

### 11.1.4: Testing for Independence in Two-Way Tables (Special Topic)

Google is constantly running experiments to test new search algorithms. For example, Google might test three algorithms using a sample of 10,000 google.com search queries. Table 6.15 shows an example of 10,000 queries split into three algorithm groups.<sup>20</sup> The group sizes were specified before the start of the experiment to be 5000 for the current algorithm and 2500 for each test algorithm.

Table 6.15: Google experiment breakdown of test subjects into three search groups.

Search algorithm	current	test 1	test 2	Total
Counts	5000	2500	2500	10000

<sup>20</sup>Google regularly runs experiments in this manner to help improve their search engine. It is entirely possible that if you perform a search and so does your friend, that you will have different search results. While the data presented in this section resemble what might be encountered in a real experiment, these data are simulated.

#### Example 11.1.4.1

What is the ultimate goal of the Google experiment? What are the null and alternative hypotheses, in regular words?

The ultimate goal is to see whether there is a difference in the performance of the algorithms. The hypotheses can be described as the following:

- $H_0$ : The algorithms each perform equally well.
- $H_A$ : The algorithms do not perform equally well.

In this experiment, the explanatory variable is the search algorithm. However, an outcome variable is also needed. This outcome variable should somehow reflect whether the search results align with the user's interests. One possible way to quantify this is to determine whether (1) the user clicked one of the links provided and did not try a new search, or (2) the user performed a related search. Under scenario (1), we might think that the user was satisfied with the search results. Under scenario (2), the search results probably were not relevant, so the user tried a second search.

Table 6.16 provides the results from the experiment. These data are very similar to the count data in Section 6.3. However, now the different combinations of two variables are binned in a two-way table. In examining these data, we want to evaluate whether there is strong evidence that at least one algorithm is performing better than the others. To do so, we apply a chi-square test to this two-way table. The ideas of this test are similar to those ideas in the one-way table case. However, degrees of freedom and expected counts are computed a little differently than before.

Table 6.16: Results of the Google search algorithm experiment.

Search algorithm	current	test 1	test 2	Total
No new search	3511	1749	1818	7078
New search	1489	751	682	2922
Total	5000	2500	2500	10000

#### What is so different about one-way tables and two-way tables?

A one-way table describes counts for each outcome in a single variable. A two-way table describes counts for combinations of outcomes for two variables. When we consider a two-way table, we often would like to know, are these variables related in any way? That is, are they dependent (versus independent)?

The hypothesis test for this Google experiment is really about assessing whether there is statistically significant evidence that the choice of the algorithm affects whether a user performs a second search. In other words, the goal is to check whether the search variable is independent of the algorithm variable.

## Expected Counts in Two-way Tables

Example 6.35 From the experiment, we estimate the proportion of users who were satisfied with their initial search (no new search) as  $\frac{7078}{10000} = 0.7078$ . If there really is no difference among the algorithms and 70.78% of people are satisfied with the search results, how many of the 5000 people in the "current algorithm" group would be expected to not perform a new search?

About 70.78% of the 5000 would be satisfied with the initial search:

$$0.7078 \times 5000 = 3539 \text{ users} \quad (11.1.4.1)$$

That is, if there was no difference between the three groups, then we would expect 3539 of the current algorithm users not to perform a new search.

### Exercise 11.1.4.1

**Exercise 6.36** Using the same rationale described in Example 6.35, about how many users in each test group would not perform a new search if the algorithms were equally helpful?<sup>21</sup>

<sup>21</sup>We would expect  $0.7078 * 2500 = 1769.5$ . It is okay that this is a fraction.

We can compute the expected number of users who would perform a new search for each group using the same strategy employed in Example 6.35 and Exercise 6.36. These expected counts were used to construct Table 6.17, which is the same as Table 6.16, except now the expected counts have been added in parentheses.

Table 6.17: The observed counts and the (Expected counts)

Search algorithm	current	test 1	test 2	Total
No new search	3511 (3539)	1749 (1769.5)	1818 (1769.5)	7078
New search	1489 (1461)	751 (730.5)	682 (730.5)	2922
Total	5000	2500	2500	10000

The examples and exercises above provided some help in computing expected counts. In general, expected counts for a two-way table may be computed using the row totals, column totals, and the table total. For instance, if there was no difference between the groups, then about 70.78% of each column should be in the first row:

$$0.7078 \times (\text{column 1 total}) = 3539 \quad (11.1.4.2)$$

$$0.7078 \times (\text{column 2 total}) = 1769.5 \quad (11.1.4.3)$$

$$0.7078 \times (\text{column 3 total}) = 1769.5 \quad (11.1.4.4)$$

Looking back to how the fraction 0.7078 was computed - as the fraction of users who did not perform a new search ( $\frac{7078}{10000}$ ) - these three expected counts could have been computed as

$$\frac{\text{row 1 total}}{\text{table total}} (\text{column 1 total}) = 3539 \quad (11.1.4.5)$$

$$\frac{\text{row 1 total}}{\text{table total}} (\text{column 2 total}) = 1769.5 \quad (11.1.4.6)$$

$$\frac{\text{row 1 total}}{\text{table total}} (\text{column 3 total}) = 1769.5 \quad (11.1.4.7)$$

This leads us to a general formula for computing expected counts in a two-way table when we would like to test whether there is strong evidence of an association between the column variable and row variable.

### Computing expected counts in a two-way table

To identify the expected count for the  $i$ th row and  $j$ th column, compute

$$\text{Expected Count}_{\text{row } i, \text{col } j} = \frac{(\text{row } i \text{ total}) \times (\text{column } j \text{ total})}{\text{table total}} \quad (11.1.4.8)$$

## The chi-square Test for Two-way Tables

The chi-square test statistic for a two-way table is found the same way it is found for a one-way table. For each table count, compute

$$\text{General formula } \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}} \quad (11.1.4.9)$$

$$\text{Row 1, Col 1 } \frac{(3511 - 3539)^2}{3539} = 0.222 \quad (11.1.4.10)$$

$$\text{Row 1, Col 2 } \frac{(1749 - 1769.5)^2}{1769.5} = 0.237 \quad (11.1.4.11)$$

$$\vdots \quad (11.1.4.12)$$

$$\text{Row 2, Col 3 } \frac{(682 - 730.5)^2}{730.5} = 3.220 \quad (11.1.4.13)$$

Adding the computed value for each cell gives the chi-square test statistic  $X^2$ :

$$X^2 = 0.222 + 0.237 + \dots + 3.220 = 6.120 \quad (11.1.4.14)$$

Just like before, this test statistic follows a chi-square distribution. However, the degrees of freedom are computed a little differently for a two-way table.<sup>22</sup> For two way tables, the degrees of freedom is equal to

$$df = (\text{number of rows minus } 1) \times (\text{number of columns minus } 1) \quad (11.1.4.15)$$

In our example, the degrees of freedom parameter is

$$df = (2 - 1) \times (3 - 1) = 2 \quad (11.1.4.16)$$

If the null hypothesis is true (i.e. the algorithms are equally useful), then the test statistic  $X^2 = 6.12$  closely follows a chi-square distribution with 2 degrees of freedom. Using this information, we can compute the p-value for the test, which is depicted in Figure 6.18.

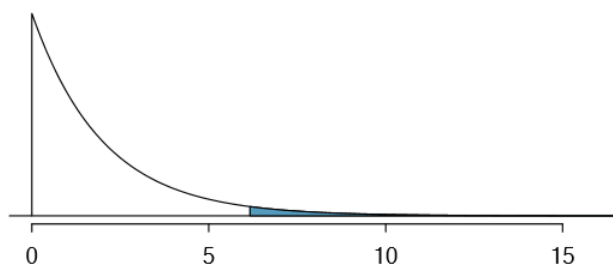


Figure 6.18: Computing the p-value for the Google hypothesis test.

### Definition: degrees of freedom for a two-way table

When applying the chi-square test to a two-way table, we use

$$df = (R - 1) \times (C - 1) \quad (11.1.4.17)$$

where R is the number of rows in the table and C is the number of columns.

<sup>22</sup>Recall: in the one-way table, the degrees of freedom was the number of cells minus 1.

Table 6.19: Pew Research poll results of a March 2012 poll.

		Congress		
	Obama	Democrats	Republicans	Total

Approve	842	736	541	2119
Disapprove	616	646	842	2104
Total	1458	1382	1383	4223

**TIP:** Use two-proportion methods for 2-by-2 contingency tables

When analyzing 2-by-2 contingency tables, use the two-proportion methods introduced in Section 6.2.

#### Example 11.1.4.1

Compute the p-value and draw a conclusion about whether the search algorithms have different performances.

##### Solution

Looking in Appendix B.3 on page 412, we examine the row corresponding to 2 degrees of freedom. The test statistic,  $X^2 = 6.120$ , falls between the fourth and fifth columns, which means the p-value is between 0.02 and 0.05. Because we typically test at a significance level of  $\alpha = 0.05$  and the p-value is less than 0.05, the null hypothesis is rejected. That is, the data provide convincing evidence that there is some difference in performance among the algorithms.

#### Example 11.1.4.1

Table 6.19 summarizes the results of a Pew Research poll.<sup>23</sup> We would like to determine if there are actually differences in the approval ratings of Barack Obama, Democrats in Congress, and Republicans in Congress. What are appropriate hypotheses for such a test?

##### Solution

- $H_0$ : There is no difference in approval ratings between the three groups.
- $H_A$ : There is some difference in approval ratings between the three groups, e.g. perhaps Obama's approval differs from Democrats in Congress.

<sup>23</sup>See the Pew Research website: [www.people-press.org/2012/03/14/romney-leads-gop-contest-trails-in-matchup-with-obama](http://www.people-press.org/2012/03/14/romney-leads-gop-contest-trails-in-matchup-with-obama). The counts in Table 6.19 are approximate.

#### Exercise 11.1.4.1

A chi-square test for a two-way table may be used to test the hypotheses in Example 6.38. As a first step, compute the expected values for each of the six table cells.<sup>24</sup>

<sup>24</sup>The expected count for row one / column one is found by multiplying the row one total (2119) and column one total (1458), then dividing by the table total (4223):  $\frac{2119 \times 1458}{4223} = 731.6$ . Similarly for the first column and the second row:  $\frac{2104 \times 1458}{4223} = 726.4$ . Column 2: 693.5 and 688.5. Column 3: 694.0 and 689.0

#### Exercise 11.1.4.1

Compute the chi-square test statistic.<sup>25</sup>

<sup>25</sup>For each cell, compute  $\frac{(\text{obs} - \text{exp})^2}{\text{exp}}$ . For instance, the first row and first column:  $\frac{(842 - 731.6)^2}{731.6} = 16.7$ . Adding the results of each cell gives the chi-square test statistic:  $X^2 = 16.7 + \dots + 34.0 = 106.4$ .

#### Exercise 11.1.4.1

Because there are 2 rows and 3 columns, the degrees of freedom for the test is  $df = (2 - 1)(3 - 1) = 2$ . Use  $X^2 = 106.4$ ,  $df = 2$ , and the chi-square table on page 412 to evaluate whether to reject the null hypothesis.<sup>26</sup>

<sup>26</sup>The test statistic is larger than the right-most column of the  $df = 2$  row of the chi-square table, meaning the p-value is less than 0.001. That is, we reject the null hypothesis because the p-value is less than 0.05, and we conclude that Americans'

approval has differences among Democrats in Congress, Republicans in Congress, and the president.

This page titled [11.1.4: Testing for Independence in Two-Way Tables \(Special Topic\)](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [David Diez, Christopher Barr, & Mine Çetinkaya-Rundel](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

- [6.4: Testing for Independence in Two-Way Tables \(Special Topic\)](#) by [David Diez, Christopher Barr, & Mine Çetinkaya-Rundel](#) is licensed [CC BY-SA 3.0](#). Original source: <https://www.openintro.org/book/os>.