Lecture Notes

# Chapter 9: Bivariate Tables

## Learning Objectives

1. Create and analyze a bivariate table.
2. Identify the properties of a bivariate relationship: existence, strength, and direction.
3. Explain how to elaborate the relationship between variables: nonspuriousness, intervening, and conditional relationships.

## Chapter Outline

1. Introduction
   1. **Cross-tabulation:** A technique for analyzing the relationship between two variables (an independent and a dependent variable) that have been organized in a table.
   2. **Bivariate analysis:** A method designed to detect and describe the relationship between two nominal or ordinal variables.
   3. This chapter focuses on not only how to detect whether two variables are associated but also how to determine the strength of the association and, when appropriate, its direction.
2. How to Construct a Bivariate Table
   1. **Bivariate table:** displays the distribution of one variable across the categories of another variable.
   2. Features of bivariate tables:
      1. The table’s title is descriptive, identifying its content in terms of the two variables.
      2. A table may have more columns and more rows, depending on how many categories the variables represent. Usually, the independent variable is the **column variable** and the dependent variable is the **row variable.**
      3. The intersection of a row and a column is called a **cell**.
      4. The column and row totals are the frequency distribution for each variable, respectively. Row and column totals are sometimes called **marginals**.
         1. The total number of cases (*N*) is the number reported at the intersection of the row and column totals.
      5. Usually referred to as a *r* × *c* table, in which *r* represents the number of rows and *c* the number of columns.
3. How to Compute Percentages in a Bivariate Table
   1. Calculating Percentages Within Each Category of the Independent Variable
      1. When the independent variable is arrayed in the columns, we compute percentages within each column separately.
      2. The frequencies within each cell and the row marginals are divided by the total of the column in which they are located, and the column totals should sum to 100%.
      3. When the independent variable is arrayed in the rows, we compute percentages within each row separately.
      4. The frequencies within each cell and the column marginals are divided by the total of the row in which they are located, and the row totals should sum to 100%.
   2. Comparing the Percentages Across Different Categories of the Independent Variable
      1. Comparisons are made by examining differences between percentage points across different categories of the independent variable.
      2. Some researchers limit their comparisons to categories with at least a 10 percentage point difference.
      3. Number of comparison to interpret table:
         1. For any 2 × 2 table, only one comparison needs to be made to interpret the table.
         2. For a larger table, more than one comparison can be made and used in interpretation.
4. The Properties of a Bivariate Relationship
   1. The Existence of the Relationship: A relationship is said to exist between two variables in a bivariate table if the percentage distributions vary across the different categories of the independent variable.
   2. The Strength of the Relationship
      1. A quick method to determine the strength of the association between two variables in a bivariate table is to examine the percentage difference across the different categories of the independent variable.
      2. The larger the percentage difference across the categories, the stronger the association.
      3. Percentage differences are a rough indicator of the strength of a relationship between two variables.
   3. The Direction of the Relationship
      1. A **positive** bivariate relationship exists when the variables vary in the same direction. Higher values of one variable go together with higher values of the other variable.
      2. In a **negative** bivariate relationship, the variables vary in opposite directions: Higher values of one variable go together with lower values of the other variable.
5. Elaboration
   1. **Elaboration**: A process designed to further explore a bivariate relationship, involving the introduction of additional variables, called **control variables.**
   2. By adding a control variable to our analysis, we are considering or “controlling” for the variable’s effect on the bivariate relationship.
      1. Each potential control variable represents an alternative explanation for the bivariate relationship under consideration.
   3. Primary goals: The introduction of additional control variables into a bivariate relationship serves three primary goals in data analysis.
      1. Elaboration allows us to test for nonspuriousness.
         1. Establishing cause-and-effect relations requires not only showing that an independent and a dependent variable are associated but also establishing the time order between them and providing theoretical and empirical evidence that the association is nonspurious, that is, it cannot be “explained away” by other variables.
      2. Elaboration clarifies the causal sequence of bivariate relationships by introducing variables hypothesized to intervene between the independent and dependent variables.
      3. Elaboration specifies the different conditions under which the original bivariate relationship might hold.
   4. Testing for Nonspuriousness: Firefighters and Property Damage
      1. **Direct causal relationship:** The relationship between two variables is said to be a direct causal relationship when it cannot be accounted for by other theoretically relevant variables.
      2. For example, in the case of firefighting:
         1. When the fire is large, more firefighters are sent to the site, and there is a great deal of property damage.
         2. Similarly, when the fire is small, fewer firefighters are at the site, and there is probably very little damage.
         3. The number of firefighters and the extent of property damage are both related to the variable *size of fire* but are not related to each other.
         4. Then the size of the fire is called a *control variable*.
      3. **Spurious relationship:** A relationship between two variables in which both the independent and dependent variables are influenced by a causally prior control variable, and there is no causal link between them.
      4. Researchers have adopted the following rule of thumb for determining whether a relationship between two variables is either direct (causal) or spurious:
6. If the bivariate relationship between the two variables remains about the same after controlling for the effect of one or more causally prior and theoretically relevant variables, then the original bivariate relationship is said to be a direct (causal relationship) association.
7. Conversely, if the original bivariate relationship decreases considerably (or vanishes), then the bivariate relationship is said to be spurious.
   * 1. Three steps for finding relationship:
8. Divide the observations into subgroups based on the control variable. We have as many subgroups as there are categories in the control variable. (In our case, there were two subgroups: small and large fires.)
9. Reexamine the relationship between the original two variables separately for the control variable subgroups.
   * + - 1. The separate tables are called **partial tables**; they display the **partial relationship** between the independent (number of firefighters) and dependent (amount of damage) variables within each specific category Variable (small vs. large fire size).
10. Compare the partial relationships with the original bivariate relationship for the total group.
    * + - 1. In a direct causal pattern, the partial relationships will be very close to the original bivariate relationship.
          2. In a spurious pattern, the partial relationship will be much weaker than the original bivariate relationship.
    1. An Intervening Relationship: Religion and Attitude Toward Abortion
       1. The research on the relationship between religious affiliation and attitudes toward abortion has shown a consistent pattern: Religious affiliation is related to the level of support for abortion.
       2. It is argued that religion is systematically related to desired family size: Catholics prefer larger numbers of children than non-Catholics.
       3. Therefore, preferred family size operates as an intervening mechanism through which the relationship between religion and abortion attitudes occurs.
       4. **Intervening variable:** A control variable that follows an independent variable but precedes the dependent variable in a causal sequence.
       5. **Intervening relationship:** One between two variables in which a control variable intervenes between the independent and dependent variables.
    2. Conditional Relationships: More on Abortion
       1. William Arney and William Trescher (1976) found that when religious participation is controlled for, there is little difference in abortion attitudes between Catholics and Protestants who attend church less than once a month.
       2. Other researchers note that age and gender may also influence the relationship between religion and abortion attitudes.
       3. **Conditional relationship:** When a bivariate relationship differs for different conditions of the control variable.
       4. Another way to describe a conditional relationship is to say that there is a statistical interaction between the control variable and the independent variable.
       5. Carol Gilligan (1982) argues that whereas men tend to be more concerned with rights and rules, women are more concerned with caring and relationships.
       6. The strength of the relationship between abortion morality and stance on legal abortion is conditioned on gender.
    3. The Limitations of Elaboration
       1. Without theory as a guide, elaboration can become a series of exercises that more closely resembles random shots in the dark than scientific analysis.
       2. Statistical analysis more complex: Even with theory as our guide, the statistical analysis is often more complex than the presentation in this section may suggest.
       3. Perilous gap between theory and analysis: When the control variable was introduced, the real nature of the relationship is not revealed easily.