Measures of Association for Categorical Variables: Chi-Square

Many research questions in clinical and behavioral science involve categorical variables that are measured on a nominal or ordinal scale. These questions usually deal with the analysis of proportions or frequencies within various categories. For instance, surveys often code responses that represent frequencies, such as the number of Yes–No responses to a series of items or the number of respondents who fall into certain age groups. We can then ask questions about the proportion of respondents that fall into each category. In descriptive studies we are often interested in how certain nominal variables are distributed. For example, we might want to determine the proportion of patients with right-sided or left-sided strokes who are functionally dependent or independent at discharge or the proportion of therapists who work in private practice versus institutional settings.

These types of categorical data are analyzed by determining if there is a difference between the proportions *observed* within a set of categories and the proportions that would be *expected* by chance. For example, if therapists are equally likely to work in private or institutional settings, then theoretically we would expect an equal proportion, or 50%, to fall into each category. The null hypothesis states that no difference exists between the actual proportions measured in a sample and this theoretical distribution. If the observed data depart significantly from these expected null values, we reject the null hypothesis.

The purpose of this chapter is to describe the use of several statistics that can be used to analyze frequencies or proportions. These statistics are based on **chi-square**, χ^2, which is a nonparametric statistic used to determine if a distribution of observed frequencies differs from theoretical expected frequencies. Chi-square has many applications in clinical research, in both experimental and descriptive analysis. We concentrate on two general uses of the test. A test of goodness of fit is used to determine if a set of observed frequencies differs from a given set of theoretical frequencies that define a specific distribution. A test that compares the proportion of therapists in private and institutional settings fits this model, based on a theoretical distribution of 50 : 50. Tests of independence are used to determine if two classification variables are independent of each other, that is, to examine the degree of association between them. For example, we could study the frequency of left- and right-sided stroke in terms of functional level at discharge to determine if these variables are related or independent of each other. We also discuss the use of a related procedure called the McNemar test, for examining frequencies of correlated samples. In addition, several other coefficients of association for categorical data will be described.

THE CHI-SQUARE STATISTIC

As we discuss the different applications of the χ^2 statistic, it is important to keep in mind two general assumptions: (1) *Frequencies represent individual counts*, not ranks or percentages. This means that data in each category represent the actual number of persons, objects, or events in those categories, not a summary statistic. (2) *Categories are exhaustive and mutually exclusive*. Therefore, every subject can be assigned to an appropriate category, but only one. Repeated measurement or assignment is not appropriate; that is, no one individual should be represented in more than one category. The characteristics being measured should be defined with enough specificity to avoid any overlaps in group assignment.

Chi-square is defined by*

$$\chi^2 = \sum \frac{(O - E)^2}{E} \tag{25.1}$$

where O represents the **observed frequency** and E represents the **expected frequency**. As the difference between observed and expected frequencies increases, the value of χ^2 will increase. If observed and expected frequencies are the same, χ^2 will equal zero.

We illustrate the application of this statistic using a simple example. Suppose we tossed a coin 100 times. The null hypothesis states that no bias exists in the coin, and we would expect a theoretical outcome of 50 heads and 50 tails. We observe 47 heads and 53 tails. Does this deviation from the null hypothesis occur because the coin is biased, or is it only a matter of chance? In other words, is the difference between the observed and expected frequencies sufficiently large to justify rejection of the null hypothesis?

We calculate χ^2 by substituting values in the term $(O - E)^2/E$ for each category. For heads,

$$\frac{(O - E)^2}{E} = \frac{(47 - 50)^2}{50} = \frac{(-3)^2}{50} = 0.18$$

For tails,

$$\frac{(O - E)^2}{E} = \frac{(53 - 50)^2}{50} = \frac{(3)^2}{50} = 0.18$$

The sum of these terms for all categories is the value of χ^2. Therefore,

---

*Although the definitional formula for χ^2 (Eq. 25.1) is used most often, there is a computational formula that may be useful: $\chi^2 = \sum \frac{O^2}{E} - N$.