

Chapter 2: Notes

December 2, 2019

Section 2.1

Variable

Variable: A characteristic that varies from one person or thing to another.

Qualitative variable: A nonnumerically valued variable.

Quantitative variable: A numerically valued variable.

Discrete variable: A quantitative variable whose possible values can be listed. In particular, a quantitative variable whose possible values can be listed. In particular, a quantitative variable with only a finite number of possible values is a discrete variable.

Continuous variable: A quantitative variable whose possible values form some interval of numbers.

A discrete variable usually involves a count of something, whereas a continuous variable usually involves a measurement of something.

Data

Data: Values of a variable.

Qualitative data: Values of a qualitative variable.

Quantitative data: Values of a quantitative variable.

Discrete data: Values of a discrete variable.

Continuous data: Values of a continuous variable.

Data are classified according to the type of variable from which they were obtained.

Section 2.2

Frequency Distribution of Qualitative Data

A frequency distribution of qualitative data is a listing of the distinct values and their frequency. A frequency distribution provides a table of the values of the observations and how often they occur.

Frequency Distribution of Qualitative Data Procedure

Step 1 List the distinct values of the observations in the data set in the first column of a table.

Step 2 For each observation, place a tally mark in the second column of the table in the row of the appropriate distinct values.

Step 3 Count the tallies for each distinct value and record totals in the third column of the table.

Relative-Frequency Distribution of Qualitative Data

A relative-frequency distribution of qualitative data is a listing of the distinct values and their relative frequencies. A relative-frequency distribution provides a table of the values of the observations and (relatively) how often they occur.

Procedure to Construct a Relative-Frequency Distribution of Qualitative Data

Step 1 Obtain a frequency distribution of the data.

Step 2 Divide each frequency by the total number of observation.

Pie Chart

A pie chart is a disk divided into wedge-shaped pieces proportional to relative frequencies of the qualitative data.

Pie Chart Procedure

Step 1 Obtain a relative-frequency distribution of the data.

Step 2 Divide a disk into wedge-shaped pieces proportional to the relative frequencies.

Step 3 Label the slices with the distinct values and their relative frequencies.

Bar Chart

A bar chart displays the distinct values of the qualitative data on a horizontal axis and the relative frequencies (or frequencies or percents) of those values on a vertical axis. The relative frequency of each distinct value is represented by a vertical axis. The relative frequency of each distinct value is represented by a vertical bar whose height is equal to the relative frequency of the value. The bars should be positioned so that they do not touch each other.

Bar Chart Procedure

Step 1 Obtain a relative-frequency distribution of the data.

Step 2 Draw a horizontal axis on which to place the bars and a vertical axis on which to display the relative frequencies.

Step 3 For each distinct value, construct a vertical bar whose height equals the relative frequency of that value.

Step 4 Label the bars with the distinct values, the horizontal axis with the name of the variable, and the vertical axis with "Relative Frequency."

Section 2.3

Single-Value Grouping

In data sets with only a few discrete values the best grouping method is usually single-value grouping, where each value constitutes its own class.

Limit Grouping

When the data set of integers has too many distinct values to be single-value grouped, limit grouping may be the best method. Using class limits, values are grouped into sequential groups.

Lower class limits: The smallest value that could go in a class.

Upper class limit: The largest value that could go in a class.

Class width: The difference between the lower limit of a class and the lower limit of the next-higher class.

Class mark: The mean of the upper and lower limits of a class.

Cutpoint Grouping

When the data set is continuous and its members are real valued the best grouping method may be cutpoint grouping. Using class cutpoints, values are grouped in sequential groups.

Lower class cutpoint: The smallest value that could go in a class.

Upper class cutpoint: The smallest value that could go in the next-higher class (equivalent to the lower cutpoint of the next-higher class).

Class width: The difference between the cutpoints of a class.

Class midpoint: The mean of the two cutpoints of a class.

Histograms

A histogram provides a graph of the values of the observations and how often they occur. A histogram displays the classes of the quantitative data on a horizontal axis and the frequencies, relative frequencies, or percents of those classes on a vertical axis. The frequency, relative frequency, or percent of each class is represented by a vertical bar whose height is equal to the frequency (relative frequency, percent) of that class. The bars should be positioned so that they touch each other.

- For single-valued grouping, we use distinct values of observations to label the bars, with each such value centered under its bar.
- For limit grouping or cutpoint grouping, we use the lower class limits (or equivalently, lower class cutpoints) to label the bars.
- Note: Some statisticians and technologies use class marks or class midpoints centered under the bars.

Histogram Procedure

Step 1 Obtain a frequency, relative-frequency, or percent distribution of the data.

Step 2 Draw a horizontal axis on which to place the bars and a vertical axis on which to display the frequencies, relative frequencies, or percents.

Step 3 For each class, construct a vertical bar whose height equals the frequency, relative frequency, or percent of that class.

Step 4 Label the bars with classes, the horizontal axis with the name of the variable, and the vertical axis with "Frequency", "Relative frequency," or "Percent".

Dotplot

A dotplot is a graph in which each observation is plotted as a dot at an appropriate place above a horizontal axis. Observations having equal values are stacked vertically.

Dotplot Procedure

Step 1 Draw a horizontal axis that displays the possible values of the quantitative data.

Step 2 Record each observation by placing a dot over the appropriate value on the horizontal axis.

Step 3 Label the horizontal axis with the name of the variable.

Stem-and -Leaf Diagram (Stemplot)

In a stem-and-leaf diagram (or stemplot), each observation is separated into two parts, namely;

Stem: consists of all but the rightmost digit

Leaf; the rightmost digit

Stem-and-Leaf Procedure

Step 1 Think of each observation as a stem, all but the rightmost digit, and a leaf, the rightmost digit.

Step 2 Write the stems from smallest to largest in a vertical column to the left of a vertical rule.

Step 3 Write each leaf to the right of the vertical rule in the row that contains the appropriate stem.

Step 4 Arrange the leaves in each row in ascending order.

Section 2.4

Distribution of a Data Set

The distribution of a data set is a table, graph, or formula that provides the values of the observations and how often they occur.

Population and Sample Data

Population data: The values of a variable for the entire population.

Sample data: The values of a variable for a sample of the population.

Population and Sample Distribution of a Variable

The distribution of a population data is called the population distribution, or the distribution of the variable. The distribution of sample data is called sample distribution.

Population and Sample Distributions

For a simple random sample, the sample distribution approximates the population distribution (i.e., the distribution of the variable under consideration). The larger the sample size, the better the approximation tends to be.