

## Statistics Corner:

# Chi-square and related statistics for $2 \times 2$ contingency tables

James Dean Brown

*University of Hawai‘i at Mānoa*

### Question:

I used to think that there was only one type of chi-square measure, but more recently, I have become confused by the variety of chi-square measures that exist. Can you explain the difference between a simple chi-square and a (1) likelihood ratio chi-square, (2) a continuity adjusted chi-square, and (3) a Mantel-Haenszel chi-square? Finally, when should each of these statistics be used and what is the difference between a Yates and Pearson correction when used for chi-square data?

### Answer:

Karl Pearson first proposed what we now call chi-square in K. Pearson (1900). Generally, chi-square (also known as Pearson’s goodness of fit chi-square, chi-square test for independence, or just simply  $\chi^2$ ) is a test of the significance of how observed frequencies differ from the frequencies that would be expected to occur by chance, cleverly called expected frequencies. This test can be applied to many designs, but it is commonly explained in terms of how it applies to  $2 \times 2$  contingency tables like the one shown in Exhibit 1 (below).

In order to tackle your question in more depth, I will address the following topics: calculating simple chi-square for a  $2 \times 2$  contingency table (using an example from the literature), calculating statistics for  $2 \times 2$  contingency tables the easy way, checking the assumptions of Pearson’s chi-square, and using variations on the chi-square theme.

## Calculating simple *chi-square* for a $2 \times 2$ contingency table

In the first study of two reported in Park, Lee, and Song (2005), the authors provide an elegant report of a  $2 \times 2$  contingency table analysis that examined the frequency of whether apologies were present or absent in American and Korean email advertising messages. They describe their results as follows:

Of 234 American email advertising messages, seven contained some form of apology (e.g., “We are sorry for anything that may cause you inconvenience”), whereas 74 of 177 Korean email advertising messages contained some form of apology. A chi-square test was conducted to examine the relationship between culture and the presence of apologies. The result showed that the frequency of apologies was significantly associated with culture,  $\chi^2(1) = 95.95$ ,  $p < .01$ ,  $\phi^2 = .23$ . A greater number of Korean email advertising messages (41.81%) included apologies than did American email advertising messages (2.99%). (p. 374)

Figure 1 shows how the data need to be laid out for the calculations of the  $\chi^2$  value that Park *et al* (2005) found for the two cultures in their study (Korean and American) and the two states of Apology (present or absent).

|         |          | Apology    |            |             |
|---------|----------|------------|------------|-------------|
|         |          | Present    | Absent     |             |
| Culture | Korean   | <b>A</b>   | <b>B</b>   | Row1 Total  |
|         | American | <b>C</b>   | <b>D</b>   | Row2 Total  |
|         |          | Col1 Total | Col2 Total | Grand Total |

**Figure 1. Layout for Culture by Apology 2 x 2 contingency table**

In Figure 2, I have filled in the data from the Park et al (2005) study (the large numbers in italic-bold print) in the appropriate cells. Notice that the row sums on the right side in the first row are for  $A + B = 74 + 103 = 177$  and in the second row are for  $C + D = 7 + 227 = 234$ . Similarly, the column sums at the bottom of the first column are for  $A + C = 74 + 7 = 81$  and at the bottom of the second column are for  $B + D = 103 + 227 = 330$ . The grand total shown at the bottom right is the sum of all four cells, or  $A + B + C + D = 74 + 103 + 7 + 227 = 411$ .

Collectively, all of these values around the edges of the contingency table are known as the marginals.

|         |          | Apology   |            |     |
|---------|----------|-----------|------------|-----|
|         |          | Present   | Absent     |     |
| Culture | Korean   | <b>74</b> | <b>103</b> | 177 |
|         | American | <b>7</b>  | <b>227</b> | 234 |
|         |          | 81        | 330        | 411 |

**Figure 2. Data for Culture by Apology 2 x 2 contingency table**

*Observed frequencies* are the frequencies that were actually found in a study and put inside the cells of the contingency table. *Expected frequencies* are estimates of the frequencies that would be found by chance in such a design (based on the marginals). Table 1 shows how the expected frequencies are calculated from the marginals for each cell. For example, the expected frequency for Cell A (Korean-Present) is calculated by multiplying the column 1 marginal times the row 1 marginal and dividing the result by the grade total, or  $(\text{Col1} \times \text{Row1}) / \text{Grand Total} = (81 \times 177) / 411 = 34.882$ . The expected frequencies for cells B, C, and D are calculated similarly, as shown in Table 1. Notice that Figure 3 shows these expected frequencies in parentheses.

**Table 1. Calculating expected frequencies**

| Cell | Culture  | Apology | Observed | Calculating Expected          | (Col. × Row / Total) = | Expected |
|------|----------|---------|----------|-------------------------------|------------------------|----------|
| A    | Korean   | Present | 74       | (Col1 × Row1) / Grand Total = | (81 × 177) / 411 =     | 34.8832  |
| B    | Korean   | Absent  | 103      | (Col2 × Row1) / Grand Total = | (330 × 177) / 411 =    | 142.1168 |
| C    | American | Present | 7        | (Col1 × Row2) / Grand Total = | (81 × 234) / 411 =     | 46.1168  |
| D    | American | Absent  | 227      | (Col2 × Row2) / Grand Total = | (330 × 234) / 411 =    | 187.8832 |

|         |          | Apology                |                          |     |
|---------|----------|------------------------|--------------------------|-----|
|         |          | Present                | Absent                   |     |
| Culture | Korean   | <b>74</b><br>(34.8832) | <b>103</b><br>(142.1168) | 177 |
|         | American | <b>7</b><br>(46.1168)  | <b>227</b><br>(187.8832) | 234 |
|         |          | 81                     | 330                      | 411 |

**Figure 3. Data for Culture by Apology 2 × 2 contingency table**

Table 2 shows how the *chi-square value* ( $\chi^2$ ) is calculated for a 2 × 2 contingency table. Intermediate values are first calculated for each cell based on the observed and expected frequencies in that cell. For example, for Cell A (Korean-Present), the value is calculated by subtracting the observed frequency minus the expected frequency and squaring the result, and then dividing the squared result by the expected frequency. In this case, that would be  $(\text{Observed} - \text{Expected})^2 / \text{Expected} = (74 - 34.8832)^2 / 34.8832 = 43.8642$ . The same process is repeated for Cells B, C, and D as shown in Figure 3. Then the four results are summed and that sum is the chi-squared value. In this case, that would be  $43.8642 + 10.7667 + 33.1793 + 8.1440 = 95.9542$ , or about 95.95 as reported in Park *et al* (2005).

**Table 2. Calculating chi-square from the observed and expected frequencies**

| Culture                         | Apology | Observed | Expected | (Observed – Expected) <sup>2</sup> / Expected = |
|---------------------------------|---------|----------|----------|---|
| Korean                          | Present | 74       | 34.8832  | $(74 - 34.8832)^2 / 34.8832 = 43.8642$          |
| Korean                          | Absent  | 103      | 142.1168 | $(103 - 142.1168)^2 / 142.1168 = 10.7667$       |
| American                        | Present | 7        | 46.1168  | $(7 - 46.1168)^2 / 46.1168 = 33.1793$           |
| American                        | Absent  | 227      | 187.8832 | $(227 - 187.8832)^2 / 187.8832 = 8.1440$        |
| <b>Sum = Chi-square value =</b> |         |          |          | <b>95.9542</b>                                  |

Clearly, the chi-square statistic is not difficult to calculate (though the process is a bit tedious). It is also fairly easy to interpret. As Park *et al* (2005) put it, “The result showed that the frequency of apologies was significantly associated with culture,  $\chi^2(1) = 95.95$ ,  $p < .01$ ,  $\phi^2 = .23$ ” (p. 374). Notice that they symbolize chi-squared as  $\chi^2(1)$  [where the (1) indicates one degree of freedom] and that this chi-square value turns out to be significant at  $p < .01$ . [To determine the degrees of freedom and whether or not

chi-square is significant requires much more information than I can supply in this short column; however further explanations are readily available in Brown, 1988, pp. 182-194, or 2001, pp. 159-169, or at the SISA website referenced in the next section]. Note that the phi-square statistic ( $\phi^2$ ) that Park *et al* (2005) report will be explained below.

## Calculating statistics for $2 \times 2$ contingency tables the easy way

Now that you understand the basic calculations and interpretation of chi-square analysis for  $2 \times 2$  contingency tables, I will show you an easier way to calculate that statistic and all of the statistics mentioned in your question at the top of this column (as well as a few bonus statistics). The first trick is to go to the very handy *Simple Interactive Statistical Analysis* (or *SISA*) below and explore a bit:

<http://www.quantitativeskills.com/sisa/>

When you are ready to focus on  $2 \times 2$  contingency table analysis go to the following URL:

<http://www.quantitativeskills.com/sisa/statistics/two2by2.htm>

When you arrive at that page, you will see a screen like the one shown in Figure 4. Go ahead and fill in the values from the Park *et al* (2005)  $2 \times 2$  contingency table as shown in Figure 4. Be sure to also check the boxes next to *Show Tables:* and *Association:*, and then click on the *Calculate* button.

**SISA**

Two by two table analysis

*For help go to SISA.*

Give four positive integer numbers.  
Non decimal numbers larger than one.

|        | Col 1: | Col 2: |
|--------|--------|--------|
| Row 1: | 74     | 103    |
| Row 2: | 7      | 227    |

Width of C.I.: 95 %

Show Tables:

Association:

Calculate

**Figure 4.** Using *SISA* to calculate statistics for the Park *et al* (2005)  $2 \times 2$  contingency table

A number of tables will appear on your screen including those shown in the first column of Figure 5. You will also see numerical output like that extracted into the second column of Figure 5 (along with a good deal of additional output). Notice that the chi-squared value and its associated probability are shown in the third line of column two, labeled as *Pearson's*.

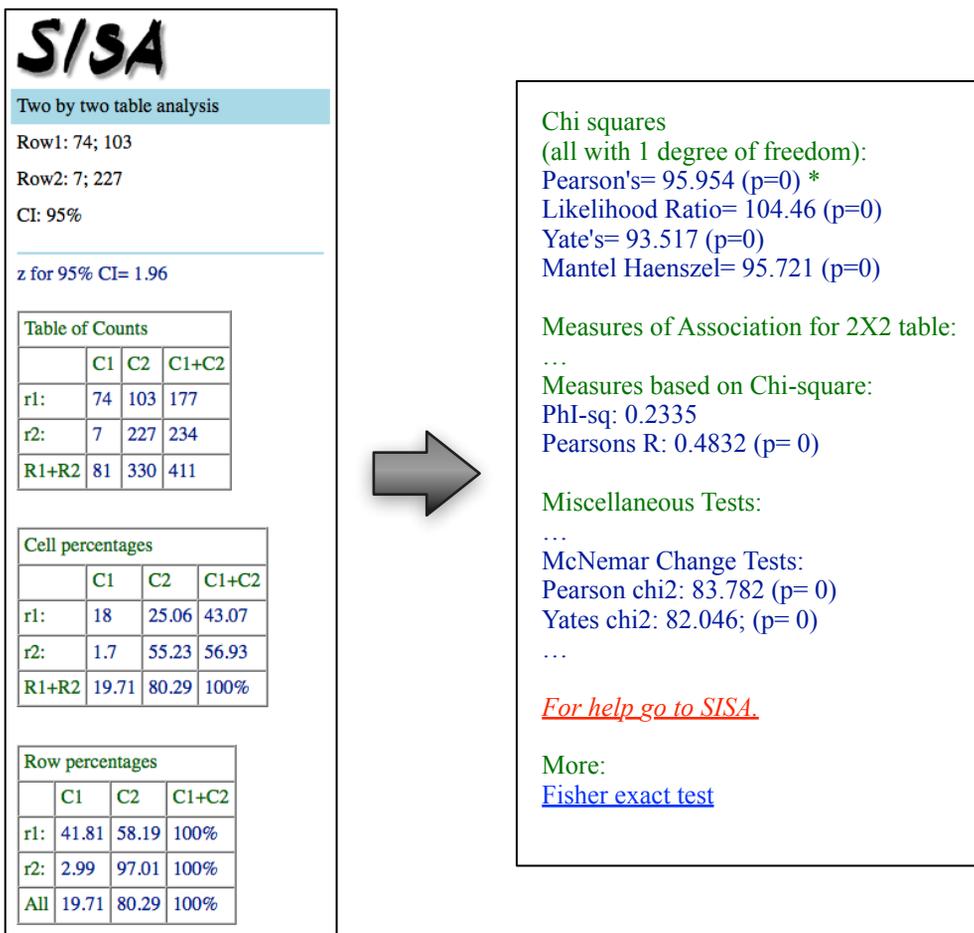


Figure 5. SISA output for the Park *et al* (2005) 2 × 2 contingency table

## Checking the assumptions of Pearson's chi-square

Pearson's chi-square for 2 × 2 contingency tables is used to analyze raw frequencies (not percentages or proportions) for two binary variables, or put more precisely, this  $\chi^2$  statistic is a reasonable test of the significance of the difference between observed and expected raw frequencies if three assumptions are met:

1. The scales are *nominal*<sup>1</sup> (i.e., they are frequencies for categorical variables)
2. Each observation is *independent* of all others
3. As a rule of thumb, the *expected frequencies* are equal to or greater than 5

For example, let's consider these assumptions in the Park *et al* (2005) study. First, the scales are clearly *nominal*: culture (American or Korean) and apology (present or absent) are definitely nominal and

<sup>1</sup> For more on nominal, ordinal, interval, and ratio scales of measurement, see Brown (2011).

binary. Second, the observations are *independent*, which means that each observation appears in one and only one cell (i.e., each advertisement is either American or Korean and has an apology present or does not, or put another way, no advertisement appears in more than one cell). Third, the smallest *expected frequency* is 34.8832, which is well over 5. So Park et al (2005) clearly met the assumptions of Pearson's chi-square. [For more on these assumptions, see Brown, 2001, pp. 168-169.]

## Using variations on the chi-square theme

Figure 5 shows selected statistics from the output that *SISA* provides. Here, I will explain the differences between these statistics, as well as when each would be appropriately applied. When the purpose of the analysis is different or the assumptions are not met, Pearson chi-square is not appropriate, but other statistics (most of which are available in the *SISA* output shown in Figure 5) have been developed for use in alternative situations as follows:

*If the scales are not nominal*, other non-parametric statistics (e.g., the *Mantel Haenszel Chi-square* is appropriate if both variables are ordinal—see Conover, 1999, pp. 192-194; Sprent & Smeeton, 2007, pp. 399-403; also see the *SISA* website) or more powerful parametric statistics may be applicable (e.g., Pearson's product-moment correlation coefficient, the *t*-test, ANOVA, regression, etc.—see Brown, 1988, 2001; Brown & Rodgers, 2002; Hatch & Lazaraton, 1991).

*If the observations are not independent*, Pearson's chi-square is not applicable. Period. This is a common violation that is ignored in second language research. Indeed, I searched for hours before finding the Park et al (2005) example that did not violate this assumption. In cases where there is a violation of this assumption, especially sequentially over time (as in a study with a dichotomous nominal variable collected from the same people on two occasions, e.g., before and after instruction), you may want to consider two other statistics: *Cochran's Q test* (see Cochran, 1950; Conover, 1999, pp. 250-258; Sprent & Smeeton, 2007, p. 215) or *McNemar's Q* (see Conover, 1999, pp. 166-170; McNemar, 1947; Sprent & Smeeton, 2007, pp. 133-135; also see *SISA* website; or to calculate this statistic: <http://vassarstats.net/propcorr.html>). In the  $2 \times 2$  case, these Cochran's Q and McNemar's Q should lead to the same result.

*If the design is larger than  $2 \times 2$* , the *likelihood ratio* (or G2) provides an alternative that can readily be used to analyze a table larger than  $2 \times 2$  and then to examine smaller components within the table in more detail (see Sprent & Smeeton, 2007, pp. 362-363; Wickens, 1989; *SISA* website).

*If an expected frequency is lower than five*, you have three alternatives: Yates correction, the Fisher exact test, or the  $N - 1$  chi-square test.

1. *Yates' correction* (Yates, 1934) is equivalent to Pearson's chi-square but with a continuity correction. In cases where an expected frequency is below 5, Yates' correction brings the result more in line with the true probability. In any case, as you can see in the second column of Figure 5, the *SISA* website will calculate this statistic for you.
2. *Fisher exact test* (Fisher, 1922) has been shown to perform accurately for  $2 \times 2$  tables with expected frequencies below 5. The Fisher exact test (aka the Fisher-Irwin test) is more difficult to calculate than Yates' correction, but given the power of our personal computers today Yates' correction can easily be replaced by the more exact Fisher exact test. Indeed, as you can see in Figure 5, you need only click on "Fisher exact test" shown in the second column for the *SISA* website to calculate this statistic for you.
3. The  $N - 1$  *chi-square test* is another option. Campbell (2007, p. 3661) compared chi-square analyses of  $2 \times 2$  tables for many different sample sizes and designs and found that a statistic

suggested by Karl Pearson's son (E. S. Pearson, 1947) called the *N - 1 chi-square test* provided the best estimates. According to Campbell, as long as the expected frequency is at least 1, this adjusted chi-square (probably the "Pearson correction" referred to in the question at the top of this column) provided the most accurate estimates of Type I error levels. However, for expected frequencies below 1, he found that Fisher's exact test performed better.

If the goal is to understand the degree of relationship between two dichotomous variables, phi-square ( $\phi^2$ ) is calculated by dividing the Pearson chi-square value by the grand total of cases. For example, in Park *et al* (2005),  $\phi^2 = \chi^2 / \text{Grand total}$ , or  $95.9542 / 411 = .2335 \approx .23$ . This statistic ranges from zero (if there is absolutely no association between the two variables) to 1.00 (if the association between the two variables is perfect). With reference to Figure 5, note that the phi square value is equal to the square of the Pearson correlation coefficient reported in Figure 5. In other words, squaring "Pearsons R" (.4832) in Exhibit 7 will yield a phi square of .2335.

## Conclusion

As the title of this column suggested, my purpose here was to explain how chi-square and related statistics can be used for analyzing  $2 \times 2$  contingency tables. To do so, I described the processes involved in calculating simple chi-square for a  $2 \times 2$  contingency table (using the Park *et al*, 2005 example from the literature), calculating statistics for  $2 \times 2$  contingency tables the easy way, checking the assumptions of Pearson's chi-square, and using variations on the chi-square theme. Along the way, I believe I addressed all parts of the original question at the top of the column.

All in all, in my experience, this family of statistics has been much abused and misused in our field—perhaps more than any other. Please consider such analyses very carefully when using them and apply them correctly. Be sure, for example, to review the correct procedures as described in some of the references listed below.

## References

- Brown, J. D. (1988). *Understanding research in second language learning: A teacher's guide to statistics and research design*. Cambridge: Cambridge University Press.
- Brown, J. D. (2001). *Using surveys in language programs*. Cambridge: Cambridge University Press.
- Brown, J. D. (2011). Statistics Corner: Questions and answers about language testing statistics: Likert items and scales of measurement? *SHIKEN: JALT Testing & Evaluation SIG Newsletter*, 15(1), 10-14. Also available on the Internet at <http://jalt.org/test/PDF/Brown34.pdf>.
- Brown, J. D., & Rodgers, T. (2002). *Doing second language research*. Oxford: Oxford University Press.
- Campbell, I. (2007). Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. *Statistics in Medicine*, 26, 3661-3675.
- Cochran W. G. (1950). The comparison of percentages in matched samples. *Biometrika*, 37, 256-266.
- Conover, W. J. (1999). *Practical nonparametric statistics* (3rd ed.). New York: Wiley.
- Fisher, R. A. (1922). On the interpretation of  $\chi^2$  from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society*, 85(1), 87-94.
- Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. Rowley, MA: Newbury House.
- McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2), 153-157.

- Park, H. S., Lee, H. E., & Song, J. A. (2005). "I am sorry to send you SPAM": Cross-cultural differences in use of apologies in email advertising in Korea and the U.S. *Human Communication Research*, 31(3), 365-398.
- Pearson, E. S. (1947). The choice of statistical tests illustrated on the interpretation of data classed in a 2 x 2 table. *Biometrika*, 34, 139-167.
- Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine, Series 5*, 50(302), 157-175.
- Sprent, P., & Smeeton, N. C. (2007). *Applied nonparametric statistical methods* (4th ed.). Boca Raton, FL: Chapman & Hall.
- Wickens, T. D. (1989). *Multiway contingency tables analysis for the social sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Yates F. (1934). Contingency tables involving small numbers and the  $\chi^2$  test. *Journal of the Royal Statistical Society Supplement*, 1, 217-235.

**Where to Submit Questions:**

Please submit questions for this column to the following e-mail or snail-mail addresses:

brownj@hawaii.edu.

JD Brown  
Department of Second Language Studies  
University of Hawai'i at Mānoa  
1890 East-West Road  
Honolulu, HI 96822  
USA

Your question can remain anonymous if you so desire.