Unit 2: Descriptive statistics (1)

Descriptive statistics are tabular, graphical, and numerical summaries of data. The purpose of descriptive statistics is to facilitate the presentation and interpretation of data. Most of the statistical presentations appearing in newspapers and magazines are descriptive in nature. Univariate methods of descriptive statistics use data to enhance the understanding of a single variable; multivariate methods focus on using statistics to understand the relationships among two or more variables. To illustrate methods of descriptive statistics, the previous example [in unit 1] in which data were collected on the age, gender, marital status, and annual income of 100 individuals will be examined.

Tabular methods

The most commonly used tabular summary of data for a single variable is a frequency distribution. A frequency distribution shows the number of data values in each of several nonoverlapping classes. Another tabular summary, called a relative frequency distribution, shows the fraction, or percentage, of data values in each class. The most common tabular summary of data for two variables is a cross tabulation, a two-variable analogue of a frequency distribution.

For a qualitative variable, a frequency distribution shows the number of data values in each qualitative category. For instance, the variable gender has two categories: male and female. Thus, a frequency distribution for gender would have two nonoverlapping classes to show the number of males and females. A relative frequency distribution for this variable would show the fraction of individuals that are male and the fraction of individuals that are female.

Constructing a frequency distribution for a quantitative variable requires more care in defining the classes and the division points between adjacent classes. For instance, if

the age data of the example above ranged from 22 to 78 years, the following six nonoverlapping classes could be used: 20–29, 30–39, 40–49, 50–59, 60–69, and 70–79. A frequency distribution would show the number of data values in each of these classes, and a relative frequency distribution would show the fraction of data values in each.

A cross tabulation is a two-way table with the rows of the table representing the classes of one variable and the columns of the table representing the classes of another variable. To construct a cross tabulation using the variables gender and age, gender could be shown with two rows, male and female, and age could be shown with six columns corresponding to the age classes 20–29, 30–39, 40–49, 50–59, 60–69, and 70–79. The entry in each cell of the table would specify the number of data values with the gender given by the row heading and the age given by the column heading. Such a cross tabulation could be helpful in understanding the relationship between gender and age.

Graphical methods

A number of graphical methods are available for describing data. A bar graph is a graphical device for depicting qualitative data that have been summarized in a frequency distribution. Labels for the categories of the qualitative variable are shown on the horizontal axis of the graph. A bar above each label is constructed such that the height of each bar is proportional to the number of data values in the category. A bar graph of the marital status for the 100 individuals in the above example is shown in Figure 1. There are 4 bars in the graph, one for each class. A pie chart is another graphical device for summarizing qualitative data. The size of each slice of the pie is proportional to the number of data values in the corresponding class. A pie chart for the marital status of the 100 individuals is shown in Figure 2.

A histogram is the most common graphical presentation of quantitative data that have been summarized in a frequency distribution. The values of the quantitative variable

are shown on the horizontal axis. A rectangle is drawn above each class such that the base of the rectangle is equal to the width of the class interval and its height is proportional to the number of data values in the class.
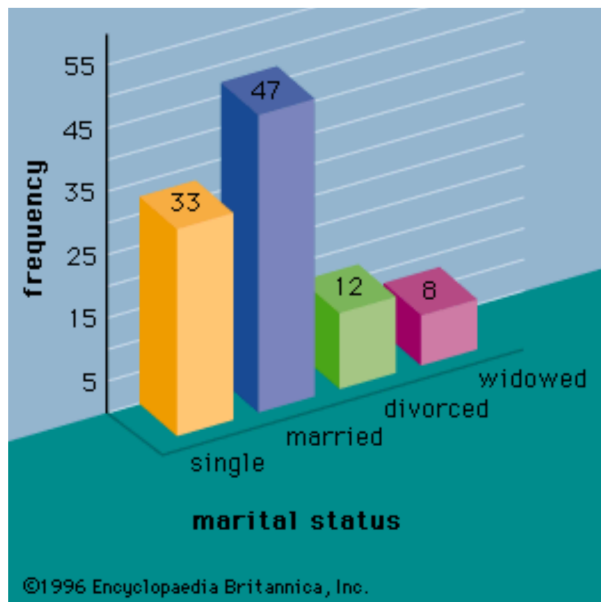

Figure 1: A bar graph showing the marital status of 100 individuals.
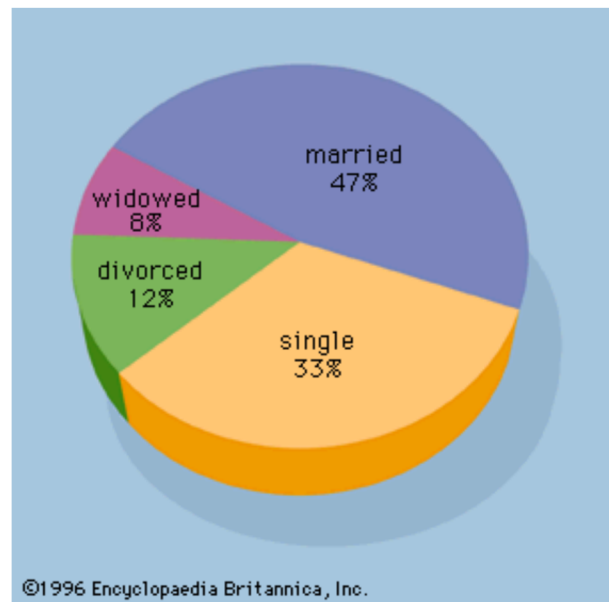
Encyclopædia Britannica, Inc.

Figure 2: A pie chart for the marital status of 100 individuals.

Encyclopædia Britannica, Inc.

Comprehension Exercises

Choose a, b, c or d which best completes each item.

1) Television, newspapers, magazines, and other media mainly present _____ statistics.

a) inferential   b) descriptive   c) computational   d) numerical

2) A frequency distribution shows the number of data values in each qualitative or quantitative _____.

a) stratum   b) group   c) category   d) level

3) The arc length of each slice (and consequently its central angle and area) of a pie chart, is proportional to the _____ of the corresponding class.

a) area   b) label   c) value   d) frequency

4) The frequency distribution of quantitative data is graphically presented by a _____.

a) bar chart   b) pie chart   c) histogram   d) scatter plot

Words to Learn

Find the Persian equivalents of the following terms and expressions.

descriptive	summary	univariate	nonoverlapping
tabular	relationship	multivariate	horizontal
graphical	illustrate	rectangle	vertical
numerical	class	interval	axis
frequency	category	device	plot
relative	one-way	depict	chart
distribution	two-way	adjacent	graph
tabulation	summarize	column	bar
pie	height	row	fraction
proportional	width	cross tabulation	draw