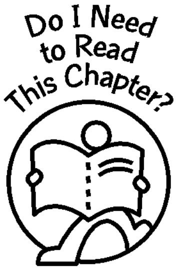
CHAPTER 1

Graphical Displays of Univariate Data



You should read this chapter if you need to review or to learn about

- The subject of statistics
- Some common graphical displays used to represent data
- Frequency distributions
- Dot plots
- Bar charts
- Histograms
- Frequency polygons
- Stem-and-leaf plots
- Pie charts
- Pareto charts

In later chapters, you will be introduced to other graphical displays, and you will recognize that these graphical displays can be combined with other measures to describe the data distribution.

1-1 Introduction

For us to have an understanding of what the subject of **statistics** is all about, we need to introduce some terminology. First we will explain what we mean by the subject of statistics.

Explanation of the term—statistics: Statistics is the science of collecting, organizing, summarizing, analyzing, and making inferences from data.

The subject of statistics is divided into two broad areas that incorporate the collecting, organizing, summarizing, analyzing, and making inferences from data. These categories are *descriptive statistics* and *inferential statistics*. These classifications are shown in **Fig. 1-1.**

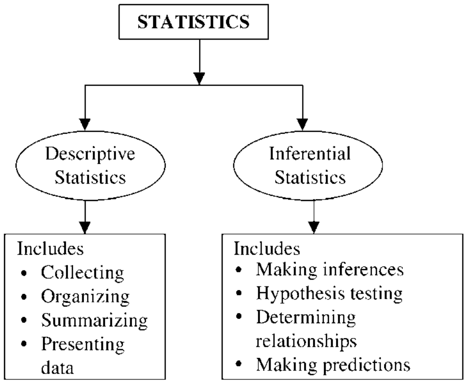


Fig. 1-1: Breakdown of the subject of statistics

In order to obtain information, **data** are collected from variables used to describe an event.

Explanation of the term—data: Data are the values or measurements that variables describing an event can assume.

Variables whose values are determined by chance are called *random variables*. There are two types of variables: **qualitative variables** and **quantitative variables.** Qualitative variables are nonnumeric in nature. Quantitative variables can assume numeric values and can be classified into two groups: **discrete variables** and **continuous variables.** A collection of values is called a *data set*, and each value is called a *data value*. **Figure 1-2** shows these relationships.

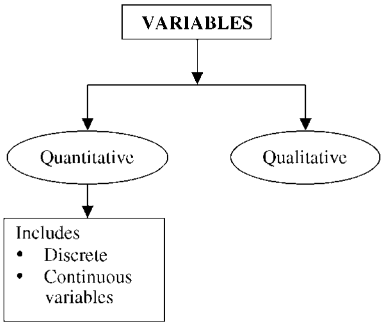


Fig. 1-2: Breakdown of the types of variables

Explanation of the term—quantitative data: Quantitative data are data values that are numeric. For example, the heights of female basketball players are quantitative data values.

Explanation of the term—qualitative data: Qualitative data are data values that can be placed into distinct categories, according to some characteristic or attribute. For example, the eye color of female basketball players is classified as qualitative data.

Explanation of the term—discrete variables: Discrete variables are variables that assume values that can be counted—for example, the number of days it rained in your neighborhood for the month of March.

Explanation of the term—continuous variables: Continuous variables are variables that can assume all values between any two given values—for example, the time it takes for you to do your Christmas shopping.

In order for statisticians to do any analysis, data must be collected. One of the things statisticians may want to do is to make some inference about a characteristic of a **population.** Sometimes it is impractical or too expensive to collect data from the entire population. In such instances, the statistician may select a representative portion of the population, called a **sample.** This is depicted in **Fig. 1-3.**

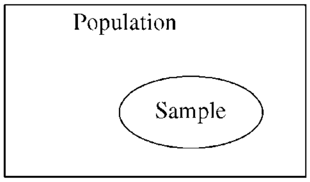


Fig. 1-3: The relationship between sample and population

Explanation of the term—population: A population consists of all elements that are being studied. For example, we may be interested in studying the distribution of ACT math scores of freshmen at a college campus. In this case, the population will be the ACT math scores of all the freshmen on that particular campus.

Explanation of the term—sample: A sample is a subset of the population. For example, we may be interested in studying the distribution of ACT math scores of freshmen at a college campus. In this case, we may select the ACT math score of every tenth freshman from an alphabetical list of the students' last names.

Explanation of the term—census: A **census** is a sample of the entire population. For example, we may be interested in studying the distribution of ACT math scores of freshmen at a college campus. In this case, we may list the ACT math scores for all freshmen on that particular campus.

Both populations and samples have characteristics that are associated with them. These are called **parameters** and **statistics,** respectively.

Explanation of the term—parameter: A parameter is a characteristic of or a fact about a population. For example, we may be interested in studying the distribution of ACT math scores of freshmen at a college campus. In this case, the average ACT math score for all freshmen on this particular campus may be 25.

Explanation of the term—statistic: A statistic is a characteristic of or a fact about a sample. For example, we may be interested in studying the distribution of ACT math scores of freshmen at a college campus. In this case, the average ACT math score for every tenth freshman from an alphabetical list of their last names may be 22.

Since parameters are descriptions of the population, a population can have many parameters. Similarly, a sample can have many statistics. These associations are shown in **Fig. 1-4.**

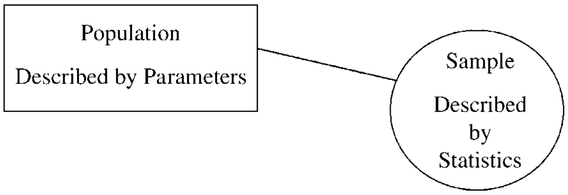


Fig. 1-4: The difference between parameters and statistics

When selecting a sample, statisticians would like to select values in such a way that there is no inherent bias. One way of doing this is by selecting a **random sample.**

Explanation of the term—random sample: A random sample of a particular size is a sample selected in such a way that each group of the same size has an equal chance of being selected. For example, in a lottery game in which six numbers are selected, this will be a random sample of size six, since each group of size six will have an equal chance of being selected.

Quick Tip

There are other types of samples that will not be discussed in this text. These include systematic, stratified, cluster, and convenience samples.

1-2 Frequency Distributions

In this section, we will deal with **frequency distributions.**

Explanation of the term—frequency distribution: A frequency distribution is an organization of raw data in tabular form, using classes (or intervals) and frequencies.

The types of frequency distributions that will be considered in this section are *categorical, ungrouped*, and *grouped* frequency distributions.

Explanation of the term—frequency count: The **frequency** or the **frequency count** for a data value is the number of times the value occurs in the data set.

Categorical or Qualitative Frequency Distributions

Explanation of the term—categorical frequency distributions: Categorical frequency distributions represent data that can be placed in specific categories, such as gender, hair color, or religious affiliation.

Example 1-1: The blood types of 25 blood donors are given below. Summarize the data using a frequency distribution.

AB	B	A	O	B
O	B	O	A	O
B	O	B	B	B
A	O	AB	AB	O
A	B	AB	O	A

Solution: We will represent the blood types as classes and the number of occurrences for each blood type as frequencies. The frequency table (distribution) in **Table 1-1** summarizes the data.

Table 1-1: Frequency Table for Example 1-1

CLASS (BLOOD TYPE)	FREQUENCY *f*
A	5
B	8
O	8
AB	4
Total	25

Quantitative Frequency Distributions—Ungrouped

Explanation of the term—ungrouped frequency distribution: An **ungrouped frequency distribution** simply lists the data values with the corresponding number of times or frequency count with which each value occurs.

Example 1-2: The following data represent the number of defectives observed each day over a 25-day period for a manufacturing process. Summarize the information with a frequency distribution.

DAY	1	2	3	4	5	6	7	8	9	10	11	12	13
Defects	10	10	6	12	6	9	16	20	11	10	11	11	9
DAY	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	
Defects	12	11	7	10	11	14	21	12	6	10	11	6	

Solution: The frequency distribution for the number of defects is shown in **Table 1-2.**

Table 1-2: Frequency Table for Example 1-2

CLASS (DEFECTS)	FREQUENCY *f*
6	4
7	1
9	2
10	5
11	6
12	3
14	1
16	1
20	1
21	1
Total	25

Quick Tip

Sometimes frequency distributions are displayed with the relative frequencies as well.

Explanation of the term—relative frequency: The relative frequency for any class is obtained by dividing the frequency for that class by the total number of observations.

$$\text{Relative frequency} = \frac{\text{frequency for class}}{\text{total number of observations}}$$

The frequency distribution in **Table 1-3** uses the data in **Example 1-2** and displays the relative frequencies and the corresponding percentages.

Table 1-3: Frequency Distribution Along with Relative Frequencies for Example 1-2

CLASS (DEFECTS)	FREQUENCY f	RELATIVE FREQUENCY	PERCENTAGE %
6	4	$4/25 = 0.16$	16
7	1	$1/25 = 0.04$	4
9	2	$2/25 = 0.08$	8
10	5	$5/25 = 0.20$	20
11	6	$6/25 = 0.24$	24
12	3	$3/25 = 0.12$	12
14	1	$1/25 = 0.04$	4
16	1	$1/25 = 0.04$	4
20	1	$1/25 = 0.04$	4
21	1	$1/25 = 0.04$	4
Total	25	1	100

Quick Tip

Sometimes frequency distributions are displayed with the *cumulative frequencies* and *cumulative relative frequencies* as well.

Explanation of the term—cumulative frequency: The cumulative frequency for a specific value in a frequency table is the sum of the frequencies for all values at or below the given value.

Explanation of the term—cumulative relative frequency: The cumulative relative frequency for a specific value in a frequency table is the sum of the relative frequencies for all values at or below the given value.

Note: The explanations given for the cumulative frequency and the cumulative relative frequency assume that the values (or classes) are arranged in ascending order from top to bottom.

The frequency distribution in **Table 1-4** uses the data in **Example 1-2** and also displays the cumulative frequencies and the cumulative relative frequencies.

Table 1-4: Frequency Distribution Along with Relative Frequencies, Cumulative Frequencies, and Cumulative Relative Frequencies for Example 1-2

CLASS (DEFECTS)	FREQUENCY	RELATIVE FREQUENCY	CUMULATIVE FREQUENCY	CUMULATIVE RELATIVE FREQUENCY
6	4	0.16	4	0.16
7	1	0.04	5	0.20
9	2	0.08	7	0.28
10	5	0.20	12	0.48
11	6	0.24	18	0.72
12	3	0.12	21	0.84
14	1	0.04	22	0.88
16	1	0.04	23	0.92
20	1	0.04	24	0.96
21	1	0.04	25	1.00

Quantitative Frequency Distributions—Grouped

Here we will discuss the idea of **grouped frequency distributions.**

Explanation of the term—grouped frequency distribution: A grouped frequency distribution is obtained by constructing classes (or intervals) for the data, and then listing the corresponding number of values (frequency count) in each interval.

Quick Tip

There are several procedures that one can use to construct a grouped frequency distribution. However, because of the many statistical software packages available today, it is not necessary to try to construct such distributions using pencil and paper. Later in the chapter, we will encounter a graphical display called the *histogram*. We will see that one can directly construct grouped frequency distributions from these graphical displays.

Quick Tip

A frequency distribution should have a minimum of 5 classes and a maximum of 20. For small data sets, one can use between 5 and 10 classes. For large data sets, one can use up to 20 classes.

Example 1-3: The weights of 30 female students majoring in Physical Education on a college campus are given below. Summarize the information with a frequency distribution using seven classes.

143	151	136	127	132	132	126	138	119	104
113	90	126	123	121	133	104	99	112	129
107	139	122	137	112	121	140	134	133	123

Solution: A grouped frequency distribution for the data using seven classes is presented in **Table 1-5.** Observe, for instance, that the upper limit value for the first class and the lower limit value for the second class have the same value, 95. The value of 95 cannot be included in both classes, so the convention that will be used here is that *the upper limit of each class is not included in the interval of values*; only the lower limit value is included in the interval. Thus, the value of 95 is included only in the interval of values for the second class.

Table 1-5: Grouped Frequency Distribution for Example 1-3

CLASS (WEIGHTS)	FREQUENCY	RELATIVE FREQUENCY	PERCENTAGE %
85–95	1	0.03	3
95–105	3	0.10	10
105–115	4	0.13	13
115–125	6	0.20	20
125–135	9	0.30	30
135–145	6	0.20	20
145–155	1	0.03	3
Total	30	≈1	≈100

Note: The *class width* for this frequency distribution is 10. It is obtained by subtracting the lower class limit for any class from the lower class limit for the next class. For the third class, the class limit = 115 – 105 = 10.

Quick Tip

In the grouped frequency distribution, observe that the relative frequency column did not add up to exactly 1 and the percentage column did not add up to exactly 100 percent. This is due to the rounding of the relative frequency values to two decimal places.

1-3 Dot Plots

Explanation of the term—dot plot: A **dot plot** is a plot that displays a dot for each value in a data set along a number line. If there are multiple occurrences of a specific value, then the dots will be stacked vertically.

Example 1-4: Construct a dot plot for the information given in **Example 1-2.**

Solution: Figure 1-5 shows the dot plot for the data set. Observe that since there are multiple occurrences of specific observations, the dots are stacked vertically. The number of dots represents the frequency count for a specific value. For instance, the value of 11 occurred 6 times, since there are 6 dots stacked above the value of 11.

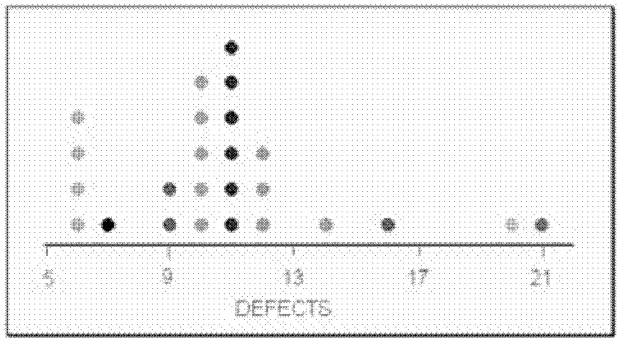


Fig. 1-5: Dot plot for Example 1-2

1-4 Bar Charts or Bar Graphs

Explanation of the term—bar chart (graph): A **bar chart** or a **bar graph** is a graph that uses vertical or horizontal bars to represent the frequencies of the categories in a data set.

Quick Tip

A bar chart (graph) is a valuable presentation tool, since it is effective at reinforcing differences in magnitude. Bar charts permit the visual comparison of data by displaying the magnitude of each category as a horizontal or vertical bar. Bar charts are useful when the data set has categories (for example, hair color, gender, etc.) and data values that are qualitative in nature. Note that the bars are equally separated.

Example 1-5: A sample of 300 college students was asked to indicate their favorite soft drink. The survey results are shown in **Table 1-6.** Display the information using a bar chart.

Table 1-6: Frequency Distribution for Example 1-6

SOFT DRINK	NUMBER OF STUDENTS
Pepsi-Cola	92
Coca-Cola	78
Dr. Pepper	48
7-Up	42
Others	40

Solution: Observe that these are categorical or qualitative data. The vertical bar chart for this information is shown in **Fig. 1-6.** The number at the top of each category represents the number of values (frequency) for that specific group (soft drink).

A horizontal bar chart for the same soft drink information is shown in **Fig. 1-7.**

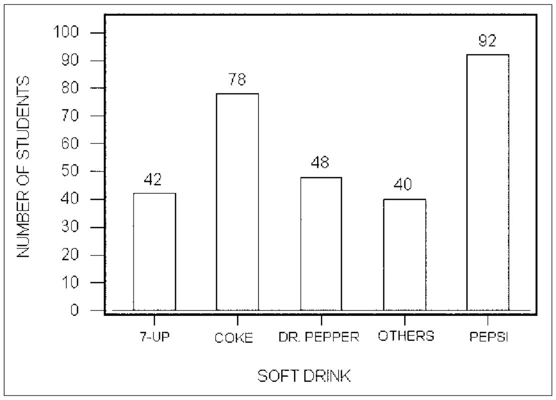


Fig. 1-6: Vertical bar chart for Example 1-5

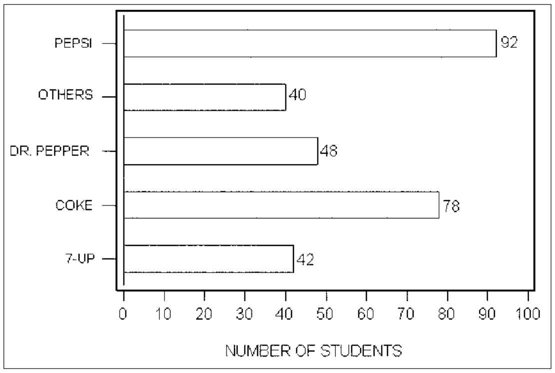


Fig. 1-7: Horizontal bar chart for Example 1-5

1-5 Histograms

Explanation of the term—histogram: A histogram is a graphical display of a frequency or a relative frequency distribution that uses classes and vertical bars (rectangles) of various heights to represent the frequencies.

Quick Tip

Histograms are useful when the data values are quantitative. A histogram gives an estimate of the shape of the distribution of the population from which the sample was taken.

Example 1-6: Display the data in **Example 1-3** with a histogram using seven classes.

Solution: A histogram with seven classes for the data is shown in **Fig. 1-8.**

The histogram shows the frequency count for each class, with each class having a width of 10.

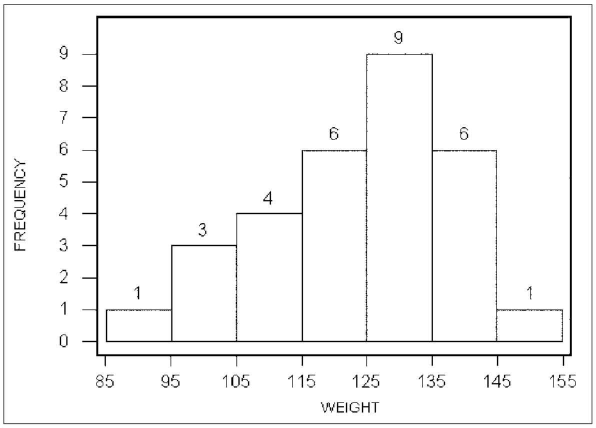


Fig. 1-8: Histogram for data in Example 1-3

Quick Tip

Observe from the histogram in *Fig. 1-8* that there is a frequency count of 1 for the interval 85–95, a frequency count of 3 for the interval 95–105, etc. From this information, we can construct the grouped frequency distribution given in *Example 1-3*.

Quick Tip

If the relative frequencies are plotted along the vertical axis to produce a relative frequency histogram, the shape of the resulting histogram will be the same as that of a histogram in which the frequencies were plotted along the vertical axis. This is true because the relative frequencies are obtained by dividing the frequency values by the total number of values in the data set.

1-6 Frequency Polygons

Explanation of the term—frequency polygon: A **frequency polygon** is a graph that displays the data using lines to connect points plotted for the frequencies. The frequencies represent the heights of the vertical bars in the histograms.

Note: A frequency polygon provides an estimate of the shape of the distribution of the population.

Example 1-7: Display a frequency polygon for the data in **Example 1-3.**

Solution: The display given in **Fig. 1-9** shows the frequency polygon superimposed on the histogram for **Example 1-6.**

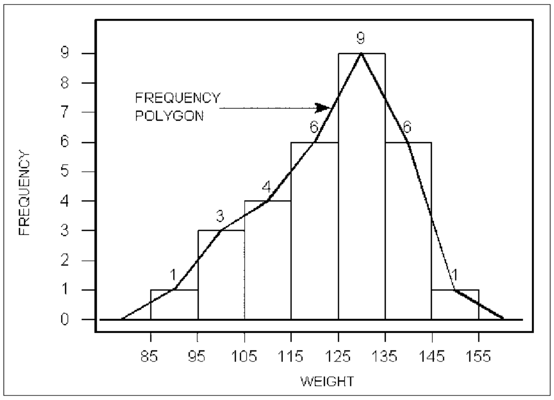


Fig. 1-9: Frequency polygon superimposed on the histogram for the data in Example 1-3

Quick Tip

Observe that the distribution is mound-shaped, with more of the values to the left of the peak. If this were a truly representative sample from the population, then one would expect that the distribution of the *population* of weights would have a similar shape. Also, observe that the line segments pass through the midpoints at the top of the rectangles and that the polygon is "tied down" to the horizontal axis at both ends. The points where the polygon is tied down correspond to the midpoints of the classes with zero frequency. In this case, the midpoints are 80 and 160. Midpoints of classes are called *class marks* or *class midpoints.*

1-7 Stem-and-Leaf Displays or Plots

Explanation of the term—stem-and-leaf plot: A **stem-and-leaf plot** is a data plot that uses part of a data value as the *stem* to form groups or classes and part of the data value as the *leaf.* A stem-and-leaf plot has an advantage over a grouped frequency distribution, since a stem-and-leaf plot retains the actual data by showing them in graphic form.

The next example will illustrate how a stem-and-leaf plot is constructed.

Example 1-8: Consider the following values: 96, 98, 107, 110, and 112.

(a) Use the last digit values as the leaves.

Solution: The data and the stems and leaves are shown in **Table 1-7.**

The corresponding stem-and-leaf plot is shown in **Table 1-8.**

(b) Use the last two digit values as the leaves.

Solution: The data and the stems and leaves are shown in **Table 1-9.**

The corresponding stem-and-leaf plot is shown in **Table 1-10.**

Table 1-7: Stems and Leaves for the Data in Example 1-8 with the Last Digit as the Leaves

DATA	STEM	LEAF
96	09	6
98	09	8
107	10	7
110	11	0
112	11	2

Table 1-8: Stem-and-Leaf Plot with the Last Digit as the Leaves for Example 1-8

STEM	LEAVES
09	6 8
10	7
11	0 2

Table 1-9: Stems and Leaves for the Data in Example 1-8 with the Last Two Digits as the Leaves

DATA	STEM	LEAF
96	0	96
98	0	98
107	1	07
110	1	10
112	1	12

Table 1-10: Stem-and-Leaf Plot with the Last Two Digits as the Leaves for Example 1-8

STEM	LEAVES
0	96 98
1	07 10 12

Example 1-9: A sample of the number of admissions to a psychiatric ward at a local hospital during the full phases of the moon is given below. Display the data using a stem-and-leaf plot with the leaves represented by the unit digits.

22	21	31	20	25	21	32	26	43	30	27
30	27	36	28	33	38	35	19	30	34	41

Solution: The stem-and-leaf display for the data is given in **Table 1-11.**

Table 1-11: Stem-and-Leaf Display for Example 1-9

STEM	LEAVES
1	9
2	0 1 1 2 5 6 7 7 8
3	0 0 0 1 2 3 4 5 6 8
4	1 3

1-8 Time Series Graphs

Data collected over a period of time can be displayed using a **time series graph.**

Explanation of the term—time series graph: A time series graph displays data that are observed over a given period of time. From the graph, one can analyze the behavior of the data over time.

Example 1-10: The data given are the number of hurricanes that occurred each year from 1970 to 1990. Display this information using a time series graph.

YEAR	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
Number	5	6	3	4	4	6	6	5	5	5	9	7	2	3	5	7	4	3	5	7	8

Solution: The time series plot for the data is shown in **Fig. 1-10.**

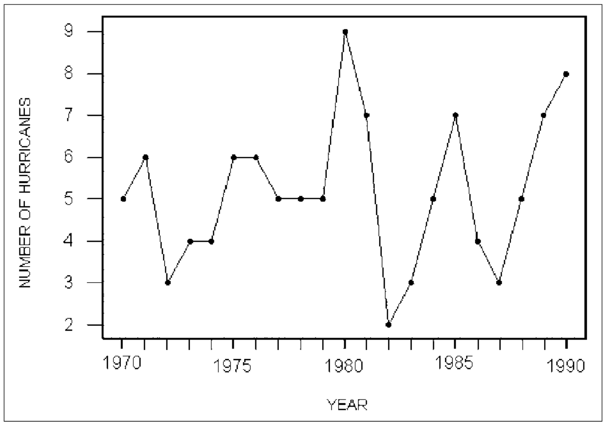


Fig. 1-10: Time series plot for Example 1-10

You can see from the graph that the largest number of hurricanes during the 20-year period occurred in 1980 and the smallest number occurred in 1982.

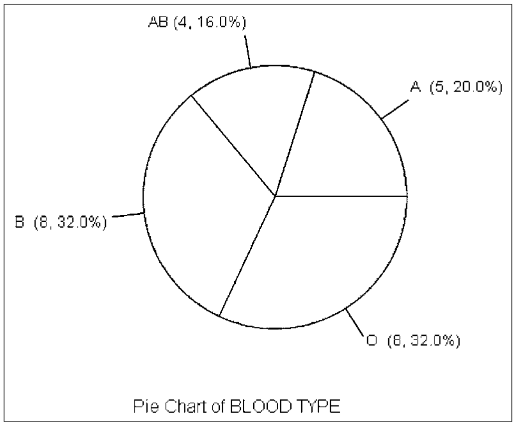


Fig. 1-11: Pie chart for the blood type data in **Example 1-1**

1-9 Pie Graphs or Pie Charts

Explanation of the term—pie graph (chart): A **pie graph** or **pie chart** is a circle that is divided into slices according to the percentage of the data values in each category.

A pie chart allows us to observe the proportions of sectors relative to the entire data set. It can be used to display either qualitative or quantitative data. However, categorical or qualitative data readily lend themselves to this type of graphical display because of the inherent categories in the data set.

Example 1-11: Present a pie chart for the blood type data given in **Example 1-1.**

Solution: The pie chart for the data is presented in **Fig. 1-11.**

Each slice of the pie chart represents a blood type category with its frequency count and the corresponding percentage for the count.

1-10 Pareto Charts

Explanation of the term—Pareto chart: A **Pareto chart** is a type of bar chart in which the horizontal axis represents categories of interest. When the bars are ordered from largest to smallest in terms of frequency counts for the categories, a Pareto chart can help you determine which of the categories make up the critical few and which are the insignificant many. A cumulative percentage line helps you judge the added contribution of each category.

Example 1-12: Display a Pareto chart for the soft drink example in **Example 1-5.**

Solution: The Pareto chart for the data is shown in **Fig. 1-12.**

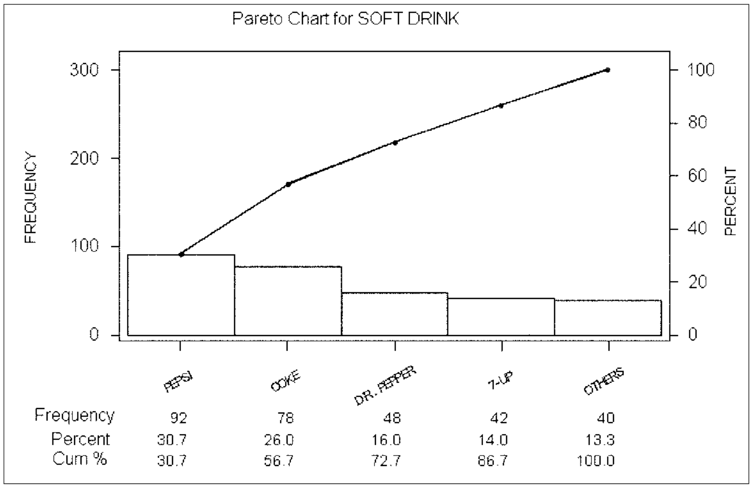


Fig. 1-12: Pareto chart for the data in **Example 1-5**

Observe that the categories have been ordered from the highest frequency to the lowest frequency.

Technology Corner

Calculators and Computer Software Packages

Statistical calculations usually involve large data sets that cannot be efficiently analyzed using pencil and paper. Modern technology eases this challenge through calculators and computers that can handle large numbers of arithmetic calculations. Most scientific calculators on the market today have statistical features that make statistical computations easy. Many specialized statistical software packages have features that perform quite sophisticated analysis of data.

Calculators are portable and relatively inexpensive. However, calculators have the drawback of not being able to deal with large amounts of data. There are many different types of calculators on the market today, so once you purchase one, in order to get familiar with its statistical features, you will need to consult the owner's manual for the calculator and practice using the appropriate keys. Specialized statistical software packages are usually more expensive than handheld calculators. These packages, however, can handle large amounts of data and usually perform a wider array of statistical calculations. Further, it is much easier to insert screen outputs from the software directly into word-processing documents. Also, it is relatively easy to transport data and analysis of data through electronic mail and floppy disks. For the more modern calculators, screen outputs can be imported into word-processing documents as well. The graphical features of statistical software packages are usually superior to the graphical features of calculators.

It's a Wrap

✔ All of the above graphical displays can be constructed with many of the software packages on the market today. As a matter of fact, all of the graphs in this chapter were constructed using the MINITAB for Windows software. A few other examples of software packages that can aid in the construction of the graphs in this chapter are Microsoft Excel, SAS, and SPSS for Windows.

Test Yourself

True/False Questions

1. A statistic is a characteristic of a population.
2. A parameter is a characteristic of a population.
3. Discrete data are data values that are measured over a continuous interval.
4. Continuous data are data that are measured over a given interval.
5. The amount of rainfall in your state for the last month is an example of discrete data.
6. The number of days it rained where you live during the last month is an example of discrete data.
7. Statistics is the science of collecting, classifying, presenting, and interpreting numerical data.
8. The subject of statistics can be broadly divided into two areas: descriptive statistics and inferential statistics.
9. A sample is the set of all possible data values for a given subject under consideration.
10. Descriptive statistics involves the collection, organization, and analysis of all data relating to some population or sample under study.
11. Statistics is concerned only with the collection, organization, display, and analysis of data.
12. A population is the set of all possible values for a given subject under consideration.
13. Inferential statistics involves making predictions or decisions about a sample from a population of values.
14. The frequency of a measurement is the number of times that measurement was observed.
15. The lower class limit for a given class is the smallest possible data value for that class.
16. The class mark for a class is the average of the upper and lower class limits for the given class.
17. The cumulative frequency of a class is the total of all class frequencies up to but not including the frequency of the present class.
18. The relative frequency for a given class is the total of all class frequencies before the class divided by the total number of entries.
19. The class midpoint for a class is computed from (upper limit – lower limit)/2, where the upper and lower limits are for the given class.

20. A frequency histogram and a relative frequency histogram for the same (grouped) frequency distribution will always have the same shape.
21. A frequency polygon for a set of data is obtained by connecting the class marks on the histogram displaying the set of data.
22. The choice of a single item from a group is called random if every item in the group has the same chance of being selected as every other item.
23. The class mark of a class is the midpoint between the lower limit of one class and the upper limit of the next class.
24. A population is part of a sample.
25. In stem-and-leaf displays, the trailing digits (digits to the right) are called the leaves.
26. The sum of the relative frequencies in a relative frequency distribution should always equal 1.
27. A population refers to the entire set of data values for a subject under consideration; a sample is a subset of the population.
28. A census is a sample of the entire population.

Completion Questions

1. A (parameter, statistic) __________ is a characteristic of a population.
2. A (parameter, statistic) __________ is a characteristic of a sample.
3. A sample of the entire population is called a (sample, population, census) __________.
4. Data that are counting numbers are called (discrete, continuous) __________ data.
5. Data that are measured over an interval are called (discrete, continuous) __________ data.
6. Drawing conclusions about a population from a sample is classified as (descriptive, inferential) __________ statistics.
7. (Descriptive, Inferential) __________ statistics is concerned with making predictions about an entire population based on information from a sample that was appropriately chosen from the population.
8. (Descriptive, Inferential) __________ statistics involves the collection, organization, summarization, and presentation of data.
9. A set of all possible data values for a subject under consideration is called a (sample, population) __________.
10. Class marks are the (lower limits, midpoints, upper limits) __________ of each class.
11. A subset of a population is called a (census, sample, small population)__________.
12. The lower class limit is the (smallest, largest) __________ possible data value for a class.
13. The (relative frequency, frequency, cumulative frequency) __________ is the number of occurrences of a measurement or data value.
14. The shape of the frequency distribution and the relative frequency distribution will always be (the same, different, skewed) __________.
15. Name three graphical methods by which you can display a set of data: (a) __________; (b) __________; (c) __________.
16. The (relative, cumulative) __________ frequency of a class is the total of all class frequencies up to and including the present class.

17. Data such as sex, eye color, race, etc., are classified as (quantitative, qualitative) _______ data.
18. The class mark of a class is defined to be the (average, minimum, maximum) _________ of the upper and lower limits of the class.
19. In a histogram there are no (gaps, values) _________ between the classes represented.
20. A pie chart or circle graph can be used to display (qualitative, quantitative, both types of) ___________ data.
21. In a stem-and-leaf plot, the trailing digits are called the (leaves, stems)___________ of the plot and the leading digits are called the (leaves, stems)___________ of the plot.
22. The choice of a single item from a group is called (random, biased)___________ if every item from a group has the same chance of being selected as every other item.

Multiple-Choice Questions

1. The section of statistics which involves the collection, organization, summarizing, and presentation of data relating to some population or sample is
   (a) inferential statistics.
   (b) descriptive statistics.
   (c) an example of a frequency distribution.
   (d) the study of statistics.
2. A subset of the population selected to help make inferences on a population is called
   (a) a population.
   (b) inferential statistics.
   (c) a census.
   (d) a sample.
3. A set of all possible data values for a subject under consideration is called
   (a) descriptive statistics.
   (b) a sample.
   (c) a population.
   (d) statistics.
4. The number of occurrences of a data value is called
   (a) the class limits.
   (b) the frequency.
   (c) the cumulative frequency.
   (d) the relative frequency.
5. A large collection of data may be condensed by constructing
   (a) classes.
   (b) a frequency polygon.
   (c) class limits.
   (d) a frequency distribution.
6. When constructing a frequency distribution for a small data set, it is wise to use
   (a) 5 to 20 classes.
   (b) 5 to 15 classes.

   (c) 5 to 10 classes.
   (d) less than 10 classes.

7. When constructing a frequency distribution for a large data set, it is wise to use
   (a) 5 to 20 classes.
   (b) 5 to 15 classes.
   (c) 5 to 10 classes.
   (d) less than 10 classes.

8. When straight-line segments are connected through the midpoints at the top of the rectangles of a histogram with the two ends tied down to the horizontal axis, the resulting graph is called
   (a) a bar chart.
   (b) a pie chart.
   (c) a frequency polygon.
   (d) a frequency distribution.

9. A questionnaire concerning satisfaction with the Financial Aid Office on campus was mailed to 50 students on a university campus. The 50 students in this survey are an example of a
   (a) statistic.
   (b) parameter.
   (c) population.
   (d) sample.

The following information relates to **Problems 10 to 15.**

The Love Your Lawn lawn care company is interested in the distribution of lawns in a certain subdivision with respect to size (in square feet) of the lawn. The following table shows the distribution of the size of the lawns in hundreds of square feet.

SIZE OF LAWN (100 SQUARE FEET)	NUMBER OF LAWNS
10–15	2
15–20	12
20–25	27
25–30	19
30–35	6
35–40	3

10. The class mark for the class 25–30 is
    (a) 24.5.
    (b) 29.5.
    (c) 4.
    (d) 27.5.

11. The relative frequency for the class 15–20 is
    (a) 0.2029.
    (b) 0.0290.
    (c) 0.1739.
    (d) 0.4058.

12. The lower class limit for the class 35–40 is
    (a) 34.5.
    (b) 35.
    (c) 37.
    (d) 39.5.
13. The upper class limit for the class 20–25 is
    (a) 24.5.
    (b) 25.
    (c) 24.
    (d) 22.
14. The cumulative frequency for the class 25–30 is
    (a) 41.
    (b) 9.
    (c) 19.
    (d) 60.
15. The cumulative relative frequency for the class 30–35 is
    (a) 0.8696.
    (b) 0.0870.
    (c) 0.1304.
    (d) 0.9565.
16. The graphical display with the relative frequencies along the vertical axis that may be constructed for quantitative data is
    (a) the pie chart.
    (b) the bar chart.
    (c) the histogram.
    (d) all of the above.
17. The cumulative relative frequency for a given class is defined to be
    (a) the proportion of values preceding the given class.
    (b) the proportion of values up to and including the given class.
    (c) the proportion of values for the given class.
    (d) the proportion of values below the given class.
18. A property of a frequency polygon is that
    (a) a histogram is always needed in the construction of the polygon.
    (b) the polygon is made up of line segments.
    (c) the end points of the polygon need not be tied down to the horizontal axis at both ends.
    (d) the polygon can be constructed on a pie chart.
19. You are given that the total number of observed values in a frequency distribution is 50 and the frequency of a given class 25–30 is 10. Also, the cumulative frequency of all classes above this given class is 40. The cumulative frequency for this class is
    (a) 10.
    (b) 50.

(c) 40.
(d) 30.

20. If a class in a frequency distribution for a sample of 50 has a frequency of 5, the cumulative relative frequency for this class
    (a) is 0.1000.
    (b) is 0.9000.
    (c) is 0.1111.
    (d) cannot be determined from the given information.

21. If the first five classes of a frequency distribution have a cumulative frequency of 50 from a sample of 58, the sixth and last class must have a frequency count of
    (a) 58.
    (b) 50.
    (c) 7.
    (d) 8.

The following information relates to **Problems 22 to 28.** *Hint:* Read the examination scores distribution from smallest value to largest value.

The table below shows the distribution of scores on a final elementary statistics examination for a large section of students.

CLASSES FOR EXAM SCORES	NUMBER OF STUDENTS
90 and over	5
80–90	12
70–80	40
60–70	18
50–60	13
40–50	6
Under 40	6

22. The class width is
    (a) 9.
    (b) 10.
    (c) 7.
    (d) 1.

23. The class mark for the class 40–50 is
    (a) 39.5.
    (b) 49.5.
    (c) 45.
    (d) 9.

24. The relative frequency for the class 80–90 is
    (a) 0.1700.
    (b) 0.0500.
    (c) 0.8300.
    (d) 0.1200.

25. The lower class limit for the class 50–60 is
    (a) 49.5.
    (b) 50.
    (c) 59.
    (d) 59.5.
26. The upper class limit for the class 70–80 is
    (a) 69.5.
    (b) 70.
    (c) 80.
    (d) 79.5.
27. The cumulative frequency for the class 60–70 is
    (a) 18.
    (b) 57.
    (c) 43.
    (d) 12.
28. The cumulative relative frequency for the class 50–60 is
    (a) 0.8800.
    (b) 0.1300.
    (c) 0.7500.
    (d) 0.2500.
29. Can a frequency distribution have overlapping classes?
    (a) Sometimes
    (b) No
    (c) Yes
    (d) All of the above
30. An organization of observed data into tabular form in which classes and frequencies are used is called
    (a) a bar chart.
    (b) a pie chart.
    (c) a frequency distribution.
    (d) a frequency polygon.
31. Given the following stem-and-leaf diagram:
    1 | 0 3
    2 | 2 2 4
    3 | 1 2 3 3 3
    4 | 1 1 2 2 2 2 5 6
    5 | 3 3 5 6
    6 | 2 4
    7 | 3

    The number that occurred the most is
    (a) 2.
    (b) 42.

(c) 33.
(d) 3.

Further Exercises

If possible, you can use any technology available to help you solve the following questions.

1. The at-rest pulse rates for 16 athletes at a meet are
   67 57 56 57 58 56 54 64 53 54 54 55 57 68 60 58
   (a) Construct a relative frequency distribution for this data set using classes 50–55, 55–60, ...
   (b) Construct a histogram for this set of data using the distribution in part (a).
2. The speeds (in mph) of 16 cars on a highway were observed to be
   58 56 60 57 52 54 54 59 63 54 53 54 58 56 57 67
   (a) Construct a relative frequency distribution for this data set using classes 52–55, 55–58, ...
   (b) Construct a stem-and-leaf plot for the data set.
3. The starting incomes for mathematics majors at a particular university were recorded for five years and are summarized in the following table:

STARTING SALARY (IN $1000)	FREQUENCY
10–15	3
15–20	5
20–25	10
25–30	7
30–35	1

   (a) Construct a histogram for the data.
   (b) Construct a table with the relative frequencies and the cumulative relative frequencies.
4. The following frequency distribution shows the distances to campus (in miles) traveled by 30 commuter students:

DISTANCE (IN MILES)	FREQUENCY
35–40	8
40–45	13
45–50	6
50–55	3

   For the class 40–45, find the following:
   (a) Lower class limit
   (b) Upper class limit
   (c) Class width
   (d) Class mark

(e) Cumulative frequency
(f) Relative frequency
(g) Cumulative relative frequency

ANSWER KEY

True/False Questions

1. F 2. T 3. F 4. T 5. F 6. T 7. T 8. T 9. F 10. T 11. F
12. T 13. F 14. T 15. T 16. T 17. F 18. F 19. F 20. T 21. T
22. T 23. F 24. F 25. T 26. T 27. T 28. T

Completion Questions

1. parameter 2. statistic 3. census 4. discrete 5. continuous 6. inferential
7. Inferential 8. Descriptive 9. population 10. midpoints 11. sample
12. smallest 13. frequency 14. the same 15. bar chart, histogram, pie chart, frequency polygon, stem-and-leaf plot (any three) 16. cumulative 17. qualitative
18. average 19. gaps 20. both types of 21. leaves, stems 22. random

Multiple-Choice Questions

1. (b) 2. (d) 3. (c) 4. (b) 5. (d) 6. (c) 7. (a) 8. (c) 9. (d)
10. (d) 11. (c) 12. (b) 13. (b) 14. (d) 15. (d) 16. (c) 17. (b) 18. (b)
19. (b) 20. (d) 21. (d) 22. (b) 23. (c) 24. (d) 25. (b) 26. (c)
27. (c) 28. (d) 29. (b) 30. (c) 31. (b)

CHAPTER 2

Data Description—Numerical Measures of Central Tendency for Ungrouped Univariate Data

Do I Need to Read This Chapter?

You should read this chapter if you need to review or to learn about

- Numerical values that measure central tendencies of a numerical data set
- How to compute these measures and investigate their properties

You will recognize that these measures deal with only one property of the data set: the centralness. Thus, you will need to combine these measures with other properties of the data set in order to fully describe it. Other properties for univariate data will be investigated in future chapters.

Get Started

A measure of central tendency for a collection of data values is a number that is meant to convey the idea of centralness for the data set. The most commonly used measures of central tendency for sample data are the mean, the median, and the mode. These measures are discussed in this chapter.

2-1 The Mean

Explanation of the term—mean: The **mean** of a set of numerical values is the average of the set of values.

Note: In the explanation of the mean, the numerical values can be population values or sample values. Hence, we can compute the mean for either population values or sample values.

Explanation of the term—population mean: If the values are from an entire population, then the mean of the values is called a **population mean.** It is usually denoted by μ (read as "mu").

Explanation of the term—sample mean: If the values are from a sample, then the mean of the values is called a **sample mean.** It is denoted by $\bar{x}$ (read as "x-bar").

Example 2-1: What is the mean for the following sample values?

3 8 6 14 0 –4 0 12 –7 0 –10

Solution: The sample mean is obtained as

$$\bar{x} = \frac{3+8+6+14+0+(-4)+0+12+(-7)+0+(-10)}{11} = 2$$

That is, the value of the sample mean is 2.

Quick Tip

When the word *mean* or *average* is used in everyday conversation, it has come to represent a typical value or the center of a set of values. Because of this, the mean is called a measure of central tendency.

Question: **Why do we use the mean as a measure of the center of a set of values?**

The following discussion will give an insight into the question. First, **Fig. 2-1** shows a plot of the data points along with the sample mean.

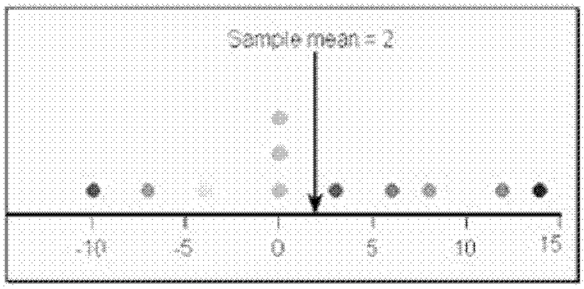


Fig. 2-1: Plot of data values for Example 2-1

Next, we compute the deviation from the sample mean for each value in the data set. That is, we compute $(x - \bar{x})$ for each value x. These deviations are given in **Table 2-1.**

Table 2-1: Deviations from the Mean for Values in Example 2-1

DATA	DEVIATIONS
3	1
8	6
6	4
14	12
0	–2
–4	–6
0	–2
12	10
–7	–9
0	–2
–10	–12

Next, a plot of the deviations from the sample mean is displayed in **Fig. 2-2.**

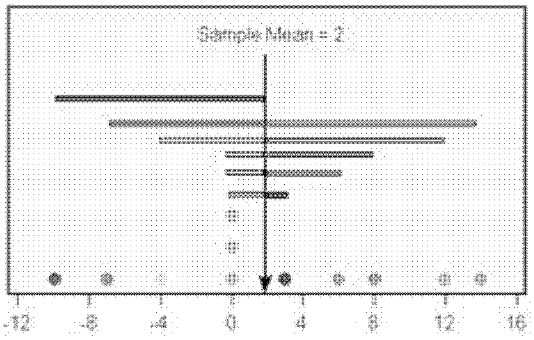


Fig. 2-2: Deviations from the sample mean for Example 2-1

When the deviations on the left and on the right of the sample means are added, disregarding the sign of the values, we see that when the "balancing point" is the sample mean, then these sums are equal in absolute value. Here in **Example 2-1,** the sum of the deviations to the right of the mean is 33. The sum of the deviations to the left of the mean is –33. However, we use the absolute value of these negative deviations; that is, we use +33. This is depicted in **Fig. 2-3.**

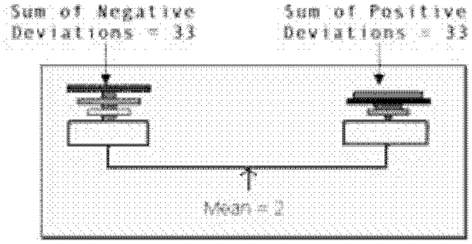


Fig. 2-3: Balanced deviations

Thus, *the mean is that central point where the sum of the negative deviations (absolute value) from the mean and the sum of the positive deviations from the mean are equal.* This is why the mean is considered a measure of central tendency.

Quick Tip

When a data set has a large number of values, we sometimes summarize it as a frequency table. The frequencies represent the number of times each value occurs.

The next example shows us how to find the mean of a set of values when the data are summarized in a frequency table.

Example 2-2: Find the mean for the following frequency table, using four decimal places.

x VALUES x_i	FREQUENCY f_i
20	2
29	4
30	4
39	3
44	2

Solution: Observe that the value of 20 has a frequency count of 2, so the total or sum can be written as 20 + 20 or 2 × 20. We can do the same for each of the values and its corresponding frequency count. The total number of values in the table is 15, which is the sum of the frequency values. Thus, we can compute the mean for the frequency distribution as

$$\bar{x} = \frac{2 \times 20 + 4 \times 29 + 4 \times 30 + 3 \times 39 + 2 \times 44}{2 + 4 + 4 + 3 + 2} = 32.0667$$

2-2 The Median

The next measure of central tendency we will consider is the **median.**

Explanation of the term—median: The median of a numerical data set is the numerical value in the middle when the data set is arranged in order.

Quick Tips

1. When the number of values in the data set is *odd,* the median will be the middle value in the ordered array.
2. When the number of values in the data set is *even,* the median will be the average of the two middle values in the ordered array.

Example 2-3: What is the median for the following sample values?

3 8 6 14 0 –4 2 12 –7 –1 –10

Solution: First of all, we need to arrange the data set in order. The ordered set is as follows:

–10 –7 –4 –1 0 2 3 6 8 12 14
↑
6th

Since the number of values is odd, the median will be the middle value in the ordered set. Thus, the median will be found in the sixth position, since we have a total of 11 values.

That is, the value of the median is 2.

Question: **Why does the middle number in an ordered data set measure central tendency?**

The following discussion will give an insight into the question. **Figure 2-4** shows a plot of the data points and the location of the sample median.

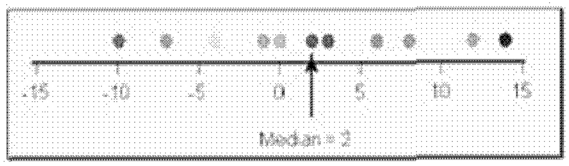


Fig. 2-4: Plot of data points for Example 2-3

A list of the values that are above the median and below the median is given in **Table 2-2.**

Table 2-2: List of Values That Are Above or Below the Median for Example 2-3

DATA	DEVIATIONS	DATA	DEVIATIONS
3	Above	2	Neither
8	Above	12	Above
6	Above	–7	Below
14	Above	–1	Below
0	Below	–10	Below
–4	Below		

When the values from above and below the median are counted, we see that if the "balancing point" is the sample median, then the number of values above the median balances (equals) the number of values below the median. This is depicted in **Fig. 2-5.**

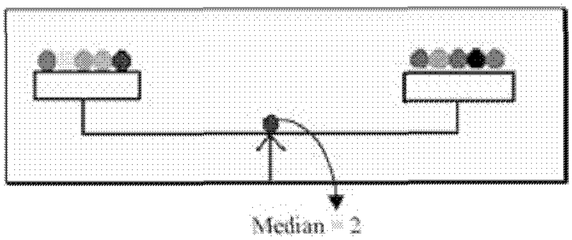


Fig. 2-5: Median as a balancing point for data values in Example 2-3

Observe that there are the same number of values above the median as there are below the median. This is why the median is considered a measure of central tendency.

Example 2-4: Find the median for the ages of the following eight college students:

23 19 32 25 26 22 24 20

Solution: First order the values. The ordered array is

19 20 22 23 24 25 26 32

Since there is an even number of ages, the median will be the average of the two middle numbers. Since the two middle numbers are located in the fourth and fifth positions, the median $= \frac{23+24}{2} = 23.5$.

2-3 The Mode

Explanation of the term—mode: The **mode** of a numerical data set is the most frequently occurring value in the data set.

Quick Tips

1. If all the elements in the data set have the same frequency of occurrence, then the data set is said to have *no mode.*
2. If the data set has one value that occurs more frequently than the rest of the values, then the data set is said to be *unimodal.*
3. If two elements of the data set are tied for the highest frequency of occurrence, then the data set is said to be *bimodal.*

Example 2-5: What is the mode for the following sample values?

3 5 1 4 2 9 6 10

Solution: We see from **Fig. 2-6** that each value occurs with a frequency of 1. Thus, the data set has no mode.

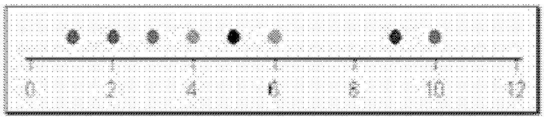


Fig. 2-6: Plot of data values for Example 2-5

Example 2-6: What is the mode for the following sample values?

3 5 1 4 2 9 6 10 5 3 4 3 9 3 6 1

Solution: **Figure 2-7** shows a plot of the data values.

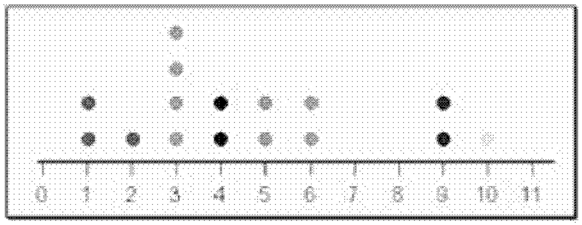


Fig. 2-7: Plot of data values for Example 2-6

Observe that the value of 3 occurs with the highest frequency. Thus, the value of the mode is 3, and this data set is unimodal.

Example 2-7: What is the mode for the following sample values?

6 10 5 3 4 3 9 3 6 1 6

Solution: **Figure 2-8** shows a plot of the data values.

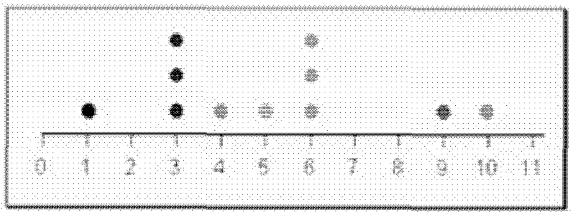


Fig. 2-8: Plot of data values for Example 2-7

Observe that the value of 3 and the value of 6 occur with the highest but equal frequency. Thus, the values of the mode are 3 and 6, and this data set is bimodal.

2-4 Shapes (Skewness)

The three most important shapes of frequency distributions are positively skewed, negatively skewed, and symmetrical.

Positively Skewed Distribution

In a positively skewed distribution, most of the data values fall to the left of the mean, and the "tail" of the distribution is to the right. In addition, the mean is to the right of the median, and the mode is to the left of the median. These properties are depicted in **Fig. 2-9.**

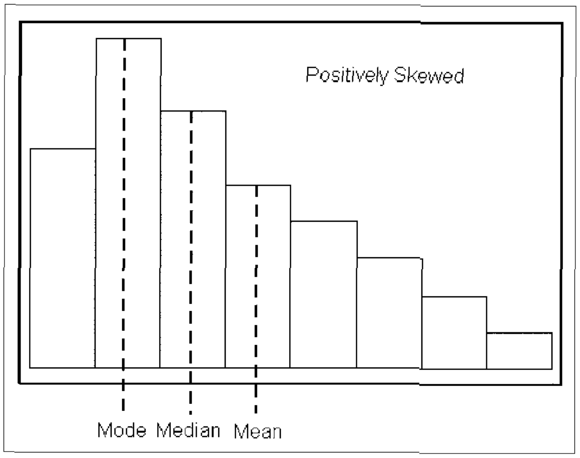


Fig. 2-9: Positively skewed distribution

Negatively Skewed Distribution

In a negatively skewed distribution, most of the data values fall to the right of the mean, and the tail of the distribution is to the left. In addition, the mean is to the left of the median, and the mode is to the right of the median. These properties are depicted in **Fig. 2-10.**

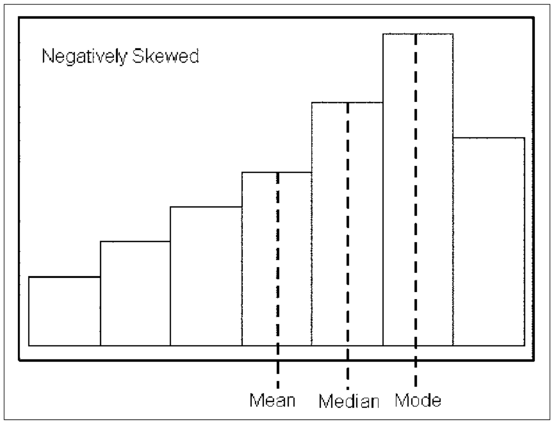


Fig. 2-10: Negatively skewed distribution

Symmetrical Distribution

In a symmetrical distribution, the data values are evenly distributed on both sides of the mean. Also, when the distribution is unimodal, the mean, median, and mode are all equal to one another and are located at the center of the distribution. These properties are depicted in **Fig. 2-11.**

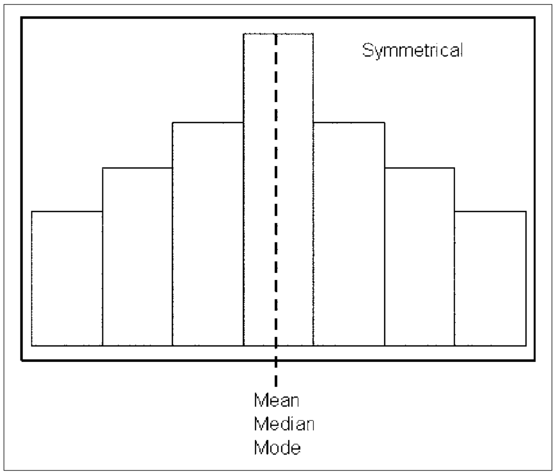
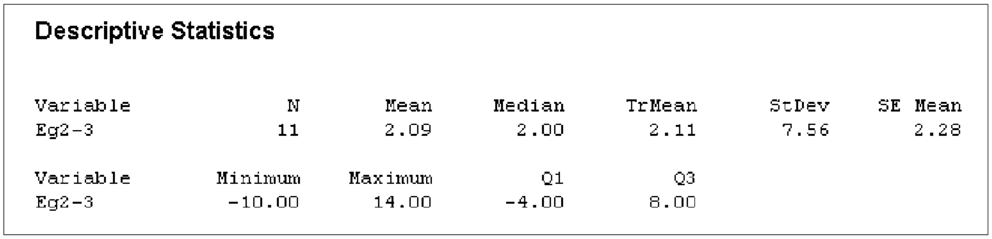


Fig. 2-11: Symmetrical distribution

Technology Corner

All of the concepts discussed in this chapter can be illustrated using most statistical software packages. All scientific and graphical calculators will aid directly in the computations. In addition, some of the newer calculators, such as the TI-83, will allow you to compute the mean and the median directly. If you own a calculator, you should consult the manual to determine what statistical features are included.

Illustration: **Figure 2-12** shows the descriptive statistics computed by the MINITAB software. **Figure 2-13** shows the 1-Var Stats (descriptive statistics) computed by the TI-83 calculator. The data used, in both cases, were from **Example 2-3.** Observe that MINITAB has given the value of the mean to two decimal places, while the TI-83 calculator gives the value to nine decimal places. The median using both technologies is 2. Observe that the mode is not displayed in either figure. One will have to use other features of the technologies to obtain the value of the mode. There are other descriptive statistics in the outputs that we will encounter later in the text.

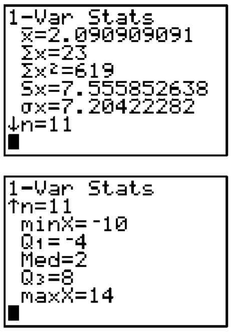


```
Descriptive Statistics

Variable            N       Mean     Median     TrMean      StDev    SE Mean
Eg2-3              11       2.09       2.00       2.11       7.56       2.28

Variable      Minimum    Maximum         Q1         Q3
Eg2-3          -10.00      14.00      -4.00       8.00
```

Fig. 2-12: MINITAB descriptive statistics output for Example 2-3



```
1-Var Stats
 x̄=2.090909091
 Σx=23
 Σx²=619
 Sx=7.555852638
 σx=7.20422282
↓n=11
```

```
1-Var Stats
↑n=11
 minX=⁻10
 Q₁=⁻4
 Med=2
 Q₃=8
 maxX=14
```

Fig. 2-13: TI-83 1-Var Stats output for Example 2-3

Quick Tip

Unlike the median and the mode, the mean is sensitive to a change in any of the values of the data set. If the same constant is added to each value in the data set, the mean of the data set will increase by the same amount. If each value in the data set is multiplied by the same constant, the mean of the data set will also be multiplied by that constant.

It's a Wrap

The three most commonly used measures of central tendency for numeric data are the

✓ Mean
✓ Median
✓ Mode

Care should always be taken when using these measures of central tendency.

Test Yourself

True/False Questions

1. The mean of a set of data always divides the data set such that 50 percent of the values lie above the mean and 50 percent lie below the mean.
2. The mode is a measure of variability.
3. The median of a set of data values is that value that occurs the most.
4. The mean is not equal to the median in a symmetrical distribution.
5. Of the mean, the median, and the mode of a data set, the mean is most influenced by an outlying value in the data set.
6. If the number of observations in a data set is odd, the median cannot be accurately found, but rather is approximated.
7. A data set with more than one mode is said to be bimodal.
8. The sum of the deviations from the mean for any data set is always 0.
9. For a negatively skewed distribution, the tail is to the right of the mean.
10. For a positively skewed distribution, the mode is less than the median, and the median is less than the mean.

Completion Questions

1. If the number of measurements in a data set is odd, the median is the __________ value when the data set is ordered from the smallest value to the largest value.
2. If the number of measurements in a data set is even, the median is the __________ of the two __________ values when the data set is ordered from the smallest value to the largest value.
3. The (mean, median, mode) __________ for a set of data is the value in the data set that occurs most frequently.
4. Two measures of central tendency are the __________ and the __________ .
5. For a symmetrical distribution, the mean, mode, and median are all (equal to, different from) __________ one another.
6. For a negatively skewed distribution, the mean is (smaller, greater) __________ than the median and the mode.
7. For a positively skewed distribution, the tail of the distribution is to the (right, left) __________ of the distribution.
8. For a positively skewed distribution, the median is (smaller, larger) __________ than the mode.

Multiple-Choice Questions

1. A student has seven statistics books open in front of him. The page numbers are as follows: 231, 423, 521, 139, 347, 400, 345. The median for this set of numbers is

   (a) 139.
   (b) 347.
   (c) 346.
   (d) 373.5.

2. A cyclist recorded the number of miles per day that she cycled for 5 days. The recordings were as follows: 13, 10, 12, 10, 11. The mean number of miles she cycled per day is
   (a) 13.
   (b) 11.
   (c) 10.
   (d) 11.2.
3. An instructor recorded the following quiz scores (out of a possible 10 points) for the 12 students present: 7, 4, 4, 7, 2, 9, 10, 6, 7, 3, 8, 5. The mode for this set of scores is
   (a) 9.5.
   (b) 7.
   (c) 6.
   (d) 3.
4. It is stated that more students are purchasing graphing calculators than any other type of calculator. Which measure is being used here?
   (a) Mean
   (b) Median
   (c) Mode
   (d) None of the above
5. Which of the following is not a measure of central tendency?
   (a) Mode
   (b) Variability
   (c) Median
   (d) Mean

Use the following frequency distribution for **Problems 6 to 8.**

x VALUES	FREQUENCY
20	2
29	4
30	4
39	3
44	2

6. The mean of the distribution is
   (a) 32.4.
   (b) 30.
   (c) 39.
   (d) 32.07.
7. The median of the distribution is
   (a) 4.

(b) 30.
(c) 29.5.
(d) 34.5.

8. The mode of the distribution is
(a) 29.
(b) 30.
(c) 29 and 30.
(d) none of the above.

9. Given the following data set:
12 32 45 14 24 31

The total deviation from the mean for the data values is
(a) 0.
(b) 26.3333.
(c) 29.5.
(d) 12.

10. The most frequently occurring value in a data set is called the
(a) spread.
(b) mode.
(c) skewness.
(d) maximum value.

11. A single numerical value used to describe a characteristic of a sample data set, such as the sample median, is referred to as a
(a) sample parameter.
(b) sample median.
(c) population parameter.
(d) sample statistic.

12. Which of the following is true for a positively skewed distribution?
(a) Mode = Median = Mean
(b) Mean < Median < Mode
(c) Mode < Median < Mean
(d) Median < Mode < Mean

13. Which of the following would be affected the most if there is an extremely large value in the data set?
(a) The mode
(b) The median
(c) The frequency
(d) The mean

14. If the number of values in a data set is even, and the numbers are ordered, then
(a) the median cannot be found.
(b) the median is the average of the two middle numbers.
(c) the median, mode, and mean are equal.
(d) none of the above answers are correct.

15. What type of distribution is described by the following information?
    Mean = 5.5 median = 5.3 mode = 4.4
    (a) Negatively skewed
    (b) Symmetrical
    (c) Bimodal
    (d) Positively skewed
16. What type of distribution is described by the following information?
    Mean = 56 median = 58.1 mode = 63
    (a) Negatively skewed
    (b) Symmetrical
    (c) Bimodal
    (d) Positively skewed
17. The mean of a set of data is the value that represents
    (a) the middle value of the data set.
    (b) the most frequently observed value.
    (c) the mean of the squared deviations of the values from the mean.
    (d) the arithmetic average of the data values.
18. The median of an ordered set of data is the value that represents
    (a) the middle or the approximate middle value of the data set.
    (b) the most frequently observed value.
    (c) the mean of the squared deviations of the values from the mean.
    (d) the arithmetic average of the data values.
19. Given the following data set: 4, 3, 7, 7, 8, 7, 4, 8, 6. What is the mean value?
    (a) 4
    (b) 5
    (c) 6
    (d) 7
20. Given the following data set: 3, 2, 7, 7, 8, 7, 3, 8, 5. What is the median value?
    (a) 5
    (b) 6
    (c) 7
    (d) 8
21. Given the following data set: 4, 5, 7, 7, 8, 6, 5, 8, 7. What is the mode?
    (a) 4
    (b) 5
    (c) 6
    (d) 7
22. A sample of 10 students was asked by their instructor to record the number of hours they spent studying for a given exam from the time the exam was announced in class. The following data values were the recorded number of hours:

    12 15 8 9 14 8 17 14 8 15

    The median number of hours spent studying for this sample is

(a) 10.
(b) 11.
(c) 12.
(d) 13.

23. The numbers of minutes spent in the computer lab by 20 students working on a project are given below:

Numbers of Minutes

30 | 0 2 5 5 6 6 6 8
40 | 0 2 2 5 7 9
50 | 0 1 3 5
60 | 1 3

The median for this data set is
(a) 400.
(b) 402.
(c) 405.
(d) 407.

24. The numbers of minutes spent in the computer lab by 20 students working on a project are given below:

Numbers of Minutes

30 | 0 2 5 5 6 6 6 8
40 | 0 2 2 5 7 9
50 | 0 1 3 5
60 | 1 3

The mode for this data set is
(a) 305.
(b) 402.
(c) 306.
(d) 300.

25. The numbers of minutes spent in the computer lab by 20 students working on a project are given below:

Numbers of Minutes

30 | 0 2 5 5 6 6 6 8
40 | 0 2 2 5 7 9
50 | 0 1 3 5
60 | 1 3

The mean for this data set is
(a) 306.0.
(b) 402.0.
(c) 403.8.
(d) 450.0.

26. A set of exam scores is given below:

Exam Scores

4 | 5 6 8
5 | 3 4 5 6 9

6 | 2 3 5 6 6 9 9
7 | 0 1 1 3 3 4 5 5 5 7 8
8 | 1 1 2 3 6 9
9 | 3 5 7 8

The mode for this data set is

(a) 75.
(b) 78.
(c) 45.
(d) 98.

Further Exercises

If possible, you can use any technology available to help you solve the following questions.

1. The at-rest pulse rates for 16 athletes at a meet are
   67 57 56 57 58 56 54 64 53 54 54 55 57 68 60 58
   Find the median, mode, and mean for this set of data.
2. The speeds (in mph) of 16 cars on a highway were observed to be
   58 56 60 57 52 54 54 59 63 54 53 54 58 56 57 67
   Find the mean, mode, and median for this set of data.
3. Estimate the mean for the following frequency distribution. *Hint:* Use the class marks as the actual observed values in each class.

CLASS	FREQUENCY
10–15	2
15–20	4
20–25	4
25–30	3
30–35	2

4. Find the mean, median, and mode for the following examination scores.

   Exam Scores

   4 | 5 6 8
   5 | 3 4 5 6 9
   6 | 2 3 5 6 6 9 9
   7 | 0 1 1 3 3 4 5 5 5 7 8
   8 | 1 1 2 3 6 9
   9 | 3 5 7 8
5. The following frequency distribution shows the scores on the exit examination for statistics majors at a four-year college for a given year.
   98 75 85 97 80 87 97 60 83 90
   Find the mean, mode, and median for this set of data.
6. The starting incomes for mathematics majors at a particular university were recorded for five years and are summarized in the following table:

STARTING SALARY (IN $1000)	FREQUENCY
10–15	3
15–20	5
20–25	10
25–30	7
30–35	1

(a) Construct a histogram for the data.
(b) Compute an approximate value for the mean by using the class mark values.

7. The numbers of 30-second radio advertising spots purchased by each of the 25 members of a local restaurant association last year are given below:

Numbers of 30-Second Spots

1 | 1 1 1 2 3 3 3 4 5 6 7
2 | 3 4 4 5 6 6
3 | 1 1 2 2 2 3
4 | 0 0 1

(a) Find the median.
(b) Find the mode.
(c) Find the mean.
(d) Describe the shape of the distribution.
(e) Construct a histogram for the data set.

ANSWER KEY

True/False Questions

1. F 2. F 3. F 4. F 5. T 6. F 7. F 8. T 9. F 10. T

Completion Questions

1. middle 2. average, middle 3. mode 4. mean, median, mode (any two)
5. equal to 6. smaller 7. right 8. larger

Multiple-Choice Questions

1. (b) 2. (d) 3. (b) 4. (c) 5. (b) 6. (d) 7. (b) 8. (c) 9. (a)
10. (b) 11. (d) 12. (c) 13. (d) 14. (b) 15. (d) 16. (a) 17. (d)
18. (a) 19. (c) 20. (c) 21. (d) 22. (d) 23. (b) 24. (c) 25. (c)
26. (a)