Keith Curtis, MD, and Scott T. Youngquist, MD, MSc

Part 21: Categoric Analysis: Pearson Chi-Square Test

Investigators from St. Paul and Minneapolis, MN, sought to determine whether a didactic and manikin-based curriculum for training helicopter emergency medical service (HEMS) providers would result in differences in the successful placement of laryngeal mask airways (LMAs) versus intubating laryngeal mask airways (ILMAs) when subsequently tested on live patients in the operating room.[1] The authors assessed differences in success rates statistically using the Pearson chi-square test (herein referred to as the chi-square test).

What Is the Chi-Square Test?

The chi-square test is a test of statistical significance performed on categorical data (ie, data that can be placed into nominal [or named] categories). Let's give a very simple example; say you want to estimate the relative frequency with which red, blue, and yellow are listed as favorite colors by people in your community. (Assume these are the only colors available for the purposes of this exercise.) If you asked people going into a movie theater on a Friday night to name their favorite of these 3 available colors and tabulated the results, you would have a dataset of categorical data in which colors define the categories of interest. Let's say you summarized your data according to Table 1. (This was a 3-dimensional rerelease of a Twilight movie; thus, the poor attendance.)

Table 1 represents the observed data. Although it appears that people going to this particular movie favor the color blue, is it possible we have just measured a random fluctuation in time and space and not a true preponderance of philoceruleanism? In an attempt to answer whether this is likely a true reflection of a larger population bias toward blue, we need to start with the following thought experiment: what would the data have looked like had it perfectly represented a population in which, in fact, an equal number of people liked red as much as blue or yellow? We can reconstruct the table according to this hypothetical expectation assuming that there is no clear favoritism simply by dividing the number of color votes (N = 33) by the number of available answers (k = 3) to decide what number (33/3 = 11) should go in each cell of the table under these assumptions (Table 2).

Table 2 is the table we would expect under the previously described assumptions (absolutely no population favorite color and no random fluctuation have occurred because of under- or oversampling of any group). The chi-square test is a way to determine whether the differences between Tables 1 and 2 are statistically significant.

How Is the Chi-Square Test Performed?

This test can be calculated using a spreadsheet or calculator although there are online calculators available to do this as well. First, take any category. Let's start with red. Subtract the expected count from the observed count, square this value, and divide the result by the expected count in the following fashion: $(\text{observed} - \text{expected})^2/\text{expected} = (11 - 11)^2/11 = 0$. Now perform the same operation for blue ($[\text{observed} - \text{expected}]^2/\text{expected} = [16 - 11]^2/11 = 2.72$) and yellow ($[\text{observed} - \text{expected}]^2/\text{expected} = [6 - 11]^2/11 = 2.72$). Next, add these numbers up as follows: $(0 + 2.72 + 2.72) = 4.55$.

This number is your chi-square test statistic. However, you need one other piece of information before you can look up the *P* value associated with this number in a table of chi-square values or using an online calculator—the degrees of freedom associated with the test. The degrees of freedom, which we will not take time to explain further, is easily calculated as the number of categories, k, minus 1 (or $k - 1$). Because there are 3 colors that make up all possible categories, the degree of freedom here is $3 - 1 = 2$. In this case, a *P* value of .10 is obtained when we input the test statistic and degrees of freedom into an online statistics calculator.

How Is the Chi-Square Test Interpreted?

The *P* value associated with this test tells you the probability of obtaining the observed cell count frequencies or ones even more extreme (even more divergent from the expected cell counts), given that the null hypothesis that the expected cell counts are really the correct ones is true. With the relatively high *P* value result obtained, we cannot reliably exclude the possibility that there is no true favorite color in the population. Alternatively, we might need a larger sample size to detect such small differences using the chi-square test. As you might expect, the closer the cell counts are to expected values under the null hypothesis, the higher the *P* value and the less likely the differ-

Table 1. Responses to a Hypothetical Survey Regarding Favorite Color

Favorite color	Red	Blue	Yellow
Counts	11	16	6

Table 2. Expected Results of a Hypothetical Survey Regarding Favorite Colors Should There Be No Favorite in the Community Sampled

Favorite color	Red	Blue	Yellow
Counts	11	11	11

Chi-Squared Test Exercise

Observed Counts

	LMA	ILMA
Success	22	20

For expected counts, we assume no difference between groups and distribute all 42 successes evenly between cells.

Expected Counts

	LMA	ILMA
Success	21	21

Chi-square test statistic = $([(22 - 21)^2/22 + (20 - 21)^2/22] = 0.095$ with degrees of freedom $= 2 - 1 = 1$.
The associated *P* value is .76.

ences between observed and expected outcomes are more than just random fluctuations.

Unfortunately, with any more than 2 categories, the chi-square test does not say anything about whether blue is the favorite color (or yellow the least favorite) even if the *P* value were $<.0001$; when the *P* value is low, it simply suggests that there is not an even distribution of favorite colors in the table, as might be expected if all the categories were the same. This is a minor point but an important one about the limits of this type of significance testing. However, by simple math, the category that deviates the most from expected cell counts is the one that contributes the most to the chi-square test statistic.

Exercise

In the LMA versus ILMA study cited earlier, the successful placement of the LMA was achieved in 22 of 22 patients and in 20 of 22 patients using the ILMA. Using the example given as a model, can you calculate the chi-square test statistic and degrees of freedom? (See sidebar.)

Reference

1. Franscone RJ, Pippert G, Heegaard W, Molinari P, Dries D. Successful training of HEMS personnel in laryngeal mask airway and intubating laryngeal mask airway placement. *Air Med J.* 2008;27:185-187.

Keith Curtis, MD is senior resident in the University of Utah's Division of Emergency Medicine in Salt Lake City, UT. Scott T. Youngquist, MD, MSc, is an assistant professor in that division, director of the Air Medical Research Institute at the University of Utah, and medical director at the Salt Lake City Fire Department. He can be reached at scott.youngquist@utah.edu.

1067-991X/$36.00

http://dx.doi.org/10.1016/j.amj.2013.04.007