# Chapter Two

Organization of Information

**Summary**

* The most basic method for organizing data is to classify the observations into a frequency distribution—a table that reports the number of observations that fall into each category of the variable being analyzed.
* Constructing a frequency distribution is usually the first step in the statistical analysis of data.
* To obtain a frequency distribution for nominal and ordinal variables, count and report the number of cases that fall into each category of the variable along with the total number of cases (N).
* To construct a frequency distribution for interval-ratio variables that have a wide range of values, first combine the scores into a smaller number of groups—known as class intervals—each containing a number of scores.
* Proportions and percentages are relative frequencies. To construct a proportion, divide the frequency (f) in each category by the total number of cases (N). To obtain a percentage, divide the frequency (f) in each category by the total number of cases (N) and multiply by 100.
* Percentage distributions are tables that show the percentage of observations that fall into each category of the variable. Percentage distributions are routinely added to almost any frequency table and are especially important if comparisons between groups are to be considered.
* Cumulative frequency distributions allow us to locate the relative position of a given score in a distribution. They are obtained by adding to the frequency in each category the frequencies of all the categories below it.
* Cumulative percentage distributions have wider applications than cumulative frequency distributions. A cumulative percentage distribution is constructed by adding to the percentages in each category the percentages of all the categories below it.
* One other method of expressing raw frequencies in relative terms is known as a rate. Rates are defined as the number of actual occurrences in a given time period divided by the number of possible occurrences. Rates are often multiplied by some power of 10 to eliminate decimal points and make the number easier to interpret.

**Outline**

* Frequency Distributions
  + - The most basic method for organizing data
    - A frequency distribution is a table that reports the number of observations that fall into each category of the variable we are analyzing
    - Constructing a frequency distribution is usually the first step in the statistical analysis of data
* Proportions and Percentages
* A proportion is a relative frequency obtained by dividing the frequency in each category by the total number of cases
* To find a proportion (p), divide the frequency (f) in each category by the total number of cases (N)
* To determine a frequency from a proportion, multiply the proportion by the total N
* A percentage is a relative frequency obtained by dividing the frequency in each category by the total number of cases and multiplying by 100
* Percentage Distributions
  + - A percentage distribution is a table showing the percentage of observations falling into each category of the variable
    - Percentage distributions (or proportions) should always show the base (N) on which they were computed
* Comparisons
* We are frequently faced with problems that call for some way to make clear and valid comparisons
* The decision to consider these groups separately or to pool them depends to a large extent on our research question
* Several types of comparisons are quite common in the social sciences
  + Between groups that have different characteristics
  + Regional differences among groups
  + Comparing changes in the same group over time
* Statistics in Practice
  + - Labor force participation among foreign born
* The Construction of Frequency Distributions
  + - Frequency distributions for nominal variables
    - Frequency distributions for ordinal-level variables
    - The major difference between frequency distributions for nominal and ordinal variables is the order in which the categories are listed
    - Frequency distributions for interval-ratio variables
    - Very often interval-ratio variables have a wide range of values, which makes simple frequency distributions very difficult to read
* Cumulative Distributions
  + - A cumulative frequency distribution shows the frequencies at or below each category of the variable
    - Cumulative frequencies are appropriate only for variables that are measured at an ordinal level or higher
    - They are obtained by adding to the frequency in each category the frequencies of all the categories below it
    - A cumulative percentage distribution shows the percentage at or below each category (class interval or score) of the variable
    - A Closer Look
      * Real limits, stated limits, and midpoints of class intervals
* Rates
  + - A rate is a number obtained by dividing the number of actual occurrences in a given time period by the number of possible occurrences
    - Calculating the poverty rate can be done by looking at the number of people in poverty in 2011/Total population in 2011
      * Crude rate
* Statistics in Practice
  + - Civilian labor force participation rates over time
    - Like percentages, rates are useful in making comparisons between different groups and over time
* Reading the Research Literature: Statistical Tables
  + - The first step in reading any statistical table is to understand what the researcher is trying to tell you
    - Tables can sometimes present data for only a subset of the sample