories.

The correct interpretation of the resulting G statistic is inextricably linked to the concept of **Degrees of Freedom (df)**. The degrees of freedom are essential because they dictate the precise shape of the reference Chi-squared distribution curve used to locate the critical value or calculate the p-value corresponding to the G statistic.

For a standard G-test of Goodness of Fit, the degrees of freedom are determined by a straightforward calculation: **df = k - 1**. In this equation, *k* represents the total number of distinct categories or cells that are being subjected to analysis. This calculation accounts for the inherent statistical constraint: once the total sample size is fixed, only k-1 categories are free to vary, as the final expected count is constrained by the requirement that the sum of all expected counts must precisely match the sum of all observed counts.

Practical Example: Testing Uniform Species Distribution

To fully appreciate the utility and mechanics of the G-test, we can examine a practical scenario concerning biological distribution. Imagine a biologist proposes a hypothesis asserting that three distinct species of turtles (labeled A, B, and C) exist in perfectly equal proportions within a specific wetland habitat. To rigorously test this claim, an independent researcher conducts a comprehensive census and records the following **observed counts (O)**:

Species A: **80** individuals
Species B: **125** individuals
Species C: **95** individuals

The total census count amounts to 300 turtles. The researcher must now methodically employ the G-test of Goodness of Fit to determine whether this collected empirical data is statistically consistent with the biologist's initial hypothesis of a perfectly uniform distribution across the three species.

Step 1: Formalizing the Hypotheses. The researcher must first translate the claim into the formal framework of statistical testing:

H0: An equal proportion of the three species exists (PA = PB = PC = 1/3).
HA: The proportion of the three species *is not* equal; the distribution deviates significantly from uniformity.

Step 2: Determining the Expected Counts (E). Under the strict assumption of the Null Hypothesis (H0)--that is, equal proportions--the expected count (E) for each species is calculated by dividing the total count by the number of categories (k). Given a total count of 300 and three categories: E = 300 / 3 = 100. Therefore, the expected count (E) is 100 for Species A, Species B, and Species C.

Calculation and Interpretation of Results

Step 3: Calculating the G Test Statistic. Utilizing the established formula, **G = 2 *** Σ, we calculate the contribution to the G statistic from each of the three species categories:

G = 2 *

Executing these calculations precisely yields the final test statistic:

G = 2 * = **10.337**

Step 4: Interpreting the G Statistic. With the test statistic G = 10.337 now calculated, the conclusive step requires comparing this value against the Chi-squared distribution, which allows us to determine the corresponding p-value. First, we must accurately establish the degrees of freedom (df), calculated as k - 1, resulting in 3 - 1 = 2 degrees of freedom.

Consulting the Chi-squared distribution table or using statistical software reveals that the p-value associated with a G statistic of 10.337, given 2 degrees of freedom, is precisely determined to be **0.005693**. If we employ the standard scientific significance level (α) of 0.05, we compare the calculated p-value to this predefined threshold. Since the calculated p-value (0.005693) is markedly smaller than the threshold of 0.05, the researcher is compelled by the evidence to **reject the Null Hypothesis**.

The decisive rejection of H0 provides the researcher with substantial statistical evidence to firmly conclude that an equal proportion of each turtle species does *not* exist in this wetland area. The observed distribution of species exhibits a statistically significant deviation from the originally hypothesized uniform distribution.

Implementing the G-test in R Statistical Software

For research projects involving extensive datasets or requiring rapid, repeatable analysis, leveraging specialized statistical software is essential. The R programming environment provides excellent tools for this purpose, specifically the **GTest()** function available within the DescTools package, which allows for the swift execution of the **G-test of Goodness of Fit**, perfectly replicating the complex manual calculations demonstrated above.

The following R code snippet illustrates how to effectively perform the G-test using the previous turtle species example. We input the vector of observed counts (x) and the vector of expected proportions (p), ensuring the parameters are correctly specified:

```
#load the DescTools library
library(DescTools)

#perform the G-test
GTest(x = c(80, 125, 95), #observed values
p = c(1/3, 1/3, 1/3), #expected proportions
correct = "none")

Log likelihood ratio (G-test) goodness of fit test

data: c(80, 125, 95)
G = 10.337, X-squared df = 2, p-value = 0.005693
```

The output generated by the R function confirms the precision of the manual calculation: the G test statistic is reported as **10.337**, and the corresponding p-value is definitively **0.005693**. Because this p-value falls well below the standard 0.05 significance level, the software's result autonomously leads to the recommendation to reject the null hypothesis.

Conclusion and Further Resources for Statistical Testing

The G-test of Goodness of Fit provides a mathematically rigorous and often superior method for assessing whether observed frequency data conforms to a theoretical model. Its foundation in likelihood ratios and its advantageous additive properties make it indispensable for modern statistical analysis, particularly when dealing with complex or large-scale ecological and biological data.

For researchers and students who need immediate calculation capabilities but lack access to dedicated statistical environments like R, utilizing online tools can significantly accelerate the workflow. We recommend using a reliable online G-test calculator to automatically perform this

analysis for various datasets, ensuring quick confirmation of results and facilitating efficient hypothesis testing.