

## 14.3: Testing Goodness of Fit with Chi-Squared

### Testing a Goodness of Fit Hypothesis

This test is used to see whether counts of subgroups or categories are approximately as expected (based on population data, underlying theory, and/or hypothesis) or are significantly different than expected. When the counts observed in the data are similar to the expected counts, the expected counts are said to be a good fit to the data. When the observed counts are significantly dissimilar to the expected counts, the expected counts are a poor fit to the data.

### Determining Expected Counts

Expected counts are set based on either what is known or expected to be true about populations. Let's review the origins for each of these possible ways of setting expected counts. When a sample is being compared to a known population, the population statistics proportions or counts are used to set expectations. When this is done, chi-squared is often being used to see if a sample is an appropriate match to, or representative of, the population from which it was drawn. This can be very useful if you want to establish that a sample represents a population on key demographics such as age, gender, ethnicity, and/or socioeconomic status before using them to test other hypotheses about the population. If the sample is not significantly different from the population on any of the demographic variables, they may be said to be a good representation of the population.

In this same way, a chi-squared goodness of fit can be used to test whether a sample is significantly different from a population or another sample when this may be relevant. For example, it can be used to see whether the counts of different ethnic identities in a sample of students is similar to (or significantly different from) the proportions of those identities in the rest of the student body they are meant to represent. If the counts are similar, it may be determined that the sample is a good representation of the sample on this specific demographic. If the counts are significantly different from the population of students, however, it may be determined that the sample is biased and/or that generalizations should be limited.

A chi-squared goodness of fit can also be used to see if a change has occurred in population. For example, suppose that a company does quality control testing each year and in the prior year, 80% of the products passed the quality control test but 20% failed and needed to be discarded. Suppose that since then the company has made substantial changes to their process to try to reduce quality issues. Suppose that after these changes, they do a sample quality control test of 100 products and find that 96 of them (equivalent to 96% of the sample) passed the quality control test and 4 (equivalent to 4%) fail and need to be discarded. The company could test whether this is a significant improvement over what would be expected before they made those changes to their process. In this example, a significant result would be desirable as it would be taken to indicate that the changes to their process resulted in significantly fewer items needing to be discarded due to quality issues.

The other way expected counts can be set is based on something theorized, hypothesized, or otherwise desired. For example, suppose that a company is launching a new drink and want to know which of two versions to release. Based on the success of other companies, they believe a sweeter version of the drink will be preferred and sell better than a less sweet, healthier version. Suppose they wanted to know whether the number of customers who preferred the sweet version would be different than the number of customers who preferred the healthier version. A chi-squared goodness of fit could be used to test whether the counts of those who preferred each version were even, indicating no clear preference for either version, or significantly uneven, indicating a preference for one version over the other. Knowing whether there is a significant preference for one version over the other can be very valuable information for the company when making their decision about which version to release.

### Computing Expected Counts from Percents

When expected counts are based on percentages, those percentages need to be converted to counts before they can be used in the goodness of fit formula. To convert from a percentage to an expected count, the following formula is used:

$$f_e = pn$$

Where:

$f_e$  stands for frequency (i.e. count) expected for a category

$p$  stands for the proportion or percent expected for a category in decimal form

$n$  stands for the sample size for the data set

For example, suppose there was a sample of 40 adults, of which 40% were expected to have children and 60% were expected to not have children. The expected counts for each subgroup would be calculated as follows:

Has Children (subgroup 1):

$$f_{e1} = .40(40) = 16$$

Does Not Have Children (subgroup 2):

$$f_{e2} = .60(40) = 24$$

Thus, the expected count for those who had children would be 16 persons and the expected count for those who did not have children would be 24 persons. Note that when you sum the expected counts for all groups, it will be equal to the total sample size.

### Determining Observed Counts

Observed counts are based on sample data. Each category for a variable is identified and the participants in each of those categories are simply counted to get the observed counts. These are called observed counts because data are examples of what has been observed. The symbol used for frequency observed is  $f_o$ . Numbers or names can be added to the subscript to differentiate subgroups. For example, if there were two subgroups,  $f_{o1}$  could be used to refer to the observed counts for subgroup 1 and  $f_{o2}$  could be used to refer to the observed counts of subgroup 2.

### The Goodness of Fit Formula

The chi-squared goodness of fit formula is fairly simple and is as follows:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

The steps to using this formula are as follows:

1. Find the difference between  $f_o$  (frequency observed in the data) and  $f_e$  (the frequency expected) for each category (i.e. subgroup).
2. Square the difference for each category.
3. Divide the squared difference by  $f_e$  for each category.
4. Sum the results of step 3 to get the  $\chi^2$  value.

Thus, the  $\frac{(f_o - f_e)^2}{f_e}$  portion must be computed for each category or subgroup before these values are summed. If there are two categories,  $\frac{(f_o - f_e)^2}{f_e}$  is computed twice, once for each category, before their results are summed. If there are three categories,  $\frac{(f_o - f_e)^2}{f_e}$  is computed three times, once for each category, before their results are summed, and so on.

### Example Using the Goodness of Fit Formula

Let's try using the formula to test whether the counts of sample members who did and did not have children was approximately equal to 40% and 60%, respectively. Suppose that in the sample of 40 adults, 22 did not have children and 18 did. These would be used for the observed counts. The data can be organized as follows:

Counts for Goodness of Fit Test

Subgroups	Observed	Expected
Has Children	18	16
Does Not Have Children	22	24

There are two ways you can organize and compute the  $\chi^2$  using the goodness of fit formula: table version and formula version. I really like the table format for organizing and conducting computations but some find the formula version easier to understand. The computations and steps are the same in both versions. The only difference is how the information is laid out. Thus, both versions are equally useful and are shown here so you can choose which the format that is the clearest to you.

Table format builds from the summary counts table above to organize and complete the chi squared formula steps. The observed and expected counts are the preparatory work which are then used in the steps of the chi-squared goodness of fit formula. For the above data, the table format of steps and results are as follows:

Table 14.1 Computations for Goodness of Fit in Table Format

Preparation			Steps		
Subgroups	Observed	Expected	Differences $f_o - f_e$	Squared $(f_o - f_e)^2$	Divided $\frac{(f_o - f_e)^2}{f_e}$
Has Children	18	16	2	4	4/16 = 0.25
Does Not Have Children	22	24	(-2)	4	4/24 = 0.1667
Total	<u>Summed</u> 40				<u>Summed</u> $\chi^2 = 0.4167$

The result when rounded to the hundredths place is:  $\chi^2 = 0.42$ .

The same preparation and steps can be used in formula format. For this format, the parts of the formula are filled in and computed using order of operations. Here is what it looks like to compute the  $\chi^2$  in formula format:

Table 14.2 Computations for Goodness of Fit in Formula Format

Has Children				Does Not Have Children
$\chi^2$	=	$\frac{(f_o - f_e)^2}{f_e}$	+	$\frac{(f_o - f_e)^2}{f_e}$
$\chi^2$	=	$\frac{(18 - 16)^2}{16}$	+	$\frac{(22 - 24)^2}{24}$
$\chi^2$	=	$\frac{(2)^2}{16}$	+	$\frac{(-2)^2}{24}$
$\chi^2$	=	$\frac{4}{16}$	+	$\frac{4}{24}$
$\chi^2$	=	0.25	+	0.1667
$\chi^2$	=		0.4167	

The result when rounded to the hundredths place is:  $\chi^2 = 0.42$ .

Notice that the steps and results are the same whether you use the table format or the formula format. This is because each is just a different way of showing the same steps and computations.

### Reading Review 14.2

1. What is an observed count based on?
2. What is an expected count based on?
3. What does  $f_o$  represent?
4. What does  $f_e$  represent?
5. What are the steps to computing  $\chi^2$  goodness of fit for two categories?

This page titled [14.3: Testing Goodness of Fit with Chi-Squared](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by .