

Chapter Outlines for:

Frey, L., Botan, C., & Kreps, G. (1999). *Investigating communication: An introduction to research methods*. (2nd ed.) Boston: Allyn & Bacon.

Chapter 11: Describing Quantitative Data

I. Introduction

- A. We are constantly bombarded by information in the form of statistics, numbers people use to describe things; some are easy to understand, others are more complex.
- B. This chapter is designed to help explain various types of statistics used to analyze quantitative (numerical) data. The goal is to help you become a competent consumer of such information.

II. Making Sense of Numbers: Statistical Data Analysis

- A. A set of acquired data is not very useful in itself; it needs to be analyzed and interpreted.
- B. **Data Analysis:** The process of examining what data mean to researchers, transforming data into useful information that can be shared.
 - 1. **Statistical data analysis:** The process of examining what quantitative data mean to researchers.
- C. “Statistics” comes from the Latin “*status*” which implies that **statistics** are used to understand the status or state of quantitative data; statistics refers to any numerical indicator of a set of data.
 - 1. There are two broad levels of statistics: Descriptive and inferential statistics.
- D. Statistics are one of the most pervasive and persuasive forms of evidence used to make arguments; one should be wary of the fact that statistics are often used in a less than honest manner (“*stiscalculation*”).
 - 1. The lack of knowledge about statistical information results either in the tendency to reject the information completely or to accept it at face value.
- E. The word statistics refers to two items:
 - 1. Products: Numerical descriptions and inferential statistics that characterize a set of quantitative data.
 - 2. Techniques: The application of procedures to produce numerical descriptions and statistical inferences.
- F. There exist two general purposes of description and inference as reflected in two types of statistical data analysis:
 - 1. **Descriptive statistical data analysis:** Used to construct simple descriptions about the characteristics of a set of quantitative data.
 - a. *Summary statistics:* Numerical indicators that summarize the data.
 - b. Conversion of raw scores into *standard scores*.
 - c. Constructing visual displays of data.
 - 2. **Inferential statistical data analysis:** Two purposes:
 - a. *Estimation:* Estimating the characteristics of a population from data gathered on a sample.
 - b. *Significance testing:* Testing for significant statistical differences between groups and significant statistical relationships between variables.

III. Describing Data Through Summary Statistics

- A. The data that comprise a relatively large data set must be condensed in some way to make sense of them.
 - 1. The data is condensed to a numerical indicator that best summarizes the data set, called a **summary statistic**, which is an efficient way to describe an entire set of quantitative data.
- B. Researchers seek a summary statistic that provides the “typical” point in a data set, the best

representation.

1. **Measures of central tendency (representative values):**

- a. Each is a summary statistic that identifies the center point of an **array** of quantitative data.
- b. **Mode**: The simplest measure of central tendency; indicates which score or value in a distribution occurs most frequently.
 - i. Most appropriate for nominal data.
 - ii. **Antimode**: Score or value that occurs the least frequently.
 - iii. **Bimodal/unimodal distribution**, that is, two modes exist as opposed to one mode.
 - iv. **Multimodal distribution**, that is, three or more modes are present.
- c. **Median**: Divides a distribution of quantitative data exactly in half (locational, positional, or order measure).
 - i. Very appropriate for describing the center point of a set of ordinal data and often is effective for describing the center point of interval/ratio data.
 - ii. Takes into account the entire set of scores in a distribution.
 - iii. Outliers: Extreme scores; median is not swayed.
 - iv. The median is not sensitive to changes in the scores that comprise the data set.
- d. **Mean**: The arithmetic average, computed by adding all the scores in a distribution of interval/ratio data and dividing by the total number of scores.
 - i. Most appropriate and effective measure of central tendency of a set of interval/ratio data because it is sensitive to every change in the data; nonresistant to extreme scores.
 - ii. **Trimmed means**: Removes the most extreme scores entirely.
 - iii. **Winsorizing**: Changes the most extreme scores to the next less extreme scores.
 - iv. **Geometric and harmonic means** are used for measuring relative change and averaging rates respectively.
 - v. The mean most often results in a fractional value.

C. Measures of Dispersion: Simply giving the central tendency summary statistic itself by itself can be misleading.

1. Measures of central tendency tell us nothing about how the scores in a distribution differ or vary from this representative score.
2. **Measures of dispersion** (measures of variability) report how much scores vary from each other or how far they are spread around the center point of a data set and across that distribution (several types).
 - a. **Variation ratio**, used only for nominal data
3. **Range**: (Span); simplest measure of dispersion, reports the distance between the highest and lowest scores in a distribution.
 - a. Calculated by subtracting the lowest number from the highest number in a distribution (extreme values).
 - b. Gives a general sense of how much the data spread out across a distribution (see Figure 11.2).
 - c. Knowing the range can also provide a general understanding of how the scores of two groups of people vary.
 - d. Nonresistant measure that is overly sensitive to extreme scores; hence, one or two extreme scores make a distribution look more dispersed than it actually is.
 - e. To compensate, researchers can divide a distribution at the median and calculate the range of values between the median and the lowest score (lowsread) and the range of scores between the median and the highest score (highspread).
 - f. **Interquartile range**: (Quartile deviation or midspread) dividing a distribution at the 25th and 75th percentiles. Dividing the interquartile range in half produces the semi-quartile range. The distribution can also be divided into 10 equal parts, or deciles.

- g. Range and its variations are the only measure of dispersion that can be applied meaningfully to ordinal data, but is limited in describing the amount of dispersion in a set of interval/ratio data.
- 4. **Variance**: The mathematical index of the average distance of scores in a distribution of interval/ratio data from the mean, in squared units.
 - a. Figure 11.3 shows how variance is calculated for this distribution, using both the definitional formula and an easier computational formula.
 - b. **Deviation score**: Amount that one score differs from the mean (deviate or error score).
 - c. Sum of squares: Adding the squared deviation scores cumulatively.
 - d. A variance score is confusing because it is expressed in units of squared deviations about the mean, which are not the same as the original unit of measurement.
 - e. **Standard Deviation (SD)**: A measure of dispersion that explains how much scores in a set of interval/ratio data vary from the mean, expressed in the original unit of measurement.
 - i. SD is found by taking the square root of the variance; that is, the number that when multiplied by itself equals the variance.
 - ii. The mean and standard deviation are the measures of central tendency and dispersion reported most frequently by researchers when analyzing interval/ratio data; they are often reported in a table, which is helpful when scores for different conditions or groups of people are compared (see Figure 11.4).

IV. Describing Data in Standard Scores

- A. **Standard score/standard normal deviates**: Researchers often report how many standard deviations a particular score is above or below the mean of its distribution.
 - 1. *Standard scores* provide a common unit of measurement that indicates how far any particular score is away from its mean.
 - a. **Z-score**: Standard score used frequently by researchers; the formula says that it can be computed for any score in a distribution by dividing its deviation score by the standard deviation for the distribution.
 - i. Researchers can meaningfully interpret an individual score within a distribution by showing how many standard deviations it is away from the mean of that distribution.
 - b. **Deviation-IQ scale**: One of the best known examples of using standard scores; used with many group-administered intelligence tests.
 - 2. Another use of standard scores has to do with the fact that different people sometimes do not use the same measurement scale in the same way (see Figure 11.5).
 - 3. Even though the same (valid and reliable) measurement scale may be used, relying on raw scores may not tell the whole story about the data; for this reason, some researchers first convert raw scores into standard scores, or even transformed standard scores.
 - a. **Raw/unstandardized scores**: Data kept in their original unit of measurement and not converted to standard scores.
 - 4. Standard scores also allow comparisons between scores on different types of scales for the same individual or for different people.
 - 5. Standard scores actually allow comparisons between different people's scores on different types of measurements (see Figure 11.6).
 - 6. Converting scores on a measurement scale to standard scores that measure how far away they are from the mean of their distribution is an important tool for describing data.

V. Describing Data Through Visual Displays

- A. Researchers, educators, and other professionals often use tables, figures, graphs, and other forms to visually display distributions of data.
1. These highlight important information in the data and/or make complex data more understandable.
 2. To construct visual displays of data and realize their benefits, one must know what types of visual displays are possible, how to construct them, and when to use them.
 3. The choice of which visual display to use is influenced by the type of data one seeks to portray.
- B. Frequency Tables
1. In one way or another, all visual displays are attempts to describe **frequency distributions**, a tally of the number of times particular values on a measurement scale occur in a data set.
 2. The most basic way is to visually display the absolute frequency of the data obtained, the actual number of times each category or measurement point occurs in a distribution, in a frequency table—a visual display in table form of the frequency of each category or measurement point on a scale for a distribution of data.
 3. **Relative frequency**: The proportion of times each category or measurement point occurs (found by dividing the number of cases for a particular category or measurement point by the total number of cases for the entire distribution).
 4. **Cumulative frequency (CF)**: The total frequency up to and including any particular category or measurement point (see Figure 11.7).
 5. **Frequency tables** are often used to describe the frequency of occurrence of categories, such as the number of times the categories associated with facilitative and inhibitive factors are mentioned; very effective for highlighting which categories are used most often, and can also be helpful for comparing at a glance the frequency of different categories for two or more groups (see Figure 11.8).
 6. Frequency tables are also helpful for showing changes over short or long periods of time; however, they are fairly simple drawings and do not rely on such graphics as shaded areas or connected lines.
- C. **Pie Charts**: Circles divided into segments proportional to the percentage of the circle that represents the frequency count for each category.
1. A favorite way of visually presenting information in the popular media (TV, newspapers, and magazines)
 2. Although used infrequently in scholarly research reports, they have been used to provide descriptive information about a research data set (see Figure 11.9)
 3. Pie charts can also be very helpful for visually showing differences between groups (see Figure 11.10).
- D. **Bar Charts**: A common and relatively simple visual depiction that use shaded “blocks” (rectangles and/or squares) to show the frequency counts for the groups that comprise a nominal independent variable.
1. Typically are drawn using vertical blocks; the independent variable groups are presented on the abscissa, the horizontal (X) axis, and the frequency of occurrence on the dependent variable is presented on the ordinate, the vertical (Y) axis
 2. **Horizontal bar charts**: Contains blocks running horizontally from left to right; sometimes are used when the categories have long names or when there are many categories.
 3. Bar charts are especially helpful for visually depicting differences between groups on a variable (see Figure 11.11).
 4. **Grouped/segmented bar charts**: A very common use of bar charts; visually shows differences between the groups formed from the combination of two or more independent variables (see Figure 11.12).
- E. **Line Graphs**: Use a single point to represent the frequency count on a dependent variable

for each of the groups and then connect these points with a single line.

1. The simplest type of line graph compares two groups with respect to a dependent variable (see Figure 11.13).
2. Line graphs are particularly useful for showing changes in groups over two or more points in time (see Figure 11.14).

F. Frequency Histograms and Frequency Polygons

1. **Frequency histograms:** Like bar graphs, uses blocks to show how frequently each point on an interval/ratio scale occurs; however, because an interval/ratio scale has equal distances between the points on a scale, the blocks touch (see Figure 11.15).
2. **Frequency polygons:** Similar to line graphs, except that a line connects the points representing the frequency counts for each point on the interval/ratio scale used to measure the independent variable (rather than each group of a nominal independent variable, as in line graphs) (see Figure 11.16).

VI. Conclusion

- A. The first step in analyzing a set of data is to understand its characteristics; once important characteristics have been determined, it is possible to go beyond description to infer conclusions about the data.