

# What is the G-test of Goodness of Fit and can you provide an example?

Authored by  
**stats writer**

April 27, 2024

## RECOMMENDED CITATION

stats writer (2024). *What is the G-test of Goodness of Fit and can you provide an example?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=139909>

The G-test of Goodness of Fit is a statistical test used to determine if there is a significant difference between the observed and expected frequencies of categorical data. It is commonly used to assess how well a set of data fits a particular theoretical distribution or expected values. The G-test calculates a test statistic known as G, which is then compared to a critical value based on the degree of freedom and desired significance level. If the calculated G value is greater than the critical value, it indicates that there is a significant difference between the observed and expected frequencies, and the null hypothesis can be rejected. An example of this test would be to determine if the distribution of hair color in a population follows the expected ratio of 3:1 for black to blonde hair. By comparing the observed frequencies to the expected frequencies, the G-test can determine if there is a significant deviation from the expected distribution.

## G-test of Goodness of Fit: Definition + Example

In statistics, the G-test of Goodness of Fit is used to determine whether or not some categorical variable follows a hypothesized distribution.

This test is an alternative to the chi-square test and is often used when outliers are present in the data or when the data you're working with is extremely large.

The G-Test of Goodness of Fit uses the following null and alternative hypotheses:

**H<sub>0</sub>:** A variable follows a hypothesized distribution.  
**H<sub>A</sub>:** A variable *does not* follow a hypothesized distribution.

The test statistic is calculated as follows:

$$G = 2 * \sum$$

**where:**

**O:** The observed count in a cell  
**E:** The expected count in a cell

If the p-value that corresponds to the test statistic is less than some  $\alpha$ , then you can reject the null hypothesis and conclude that the variable under study does not follow the hypothesized distribution.

The following example shows how to perform a G-test of Goodness of Fit in practice.

**Example: G-test of Goodness of Fit**

A biologist claims that an equal proportion of three species of turtles exist in a certain area. To test this claim, an independent researcher counts the number of each type of species and finds the following:

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

The independent researcher can use the following steps to perform a G-test of Goodness of Fit to determine if the data she collected is consistent with the biologist's claim.

## Step 1: State the null and alternative hypotheses.

The researcher will perform the G-test of Goodness of Fit using the following hypotheses:

**H0:** An equal proportion of three species of turtles exist in this area.  
**HA:** An equal proportion of three species of turtles *does not* exist in this area.

The formula to calculate the test statistic is as follows:

$$G = 2 * \sum$$

In this example, there are 300 total observed turtles. If there was an equal proportion of each species, we would expect to observe 100 turtles from each species. Thus, we can calculate the test statistic as:

$$G = 2 * = 10.337$$

## Step 3: Calculate the p-value of the test statistic.

According to the , the p-value associated with a test statistic of 10.337 and #categories-1 = 3-1 = 2 degrees of freedom is 0.005693.

Since this p-value is less than .05 the researcher would

**reject the null hypothesis. This means she has sufficient evidence to say that an equal proportion of each species of turtle *does not* exist in this particular area.**

**Bonus: G-test of Goodness of Fit in R**

**You can use the `Gtest()` function from the `DescTools` package to quickly perform a G-test of Goodness of Fit in R.**

**The following code shows how to perform a G-test for the previous example:**

```
#load the DescTools librarylibrary(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
```

**Notice that the G test statistic is 10.337 and the corresponding p-value is 0.005693. Since this p-value is less than .05, we would reject the null hypothesis.**

**This matches the results that we calculated by hand.**

**Feel free to use this to automatically perform a G-test for any dataset.**

ARABPSYCHOLOGY.COM