

Lesson 2

Graphical Descriptions of Data

Learning Objectives

Upon completion of this lesson, you should be able to do the following:

1. Generate pie charts, bar charts, stem and leaf diagrams, and histograms from a data set.
2. Describe aspects of distributions represented by the various graphs.
3. Compare the information related about the data set by each of the types of graphs.

Key Words

frequency, relative frequency, qualitative, quantitative, nominal, categorical data, ordinal, interval, ratio, pie charts, bar charts, stem and leaf diagrams, histograms

Concepts

Graphical descriptions of data are very useful to recognize the general shape of a set of data. There are several types of graphs that are appropriate for different kinds of variables and relate various types of data measured about the variable.

Many graphs relate information about the frequency, or relative frequency, of the values of the variables. The **frequency** of a category or specific value of a data set tells the number of times the category or value was observed in the sample. The **relative frequency** of a category or specific value of a variable tells the proportion of the data that occur in a certain category or have a specific value. The relative frequency of a category or value is the frequency divided by the total number of observations in the data set. The relative frequency of a category or value reveals how big is the frequency of the category or value relative to the size of the whole data set.

Two Types of Variables

Data can either be in the form of words or numbers. There are two types of variables. One type results in numerical responses. The other type results in responses described in words. **Qualitative** variables have possible responses that are words. Examples are one's hometown or religious preference. **Quantitative** variables have responses that are numbers. A person's GPA and daily bank checking account balance are two examples of quantitative variables.

Four Types of Data

Although all data are words or numbers there are four specific types of data. **Nominal** data are observations that can be classified by categories, such as the religious preference of each of the individuals in a group of people. Nominal data are sometimes referred to as **categorical data** and are based on qualitative variables.

Ordinal data are measurements that can be ordered with respect to the other observations in the set, or in other words, measurements that can be rank ordered. Ordinal data can be either qualitative or quantitative.

Interval data are measurements that not only can be rank ordered but the differences between data values tell how much more or less the value is compared to another. Ordinal data contains information about whether a data value is more or less than another value. Interval data has additional information about how much more or less.

Ratio data measures one value of a variable with respect to the value of another variable. Interval and ratio data can only result from quantitative variables.

Various Types of Graphs

One objective of statistics is to describe populations from observed samples. In order to form conclusions about populations from sample information first the sample is described. Graphs can provide a visual summary of the sample information. The graphs discussed here are based on frequency and relative frequency information.

Bar charts represent the frequency or relative frequency of the categories or values of a data set with the height of a bar or line. Each category or value has an associated line or bar whose height represents the frequency or relative frequency of the specific category or certain value.

Pie charts represent the relative frequency of a specific category by assigning a wedge of the pie to each category. The size of the wedge represents the relative frequency of the category. If a category has relative frequency of one-sixth then one-sixth of the 360 degrees in the circle of the graph would be assigned to that wedge of the pie chart, that is, one-sixth times 360 or 60.

Stem and leaf diagrams summarize frequency and relative frequency information by assigning the first part of the data values as the stem and listing all the second parts as the leaves to represent the frequency for each stem. Unlike the other visual techniques, stem and leaf diagrams can be used to regenerate the original data values, not just the relative frequency information about the data. This method is only appropriate for data resulting from quantitative variables.

Histograms represent the relative frequency of categories of data values with the area in the bar. Recall that bar charts use height of lines or bars to represent this information. A histogram uses a height scale that is relative frequency divided by the width of the bar to force the area to be equal to the relative frequency.