14.4: Example of How to Test a Hypothesis Using Chi-squared Goodness of Fit

Suppose a café is preparing to offer four different flavors of latte for the fall season but they think the preferences of customers for the four flavors will be uneven. They do a test run wherein they have 120 participants try each latte and select their favorite. Let's test the hypothesis that the counts of customers who prefer each of the four different flavors will be uneven using Data Set 14.1. We will use this information to follow the steps in hypothesis testing.

Data Set 14.1. Preferred Drink Flavor.

Counts for Goodness of Fit Test	
Flavors	Observed Counts
No Flavor Added	30
Vanilla	39
Chocolate	36
Pumpkin	15

Steps in Hypothesis Testing

In order to test a hypothesis, we must follow these steps:

1. **State the hypothesis.**

A summary of the research hypothesis and corresponding null hypothesis in sentence and symbol format are shown below. However, researchers often only state the research hypothesis using a format like this: *It is hypothesized that the counts of preferences will be uneven across different drink flavors.*

Non-Directional Hypothesis for Chi-squared Goodness of Fit

Research hypothesis	Counts of preference for different drink flavors will be significantly uneven.	$H_A : f_{\text{observed}} \neq f_{\text{expected}}$
Null hypothesis	Counts of preference for different drink flavors will *not* be significantly uneven.	$H_0 : f_{\text{observed}} = f_{\text{expected}}$

2. **Choose the inferential test (formula) that best fits the hypothesis.**

The counts of categories for a qualitative variable are being tested so the appropriate test is chi-squared goodness of fit.

3. **Determine the critical value.**

In order to determine the critical value for chi-square, we need to know the alpha level and the degrees of freedom. The alpha level is often set at .05 unless there is reason to adjust it such as when multiple hypotheses are being tested in one study or when a Type I Error could be particularly problematic. The default alpha level can be used for this example because only one hypothesis is being tested and there is no clear indication that a Type I Error would be especially problematic. Thus, alpha can be set to 5%, which can be summarized as α = .05.

The degrees of freedom for chi-squared goodness of fit are computed using the following formula:

$$df = k - 1$$

Where k stands for the number of categories. In the current hypothesis there are 4 drink flavors so k = 4. Thus, the calculation for df for this example is as follows:

$$df = 4 - 1$$
$$df = 3$$

The alpha level and df are used to determine the critical value for the test. Below is the χ^2 critical values tables that fits the current hypothesis and data. Under the conditions of an alpha level of .05 and df = 3, the critical value is 7.815.

Chi-Squared Critical Values

df	α = .05	α = .01
1	3.841	6.635
2	5.991	9.210
3	7.815	11.345
4	9.488	13.277
5	11.070	15.086
6	12.592	16.812
7	14.067	18.475

The critical value represents the value which must be exceeded in order to declare a result significant. It represents the threshold of evidence needed to be confident a hypothesis is true. The obtained χ^2-value must be greater than 7.815 to be declared significant when using Data Set 14.1.

4. **Calculate the test statistic.**

In order to use a goodness of fit test, we first must find the observed and expected counts. Observed counts are based on a data set, however, expected counts must be computed. To find the expected counts, we need to know what the hypothesized proportions are. In the hypothesis, it states that the counts of the four groups will be *uneven*. This means that the null will state the counts *are even*. There are four categories being compared; if they had even counts (as stated by the null), it would mean that 25% of the sample would be in each of the four categories (because 25% is one-fourth). Now, we must find the total sample size and multiply it by 25% (which is 0.25 when written in decimal form) to find the expected counts.

Counts for Goodness of Fit Test

Flavors	Observed Counts	Expected Counts
No Flavor Added	30	.25 × 120 = 30
Vanilla	39	.25 × 120 = 30
Chocolate	36	.25 × 120 = 30
Pumpkin	15	.25 × 120 = 30
Total	120	

Notice that all the expected counts are the same. This will occur anytime we are testing whether counts are even or not. Now that we have the observed and expected counts for all categories, we can plug these values into the formula and solve. The computations for this example, shown in formula format, are as follows:

		No Flavor		Vanilla		Chocolate		Pumpkin
χ^2	=	$\frac{(f_o - f_e)^2}{f_e}$	+	$\frac{(f_o - f_e)^2}{f_e}$	+	$\frac{(f_o - f_e)^2}{f_e}$	+	$\frac{(f_o - f_e)^2}{f_e}$
χ^2	=	$\frac{(30-30)^2}{30}$	+	$\frac{(39-30)^2}{30}$	+	$\frac{(36-30)^2}{30}$	+	$\frac{(15-30)^2}{30}$
χ^2	=	$\frac{(0)^2}{30}$	+	$\frac{(9)^2}{30}$	+	$\frac{(6)^2}{30}$	+	$\frac{(-15)^2}{30}$
χ^2	=	$\frac{0}{30}$	+	$\frac{81}{30}$	+	$\frac{36}{30}$	+	$\frac{225}{30}$
χ^2	=	0.00	+	2.70	+	1.20	+	7.50
χ^2	=	11.40						

5. **Apply a decision rule and determine whether the result is significant.**

Assess whether the obtained value for χ^2 exceeds the critical value as follows:

The critical value is 7.815

The obtained χ^2-value is 11.40

The obtained χ^2-value exceeds (i.e. is greater than) the critical value, thus, the result is significant.

Keep in mind that obtained values are often rounded to the hundredths place when reported.

6. **Calculate the effect sizes and any secondary analyses.**

The chi-squared goodness of fit test is an omnibus test which can tell us whether, overall, the observed counts are different from the expected counts; it does not, however, always allow us to determine which category counts are different from their expected counts and which are not when not all counts are different than expected. Thus, post-hoc tests are sometimes desired when a chi-squared goodness of fit result is significant. When two categories are being compared, a post-hoc test is not generally used. However, when three or more category counts are being compared, various post-hoc tests (such as using a chi-squared test for independence with a Bonferroni correction to compare each pair of categories) may be desired and used. However, as the focus of this chapter is the omnibus test, we will not go into a detailed review of these secondary analyses. For our purposes, therefore, secondary analyses will not be used to test the hypothesis and the use of both the chi-squared goodness of fit test and the chi-squared test for independence will only be reviewed for use as omnibus tests.

7. **Report the results in American Psychological Associate (APA) format.**

Results for inferential tests are often best summarized using a paragraph that states the following:

a. the hypothesis and specific inferential test used,
b. the main results of the test and whether they were significant,
c. any additional results that clarify or add details about the results,
d. whether the results support or refute the hypothesis.

There are no means or standard deviations to report for chi-squared because it is non parametric. It is, however, necessary to include the observed counts for each category in the results. Finally, it is customary to reported the total sample size using "N =" to the right of the df in the parenthesis of the evidence string. Following this, the results for our hypothesis with Data Set 14.1 can be written as shown in the summary example below.

APA Formatted Summary Example

A chi-squared goodness of fit was used to test the hypothesis that the counts of preference for different drink flavors would be significantly uneven. Consistent with the hypothesis, the counts of preference for non-flavored (n = 30), vanilla (n = 39), chocolate (n = 36), and pumpkin (n = 15) were significantly uneven, $\chi^2(3, N = 120) = 11.40$, p < .05.

As always, the APA-formatted summary provides a lot of detail in a particular order. For a brief review of the structure for the APA-formatted summary of the omnibus test results, see the summary below.

Anatomy of the Evidence String

The following breaks down what each part represents in the evidence string for the chi-squared results in the APA-formatted paragraph above:

Symbol for the test	Degrees of Freedom and Total Sample Size	Obtained Value	p-Value
χ^2	$(3, N = 120)$	= 11.40	p < .05.

Reading Review 14.3

1. How is df calculated for a chi-squared goodness of fit test?
2. What is reported within the parenthesis next to df in the evidence string for chi-squared?

3. How are expected counts for each category calculated when testing whether counts are even or uneven using a chi-squared goodness of fit test?
4. What detail about each category should be included in the APA-formatted summary for a chi-squared goodness of fit test?
