Chapter 14: Nonparametric Statistics and the Chi-Square Test of Independence

I. Overview

a. Most statistical procedures in this book assume that the data to be analyzed are normally distributed.

b. In the real world, distributions of data are often not normally distributed, so there is another class of statistics called *nonparametric* statistics.
   i. Do not rely on the assumption of a normal distribution.

c. There are a wide variety of nonparametric statistics. In this chapter, we focus on one: The chi-square test of independence.

II. Chi-Square Test of Independence in Depth

a. Purpose: To determine whether cases are distributed among the categories of two variables in the numbers one would expect by chance alone.

b. Types of variables: Two categorical variables
   i. e.g., gender and level of mathematics class enrolled in.

c. Procedure: A *contingency table* is produced that shows how many cases in each group on one variable fall into each of the categories on the second variable. These numbers are then compared to the number of cases one would expect to see in each category by chance alone.

   i. e.g., Suppose that in one high school, there are 100 students in eleventh grade. Half of these students are girls, the other half are boys. Half of these 100 students are taking advanced math, and half are taking basic math.

      1. Based on the number of boys and girls in the sample, and the number of students in each level of math, we would *expect* there to be 25 students in each cell of this contingency table, by chance alone. So the numbers in the table below are the *expected* values.

	Basic math	Advanced math
Girl	25	25
Boy	25	25

   ii. Now suppose that I examine the actual number of cases that appear in each cell of the table. The table below contains the *observed* number of students in each cell:

	Basic math	Advanced math
Girl	15	35

Boy	35	15

iii. As we can see, the observed values in the cells look quite different from the expected values. There are more girls and fewer boys than we would expect by chance alone in the advanced math and vice-versa in the basic math.

iv. Now we need to determine whether the difference between the observed and expected values is statistically significant.

1. To do this, we calculate the χ^2 value by subtracting the expected value from the observed value in each cell, squaring that, and dividing by the expected value for the cell. I follow this procedure for each cell and then add up these values for each cell. This produces the *observed* chi-square value. The formula is as follows:

$$\chi^2 = \sum\left(\frac{(O-E)^2}{E}\right)$$

2. Once we have calculated the χ^2 value we look in Appendix E to determine whether the value is statistically significant.

a. To use Appendix E, use $(R-1)(C-1)$ to calculate the degrees of freedom.

i. *R* represents the number of rows in the contingency table, *C* represents the number of columns.\

v. Interpret the results.

1. If the χ^2 value that you calculated is statistically significant, this tells you that the observed values in your contingency table differ from the expected values. It does not tell you exactly what is causing this difference, but you can usually get a pretty good idea by looking at the values in your contingency table.

III. Summary

a. Distributions of data are not always normally distributed.

b. There are several nonparametric statistics researchers can use when their data are not normally distributed.

c. One of these is the chi-square test of independence (χ^2).

i. This statistic is used to determine whether expected frequencies differ from observed frequencies on two categorical variables.