# Chapter Eleven

The Chi-Square Test and Measures of Association

**Summary**

• The chi-square test is an inferential statistical technique designed to test for a significant relationship between nominal or ordinal variables organized in a bivariate table. The test is conducted by testing the null hypothesis that no association exists between two cross-tabulated variables in the population, and therefore, the variables are statistically independent.

• The obtained chi-square (c2) statistic summarizes the differences between the observed frequencies (fo) and the expected frequencies (fe)—the frequencies we would have expected to see if the null hypothesis were true and the variables were not associated. The Yates’s correction for continuity is applied to all 2 ´ 2 tables.

• The sampling distribution of chi-square tells the probability of getting values of chi-square, assuming no relationship exists in the population. The shape of a particular chi-square sampling distribution depends on the number of degrees of freedom.

• Measures of association are single summarizing numbers that reflect the strength of the relationship between variables, indicate the usefulness of predicting the dependent from the independent variable, and often show the direction of the relationship.

• Proportional reduction of error (PRE) underlies the definition and interpretation of several measures of association. PRE measures are derived by comparing the errors made in predicting the dependent variable while ignoring the independent variable with errors made when making predictions that use information about the independent variable.

• Measures of association may be symmetrical or asymmetrical. When the measure is symmetrical, its value will be the same regardless of which of the two variables is considered the independent or dependent variable. In contrast, the value of asymmetrical measures of association may vary depending on which variable is considered the independent variable and which the dependent variable.

• Lambda is an asymmetrical measure of association suitable for use with nominal variables. It can range from 0.0 to 1.0 and gives an indication of the strength of an association between the independent and the dependent variables.

• Gamma is a symmetrical measure of association suitable for ordinal variables or for dichotomous nominal variables. It can vary from 0.0 to ±1.0 and reflects both the strength and direction of the association between two variables.

• Kendall’s tau-b is a symmetrical measure of association suitable for use with ordinal variables. Unlike gamma, it accounts for pairs tied on the independent and dependent variable. It can vary from 0.0 to ±1.0. It provides an indication of the strength and direction of the association between two variables.

• Cramer’s V is a measure of association for nominal variables. It is based on the value of chi-square and ranges between 0.0 to 1.0. Because it cannot take negative values, it is considered a nondirectional measure.

**Outline**

* The Concept of Chi-Square as a Statistical Test
  + Chi-square test is an inferential statistical technique designed to test for significant relationships between two nominal or ordinal variables organized in a bivariate table
  + One of the most widely used tests in the social sciences
  + It requires no assumptions about the shape of the population distribution from which a sample is drawn
  + It can be applied to nominally or ordinally measured variables
* The Concept of Statistical Independence
  + Statistical independence is the absence of association between two cross-tabulated variables
  + The percentage distributions of the dependent variable within each category of the independent variable are identical
* The Structure of Hypothesis Testing With Chi-Square
  + The Assumptions
    - It assumes random sampling
    - It can be applied to variables measured at a nominal and/or an ordinal level of measurement
  + Stating the research and the null hypotheses
    - The research hypothesis (H1) proposes that the two variables are related in the population
    - The null hypothesis (H0) states that no association exists between two cross-tabulated variables in the population, and therefore, the variables are statistically independent
  + Calculating the expected frequencies
    - The chi-square test is based on cell-by-cell comparisons between the expected frequencies and the frequencies actually observed
    - The most important element in using chi-square to test for the statistical significance of cross-tabulated data is the determination of the expected frequencies
  + Calculating the obtained chi-square
    - The chi-square (obtained) is the test statistic that summarizes the differences between the observed (fo) and the expected (fe) frequencies in a bivariate table
  + The sampling distribution of chi-square
    - The chi-square sampling distributions depend on the degrees of freedom
    - Like the t distribution, the sampling distribution of chi-square is a family of distributions
  + Making a final decision
    - The size of the calculated chi-square is directly proportional to the size of the sample, independent of the strength of the relationship between the variables
    - Chi-square is sensitive to small expected frequencies in one or more of the cells in the table
* Focus on Interpretation
  + Education and health assessment
* Reading the Research Literature
  + Violent offense onset by gender, race, and age
* Proportional Reduction of Error: A Brief Introduction
  + Measure of association is a single summarizing number that reflects the strength of a relationship, indicates the usefulness of predicting the dependent variable from the independent variable, and often shows the direction of the relationship
  + It is based on the concept of the proportional reduction of error
  + Proportional reduction of error (PRE) is the concept that underlies the definition and interpretation of several measures of association
  + PRE measures are derived by comparing the errors made in predicting the dependent variable while ignoring the independent with errors made when making predictions that use information about the independent variable
* Lambda: A Measure of Association for Nominal Variables
  + Lambda is an asymmetrical measure of association suitable for use with nominal variables and may range from 0.0 to 1.0
  + It provides us with an indication of the strength of an association between the independent and dependent variables
  + Lambda will vary depending on which variable is considered the independent variable and which the dependent variable
* Cramer’s *V*: A Chi-Square-Related Measure of Association for Nominal Variables
  + Cramer’s *V* is an alternative measure of association that can be used for nominal variables
  + It is based on the value of chi-square and ranges between 0 to 1, with 0 indicating no association and 1 indicating perfect association
  + It is a nondirectional measure
* Focus on Interpretation
  + Gamma and Kendall’s tau-*b*
  + Gamma and Kendall’s tau-*b* are symmetrical measures of association suitable for use with ordinal variables or with dichotomous nominal variables
  + Their value will be the same regardless of which variable is the independent variable or the dependent variable
  + Both can vary from 0.0 to ±1.0 and provide us with an indication of the strength and direction of the association between the variables
* Using Ordinal Measures with Dichotomous Variables
  + An ordinal measure of association might be preferable for tables when an association cannot be detected by lambda
  + Dichotomous variables can be treated as ordinal variables for most purposes
* Focus on Interpretation
  + The gender gap in gun control