

## 2.4: G–Test of Goodness-of-Fit

### Learning Objectives

- To study the use of  $G$ –test of goodness-of-fit (also known as the likelihood ratio test, the log-likelihood ratio test, or the  $G^2$  test) when you have one nominal variable
- To see whether the number of observations in each category fits a theoretical expectation, and the sample size is large

### When to use it

Use the  $G$ –test of goodness-of-fit when you have one nominal variable with two or more values (such as male and female, or red, pink and white flowers). You compare the observed counts of numbers of observations in each category with the expected counts, which you calculate using some kind of theoretical expectation (such as a 1 : 1 sex ratio or a 1 : 2 : 1 ratio in a genetic cross).

If the expected number of observations in any category is too small, the  $G$ –test may give inaccurate results, and you should use an exact test instead. See the web page on small sample sizes for discussion of what "small" means.

The  $G$ –test of goodness-of-fit is an alternative to the chi-square test of goodness-of-fit; each of these tests has some advantages and some disadvantages, and the results of the two tests are usually very similar. You should read the section on "Chi-square vs.  $G$ –test" near the bottom of this page, pick either chi-square or  $G$ –test, then stick with that choice for the rest of your life. Much of the information and examples on this page are the same as on the chi-square test page, so once you've decided which test is better for you, you only need to read one.

### Null hypothesis

The statistical null hypothesis is that the number of observations in each category is equal to that predicted by a biological theory, and the alternative hypothesis is that the observed numbers are different from the expected. The null hypothesis is usually an extrinsic hypothesis, where you know the expected proportions before doing the experiment. Examples include a 1 : 1 sex ratio or a 1 : 2 : 1 ratio in a genetic cross. Another example would be looking at an area of shore that had 59% of the area covered in sand, 28% mud and 13% rocks; if you were investigating where seagulls like to stand, your null hypothesis would be that 59% of the seagulls were standing on sand, 28% on mud and 13% on rocks.

In some situations, you have an intrinsic hypothesis. This is a null hypothesis where you calculate the expected proportions after the experiment is done, using some of the information from the data. The best-known example of an intrinsic hypothesis is the Hardy-Weinberg proportions of population genetics: if the frequency of one allele in a population is  $p$  and the other allele is  $q$ , the null hypothesis is that expected frequencies of the three genotypes are  $p^2$ ,  $2pq$ , and  $q^2$ . This is an intrinsic hypothesis, because you estimate  $p$  and  $q$  from the data after you collect the data, you can't predict  $p$  and  $q$  before the experiment.

### How the test works

Unlike the exact test of goodness-of-fit, the  $G$ –test does not directly calculate the probability of obtaining the observed results or something more extreme. Instead, like almost all statistical tests, the  $G$ –test has an intermediate step; it uses the data to calculate a test statistic that measures how far the observed data are from the null expectation. You then use a mathematical relationship, in this case the chi-square distribution, to estimate the probability of obtaining that value of the test statistic.

The  $G$ –test uses the log of the ratio of two likelihoods as the test statistic, which is why it is also called a likelihood ratio test or log-likelihood ratio test. (Likelihood is another word for probability.) To give an example, let's say your null hypothesis is a 3 : 1 ratio of smooth wings to wrinkled wings in offspring from a bunch of *Drosophila* crosses. You observe 770 flies with smooth wings and 230 flies with wrinkled wings. Using the binomial equation, you can calculate the likelihood of obtaining exactly 770 smooth-winged flies, if the null hypothesis is true that 75% of the flies should have smooth wings ( $L_{null}$ ); it is 0.01011. You can also calculate the likelihood of obtaining exactly 770 smooth-winged flies if the alternative hypothesis that 77% of the flies should have smooth wings ( $L_{alt}$ ); it is 0.02997. This alternative hypothesis is that the true proportion of smooth-winged flies is exactly equal to what you observed in the experiment, so the likelihood under the alternative hypothesis will be higher than for the null hypothesis. To get the test statistic, you start with  $L_{null}/L_{alt}$ ; this ratio will get smaller as  $L_{null}$  gets smaller, which will happen as the observed results get further from the null expectation. Taking the natural log of this likelihood ratio, and multiplying it by  $-2$ , gives the log-likelihood ratio, or  $G$ -statistic. It gets bigger as the observed data get further from the null expectation. For the fly example, the test statistic is  $G = 2.17$ . If you had observed 760 smooth-winged flies and 240 wrinkled-wing flies, which is closer

to the null hypothesis, your  $G$ -value would have been smaller, at 0.54; if you'd observed 800 smooth-winged and 200 wrinkled-wing flies, which is further from the null hypothesis, your  $G$ -value would have been 14.00.

You multiply the log-likelihood ratio by  $-2$  because that makes it approximately fit the chi-square distribution. This means that once you know the  $G$ -statistic and the number of degrees of freedom, you can calculate the probability of getting that value of  $G$  using the chi-square distribution. The number of degrees of freedom is the number of categories minus one, so for our example (with two categories, smooth and wrinkled) there is one degree of freedom. Using the **CHIDIST** function in a spreadsheet, you enter **=CHIDIST(2.17, 1)** and calculate that the probability of getting a  $G$ -value of 2.17 with one degree of freedom is  $P = 0.140$ .

Directly calculating each likelihood can be computationally difficult if the sample size is very large. Fortunately, when you take the ratio of two likelihoods, a bunch of stuff divides out and the function becomes much simpler: you calculate the  $G$ -statistic by taking an observed number ( $O$ ), dividing it by the expected number ( $E$ ), then taking the natural log of this ratio. You do this for the observed number in each category. Multiply each log by the observed number, sum these products and multiply by 2. The equation is:

$$G = 2 \sum \left[ O \times \ln \left( \frac{O}{E} \right) \right] \quad (2.4.1)$$

The shape of the chi-square distribution depends on the number of degrees of freedom. For an extrinsic null hypothesis (the much more common situation, where you know the proportions predicted by the null hypothesis before collecting the data), the number of degrees of freedom is simply the number of values of the variable, minus one. Thus if you are testing a null hypothesis of a 1 : 1 sex ratio, there are two possible values (male and female), and therefore one degree of freedom. This is because once you know how many of the total are females (a number which is "free" to vary from 0 to the sample size), the number of males is determined. If there are three values of the variable (such as red, pink, and white), there are two degrees of freedom, and so on.

An intrinsic null hypothesis is one where you estimate one or more parameters from the data in order to get the numbers for your null hypothesis. As described above, one example is Hardy-Weinberg proportions. For an intrinsic null hypothesis, the number of degrees of freedom is calculated by taking the number of values of the variable, subtracting 1 for each parameter estimated from the data, then subtracting 1 more. Thus for Hardy-Weinberg proportions with two alleles and three genotypes, there are three values of the variable (the three genotypes); you subtract one for the parameter estimated from the data (the allele frequency,  $p$ ); and then you subtract one more, yielding one degree of freedom. There are other statistical issues involved in testing fit to Hardy-Weinberg expectations, so if you need to do this, see Engels (2009) and the older references he cites.

---

This page titled [2.4: G-Test of Goodness-of-Fit](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [John H. McDonald](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.