Lecture Notes

# Chapter 2: The Organization and Graphic Presentation of Data

## Learning Objectives

1. Construct and analyze frequency, percentage, and cumulative distributions.
2. Calculate proportions and percentages.
3. Compare and contrast frequency and percentage distributions for nominal, ordinal, and interval-ratio variables.
4. Construct and interpret a pie chart, bar graph, histogram, the statistical map, line graph, and time-series chart.

## Chapter Outline

1. Introduction
   1. Demographers examine the size, composition, and distribution of human populations. Changes in the birth, death, and migration rates of a population affect its composition and social characteristics.
   2. In order to examine a large population, researchers often have to deal with very large amounts of data.
   3. To make sense out of these data, a researcher must organize and summarize the data in some systematic fashion.
   4. Two such methods used by social scientists: The creation of frequency distributions and the use of graphic presentation.
2. Frequency Distributions
   1. The most basic way to organize data is to classify the observations into a frequency distribution.
   2. **A frequency distribution** is a table that reports the number of observations that fall into each category of the variable we are analyzing.
   3. Constructing a frequency distribution is usually the first step in the statistical analysis of data.
3. Proportions and Percentages
   1. Frequency distributions are helpful in presenting information in a compact form.
      1. However, when the number of cases is large, the frequencies may be difficult to grasp.
   2. To standardize these raw frequencies, we can translate them into relative frequencies, that is, proportions or percentages.
   3. A **proportion** is a relative frequency obtained by dividing the frequency in each category by the total number of cases.
4. To find a proportion (*p*), divide the frequency (*f*) in each category by the total number of cases (*N*).



where *f* = frequency, *N* = total number of cases.

1. Proportions should always sum to 1.00 (allowing for some rounding errors).
2. To determine a frequency from a proportion, we simply multiply the proportion by the total *N*:



1. The obtained frequency differs somewhat from the actual frequency due to rounding off of the proportion.
2. If we use the actual proportion instead of the rounded proportion, we obtain the correct frequency.
3. A **percentage** is a relative frequency obtained by dividing the frequency in each category by the total number of cases and multiplying by 100.
4. In most statistical reports, frequencies are presented as percentages rather than proportions.
5. To calculate a percentage, multiply the proportion by 100:



or



1. Percentage Distributions
   1. Percentages are usually displayed as percentage distributions.
      1. A **percentage distribution** is a table showing the percentage of observations falling into each category of the variable.
      2. Percentage distributions (or proportions) should always show the base (*N*) on which they were computed.
2. The Constructions of Frequency Distributions
   1. To construct frequency distributions:
3. We can use statistical software.
4. Go through the process to understand how frequency distributions are actually constructed.
5. For nominal and ordinal variables, constructing a frequency distribution is quite simple.
6. To do so, count and report the number of cases that fall into each category of the variable along with the total number of cases (N).
   1. Frequency Distributions for Nominal Variables
      1. For the nominal variable, gender:
         1. First, tally the number of males, then the number of females.
         2. The tally results are then used to construct the frequency distribution.
         3. The table has a title describing its content.
         4. Its categories (male and female) and their associated frequencies are clearly listed.
         5. In addition, the total number of cases (*N*) is also reported.
      2. Adding percentage distributions
7. The percentage column is the percentage distribution for this variable.
8. To convert the frequency column to percentages, simply divide each frequency by the total number of cases and multiply by 100.
9. Percentage distributions are routinely added to almost any frequency table and are especially important if comparisons with other groups are to be considered.
10. Immediately, we can see that it is easier to read the information.
11. Based on this frequency distribution, we can also conclude that the majority of sample respondents are female.
    1. Frequency Distributions for Ordinal Variables
       1. To construct a frequency distribution for ordinal-level variables, follow the same procedures outlined for nominal-level variables.
       2. Difference between frequency distributions for nominal and ordinal variables:
          1. The major difference is the order in which the categories are listed.
          2. The categories for nominal-level variables do not have to be listed in any particular order.
          3. Because the categories or values of ordinal variables are rank-ordered, however, they must be listed in a way that reflects their rank from the lowest to the highest or from the highest to the lowest.
    2. Frequency Distributions for Interval-Ratio Variables
       1. Very often interval-ratio variables have a wide range of values, which makes simple frequency distributions very difficult to read.
       2. For a more concise picture, the large number of different scores could be reduced into a smaller number of groups, each containing a range of scores.
       3. Frequency tells us the number of respondents who fall into each of the intervals.
       4. The percentage distribution displays the relative frequency of each interval and emphasizes this pattern as well.
12. Cumulative Distributions
    1. A **cumulative frequency distribution** shows the frequencies at or below each category of the variable.
       1. Cumulative frequencies are appropriate only for variables that are measured at an ordinal level or higher.
       2. They are obtained by adding to the frequency in each category the frequencies of all the categories below it.
       3. It shows the cumulative frequencies based on the frequency distribution.
       4. The cumulative frequency column, denoted by Cf, shows the number of persons at or below each interval.
       5. To construct a cumulative frequency distribution, start with the frequency in the lowest class interval (or with the lowest score, if the data are ungrouped) and add to it the frequencies in the next highest class interval.
       6. Continue adding the frequencies until you reach the last class interval. The cumulative frequency in the last class interval will be equal to the total number of cases (*N*).
    2. A **cumulative percentage distribution** shows the percentage at or below each category (class interval or score) of the variable.
       1. This has wider applications than the cumulative frequency distribution.
       2. It is constructed using the same procedure as for a cumulative frequency distributions except the percentages--rather than the frequencies--for each category are added to the total percentages for all the previous categories.
       3. These are useful when comparing differences between groups.
13. Rates
    1. A **rate** is obtained by dividing the number of actual occurrences in a given time period by the number of possible occurrences.



* 1. Rates are often expressed as rates per thousand or hundred thousand to eliminate decimal points and make the number easier to interpret.
  2. A crude rate is based on the total population.

1. Graphic Presentation of Data
   1. Statistical graphs summarize hundreds or thousands of numbers.
   2. Graphs communicate information visually, rather than in words or numbers and are often utilized in news stories, research reports, and government documents.
   3. Information presented graphically may seem more accessible than the same information when presented in frequency distributions or in other tabular forms.
2. The Pie Chart
   1. A **pie chart** shows the differences in frequencies or percentages among the categories of a nominal or an ordinal variable.
   2. The categories are displayed as segments of a circle whose pieces add up to 100% of the total frequencies.
   3. Helps interpret information easily.
3. The Bar Graph
   1. The **Bar Graph** provides an alternative way to graphically present nominal or ordinal data. It shows the differences in frequencies or percentages among categories of a nominal or an ordinal variable.
   2. The categories are displayed as rectangles of equal width with their height proportional to the frequency or percentage of the category.
   3. To construct a bar graph:
      1. First label the categories of the variables along the horizontal axis.
      2. For these categories, construct rectangles of equal width with the height of each proportional to the frequency or percentage of the category.
      3. Note that a space separates each of the categories to make clear that they are nominal categories.
   4. Bar graphs are often used to compare one or more categories of a variable among different groups.
4. The Histogram
   1. The **histogram** is used to show the differences in frequencies or percentages among categories of an interval-ratio or ordinal variable.
      1. The categories are displayed as contiguous bars.
      2. Width is proportional to the width of the category.
      3. Height proportional to the frequency or percentage of that category.
   2. Difference between a histogram and a bar graph:
      1. A histogram looks very similar to a bar graph except that the bars are contiguous to each other (touching) and may not be of equal width.
      2. In a bar graph, the spaces between the bars visually indicate that the categories are separate.
      3. In a histogram, the touching bars indicate that the categories or intervals are ordered from low to high in a meaningful way.
   3. To construct the histogram:
      1. Arrange the intervals along the horizontal axis and the frequencies (or percentages) along the vertical axis.
      2. For each age category, construct a bar with the height corresponding to the frequency in the population in that category.
      3. The width of each bar corresponds to the number that the interval represents.
      4. The bar for each category is touching the bar associated with the category above and below.
      5. The area that each bar occupies tells us the number of individuals that falls into a given interval.
5. The Statistical Map
   1. A **statistical map** presents geographic data patterns or variations such as population distribution, voting patterns, crime rates, or labor force composition.
   2. Maps may also display geographical patterns on the level of cities, counties, city blocks, census tracts, and other units.
   3. One’s choice of whether to display variations on the state level or for smaller units will depend on the research question one wishes to explore.
6. The Line Graph
   1. The **line graph** is another way to display interval-ratio distributions.
   2. It shows the differences in frequencies or percentages among categories of an interval-ratio variable.
   3. Compared with histograms, line graphs are better suited for comparing how a variable is distributed across two or more groups or across two or more time periods.
   4. Points representing the frequencies of each category are placed above the midpoint of the category and are joined by a straight line.
7. The Time Series Chart
   1. A **time-series chart** displays changes in a variable at different points in time.
   2. It involves two variables:
      1. Time labeled across the horizontal axis.
      2. Values (frequencies, percentages, or rates) are labeled along the vertical axis.
   3. To construct a time-series chart, use a series of dots to mark the value of the variable at each time interval and then join the dots by a series of straight lines.