CHAPTER 3

NUMERICAL DESCRIPTIVE MEASURES

Prem Mann, *Introductory Statistics*, *7/E*

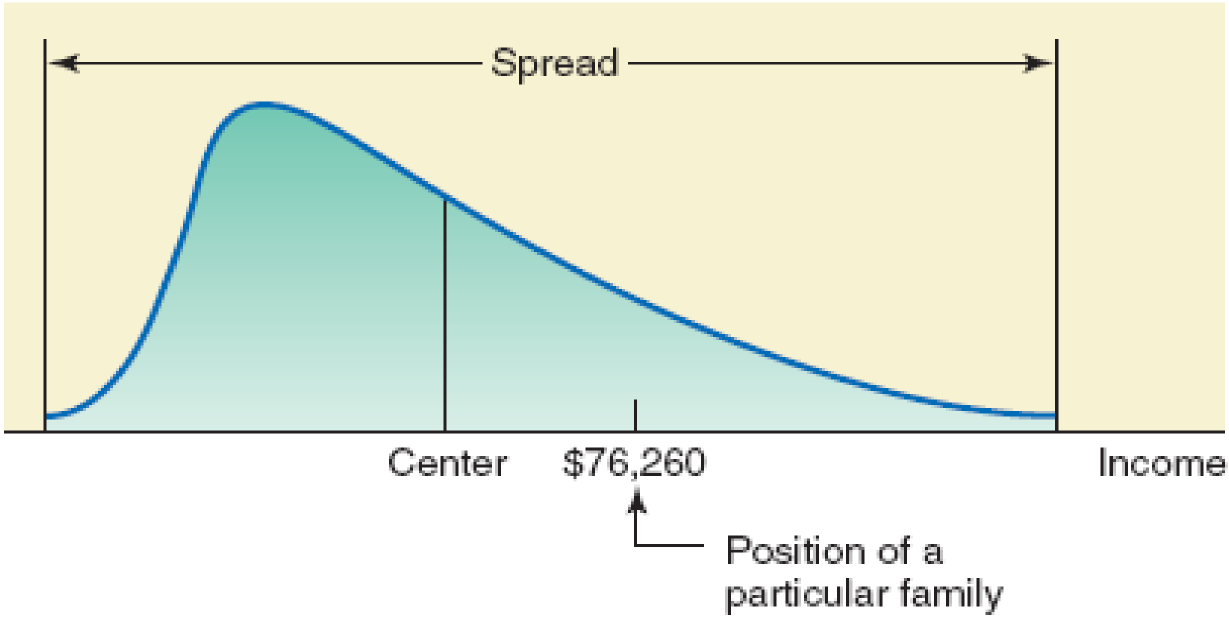
Opening Example

During the 2008 season, among all baseball teams, the New York Yankees drew the highest average of spectators to the games. (See Case Study 3-1). Despite baseball being "America's National Pastime," attendance at Major League Baseball games varies from team to team. What may attract fans to baseball fields? Is it the number of championships won, market size, fan loyalty, or just love for the team or game?

3.1 Measures of Central Tendency for Ungrouped Data

- Mean
- Median
- Mode
- Relationships among the Mean, Median, and Mode

Figure 3.1



Prem Mann, *Introductory Statistics*, *7/E*

Mean

The ***mean for ungrouped data*** is obtained by dividing the sum of all values by the number of values in the data set. Thus,

Mean for population data: $$\mu = \frac{\sum x}{N}$$

Mean for sample data: $$\bar{x} = \frac{\sum x}{n}$$

Example 3-1

Table 3.1 lists the total sales (rounded to billions of dollars) of six U.S. companies for 2008.

Table 3.1 2008 Sales of Six U.S. Companies

Company	Total Sales (billions of dollars)
General Motors	149
Wal-Mart Stores	406
General Electric	183
Citigroup	107
Exxon Mobil	426
Verizon Communication	97

Find the 2008 mean sales for these six companies.

Example 3-1: Solution

$$\sum x = x_1 + x_2 + x_3 + x_4 + x_5 + x_6$$

$$= 149 + 406 + 183 + 107 + 426 + 97 = 1368$$

$$\bar{x} = \frac{\sum x}{n} = \frac{1368}{6} = 228 = \$228 \text{ Billion}$$

Thus, the mean 2008 sales of these six companies was 228, or $228 billion.

Example 3-2

The following are the ages (in years) of all eight employees of a small company:

53 32 61 27 39 44 49 57

Find the mean age of these employees.

Example 3-2: Solution

$$\mu = \frac{\sum x}{N} = \frac{362}{8} = 45.25 \text{ years}$$

Thus, the mean age of all eight employees of this company is 45.25 years, or 45 years and 3 months.

Example 3-3

Table 3.2 lists the total philanthropic givings (in million dollars) by six companies during 2007.

Corporation	Money Given in 2007 (millions of dollars)
CVS	22.4
Best Buy	31.8
Staples	19.8
Walgreen	9.0
Lowe's	27.5
Wal-Mart	337.9

Example 3-3

Notice that the charitable contributions made by Wal-Mart are very large compared to those of other companies. Hence, it is an outlier. Show how the inclusion of this outlier affects the value of the mean.

Example 3-3: Solution

If we do not include the charitable givings of Wal-Mart (the outlier), the mean of the charitable contributions of the fiver companies is

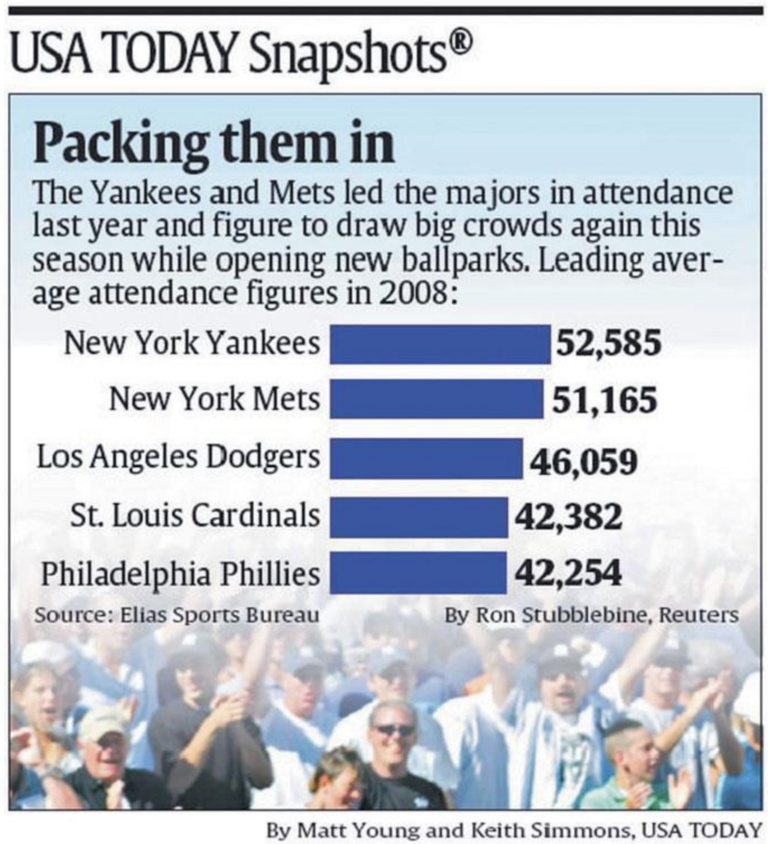
$$\text{Mean} = \frac{22.4 + 31.8 + 19.8 + 9.0 + 27.5}{5} = \$22.1 \text{ million}$$

Example 3-3: Solution

Now, to see the impact of the outlier on the value of the mean, we include the contributions of Wal-Mart and find the mean contributions of the six companies. This mean is

$$\textbf{Mean} = \frac{22.4 + 31.8 + 19.8 + 9.0 + 27.5 + 337.9}{6} = \$74.73 \text{ million}$$

Case Study 3-1 Average Attendance at Baseball Games



Median

Definition

The ***median*** is the value of the middle term in a data set that has been ranked in increasing order.

The calculation of the median consists of the following two steps:

1. Rank the data set in increasing order.
2. Find the middle term. The value of this term is the median.

Example 3-4

The following data give the prices (in thousands of dollars) of seven houses selected from all houses sold last month in a city.

312 257 421 289 526 374 497

Find the median.

Example 3-4: Solution

First, we rank the given data in increasing order as follows:

257   289   312   374   421   497   526

Since there are seven homes in this data set and the middle term is the fourth term,

257   289   312   374   421   497   526

$\uparrow$

Median

Thus, the median price of a house is 374.

Example 3-5

Table 3.3 gives the 2008 profits (rounded to billions of dollars) of 12 companies selected from all over the world.

Table 3.3 Profits of 12 Companies for 2008

Find the median of these data.

Company	2008 Profits (billions of dollars)
Merck & Co	8
IBM	12
Unilever	7
Microsoft	17
Petrobras	14
Exxon Mobil	45
Lukoil	10
AT&T	13
Nestlé	17
Vodafone	13
Deutsche Bank	9
China Mobile	11

Example 3-5: Solution

First we rank the given profits as follows:

7 8 9 10 11 12 13 13 14 17 17 45

There are 12 values in this data set. Because there is an even number of values in the data set, the median is given by the average of the two middle values.

Example 3-5: Solution

The two middle values are the sixth and seventh in the foregoing list of data, and these two values are 12 and 13.

$$\textbf{Median} = \frac{\mathbf{12+13}}{\mathbf{2}} = \frac{\mathbf{25}}{\mathbf{2}} = \mathbf{12.5} = \mathbf{\$12.5\ billion}$$

Thus, the median profit of these 12 companies is $12.5 billion.

Median

The median gives the center of a histogram, with half the data values to the left of the median and half to the right of the median. The advantage of using the median as a measure of central tendency is that it is not influenced by outliers. Consequently, the median is preferred over the mean as a measure of central tendency for data sets that contain outliers.

Case Study 3-2 The Gender Pay Gap

Mode

Definition

The ***mode*** is the value that occurs with the highest frequency in a data set.

Example 3-6

The following data give the speeds (in miles per hour) of eight cars that were stopped on I-95 for speeding violations.

77 82 74 81 79 84 74 78

Find the mode.

Example 3-6: Solution

In this data set, 74 occurs twice and each of the remaining values occurs only once. Because 74 occurs with the highest frequency, it is the mode. Therefore,

Mode = **74 miles per hour**

Mode

- A major shortcoming of the mode is that a data set may have none or may have more than one mode, whereas it will have only one mean and only one median.
  - Unimodal: A data set with only one mode.
  - Bimodal: A data set with two modes.
  - Multimodal: A data set with more than two modes.

Example 3-7 (Data set with no mode)

Last year’s incomes of five randomly selected families were \$76,150. \$95,750, \$124,985, \$87,490, and \$53,740. Find the mode.

Example 3-7: Solution

Because each value in this data set occurs only once, this data set contains **no mode**.

Example 3-8 (Data set with two modes)

Refer to the data on 2008 profits of 12 companies given in Table 3.3 of Example 3-5. Find the mode for these data.

Example 3-8: Solution

In the data given in Example 3-5, each of two values 13 and 17 occurs twice, and each of the remaining values occurs only once. Therefore, that data set has two modes: **$13 billion** and **$ 17 billion**.

Example 3-9 (Data set with three modes)

The ages of 10 randomly selected students from a class are 21, 19, 27, 22, 29, 19, 25, 21, 22 and 30 years, respectively. Find the mode.

Example 3-9: Solution

This data set has three modes: **19**, **21** and **22**. Each of these three values occurs with a (highest) frequency of 2.

Mode

One advantage of the mode is that it can be calculated for both kinds of data - quantitative and qualitative - whereas the mean and median can be calculated for only quantitative data.

Example 3-10

The status of five students who are members of the student senate at a college are senior, sophomore, senior, junior, senior. Find the mode.

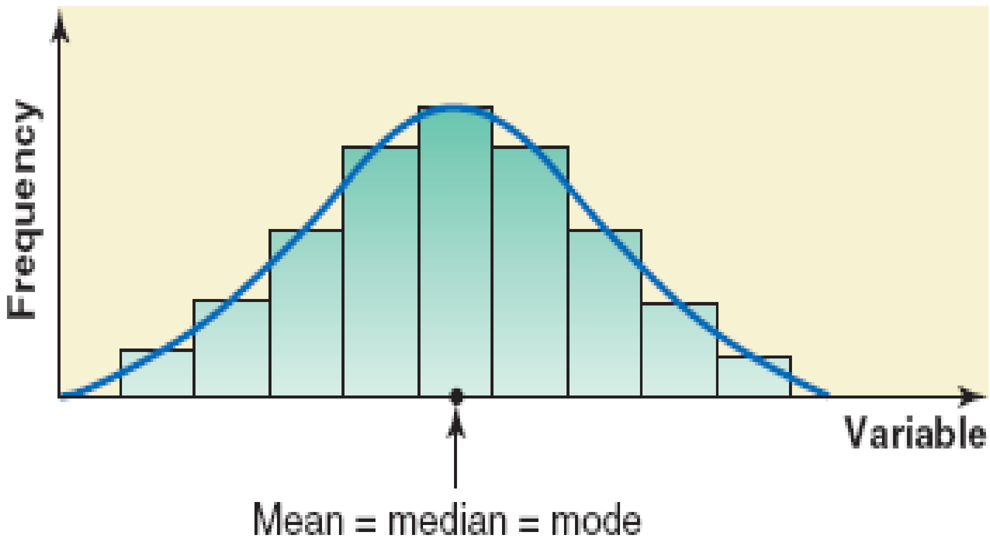
Example 3-10: Solution

Because **senior** occurs more frequently than the other categories, it is the mode for this data set. We cannot calculate the mean and median for this data set.

Relationships among the Mean, Median, and Mode

1. For a symmetric histogram and frequency curve with one peak (Figure 3.2), the values of the mean, median, and mode are identical, and they lie at the center of the distribution.

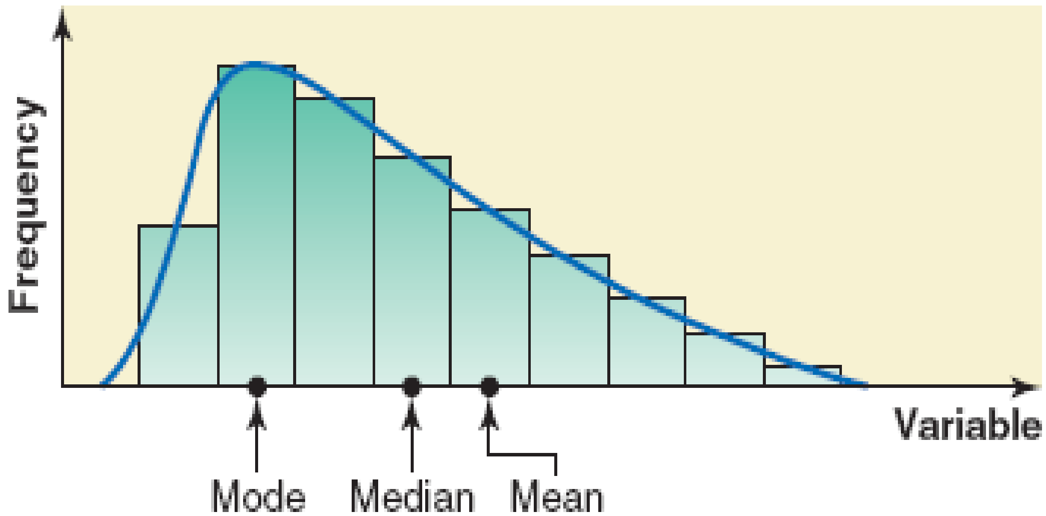
Figure 3.2 Mean, median, and mode for a symmetric histogram and frequency curve.



Relationships among the Mean, Median, and Mode

2. For a histogram and a frequency curve skewed to the right (Figure 3.3), the value of the mean is the largest, that of the mode is the smallest, and the value of the median lies between these two. (Notice that the mode always occurs at the peak point.) The value of the mean is the largest in this case because it is sensitive to outliers that occur in the right tail. These outliers pull the mean to the right.

Figure 3.3 Mean, median, and mode for a histogram and frequency curve skewed to the right.



Relationships among the Mean, Median, and Mode

3. If a histogram and a distribution curve are skewed to the left (Figure 3.4), the value of the mean is the smallest and that of the mode is the largest, with the value of the median lying between these two. In this case, the outliers in the left tail pull the mean to the left.

Figure 3.4 Mean, median, and mode for a histogram and frequency curve skewed to the right.