Summarizing Data: Measures of Location

When we are about to describe a set of data, it is a sound advice to say neither too little nor too much. Thus, depending on the nature of the data and the purpose we have in mind, statistical descriptions can be very brief or very elaborate. Sometimes we present data just as they are and let them speak for themselves; on other occasions we may just group the data and present their distribution in tabular or graphical form. Most of the time, though, we have to describe data in various other ways.

It is often appropriate to summarize data by means of a few well-chosen numbers that, in their way, are descriptive of the entire set. Exactly what sort of numbers we choose depends on the particular characteristics we want to describe. In one study we may be interested in a value that somehow describes the middle or the most typical of a set of data; in another we may be interested in the value that is exceeded only by 25% of the data; and in still another we may be interested in the length of the interval between the smallest and the largest values among the data. The statistical measures cited in the first two situations come under the heading of **measures of location** and the one cited in the third situation fits the definition of a **measure of variation**.

In this chapter, we shall concentrate on measures of location, and in particular on **measures of central location**, which in some way describe the center or the middle of a set of data.

Populations and Samples

If a set of data consists of all conceivably possible (or hypothetically possible) observations of a given phenomenon, we call it a **population**; if a set of data consists of only a part of these observations, we call it a **sample**.

The Mean

The most popular measure of central location is what the lay person calls an "average" and what the statistician calls an arithmetic mean, or simply a mean. It is defined as follows:

The mean of n numbers is their sum divided by n

EXAMPLE: From 1990 through 1994, the combined seizure of drugs the Drug Enforcement Administration, Custom's Service added up to 1,794, 3,030, 2,551, 3,514, and 2,824 pounds. Find the mean seizure of drugs for the given five-year period.

Solution: The total for the five years is

$$1,794 + 3,030 + 2,551 + 3,514 + 2,824 = 13,713 \text{ pounds}$$

Hence the mean is

$$\frac{13,713}{5} = 2,742.6 \text{ pounds}$$

In short, the mean is

$$\frac{1,794 + 3,030 + 2,551 + 3,514 + 2,824}{5} = \frac{13,713}{5} = 2,742.6 \text{ pounds}$$

EXAMPLE: In the 9th through 97th Congress of Egypt, there were, respectively, 67, 71, 78, 82, 96, 110, 104, and 92 Representatives at least 60 years old at the beginning of the first session. Find the mean.

Solution: The total of these figures is

$$67 + 71 + 78 + 82 + 96 + 110 + 104 + 92 = 700$$

Hence the mean is

$$\frac{700}{8} = 87.5$$

In short, the mean is

$$\frac{67 + 71 + 78 + 82 + 96 + 110 + 104 + 92}{8} = \frac{700}{8} = 87.5$$

EXAMPLE: Find the mean of the following numbers

(a) $1, 1$

(b) $\underbrace{1, 1, \ldots, 1}_{2012 \text{ times}}$

(c) $1, 2$

(d) $\underbrace{1, 1, \ldots, 1}_{2011 \text{ times}}, 2$

(e) $\underbrace{1, 1, \ldots, 1}_{2011 \text{ times}}, 989$

(f) $1, \underbrace{2, 2, \ldots, 2}_{2011 \text{ times}}$

(g) $\underbrace{1, 1, \ldots, 1}_{1006 \text{ times}}, \underbrace{2, 2, \ldots, 2}_{1006 \text{ times}}$

(h) $-1006, -1005, -1004, \ldots, -3, -2, -1, 0, 1, 2, 3, \ldots, 1004, 1005, 1006$

(i) $-1006, -1005, -1004, \ldots, -3, -2, -1, 1, 2, 3, \ldots, 1004, 1005, 1006$

(j) $1, 2, 3, 4, 5, 6$

(k) $1, 2, 3, \ldots, 2012$

(l) $0, 1, 2, 3, \ldots, 2012$

Solution:

(a) $\dfrac{1+1}{2} = \dfrac{2}{2} = 1$

(b) $\dfrac{\overbrace{1+1+\ldots+1}^{2012 \text{ times}}}{2012} = \dfrac{2012}{2012} = 1$

(c) $\dfrac{1+2}{2} = \dfrac{3}{2} = 1.5$

(d) $\dfrac{\overbrace{1+1+\ldots+1}^{2011 \text{ times}}+2}{2012} = \dfrac{2011+2}{2012} = \dfrac{2013}{2012} \approx 1.000497$

(e) $\dfrac{\overbrace{1+1+\ldots+1}^{2011 \text{ times}}+989}{2012} = \dfrac{2011+989}{2012} = \dfrac{3000}{2012} \approx 1.49$

(f) $\dfrac{1+\overbrace{2+2+\ldots+2}^{2011 \text{ times}}}{2012} = \dfrac{1+2\cdot 2011}{2012} = \dfrac{1+4022}{2012} = \dfrac{4023}{2012} \approx 1.9995$

(g) We have

$$\frac{\overbrace{1+1+\ldots+1}^{1006 \text{ times}}+\overbrace{2+2+\ldots+2}^{1006 \text{ times}}}{2012} = \frac{1\cdot 1006+2\cdot 1006}{2012} = \frac{(1+2)\cdot 1006}{2\cdot 1006} = \frac{1+2}{2} = \frac{3}{2} = 1.5$$

or

$$\frac{\overbrace{1+1+\ldots+1}^{1006 \text{ times}}+\overbrace{2+2+\ldots+2}^{1006 \text{ times}}}{2012} = \frac{\overbrace{(1+2)+(1+2)+\ldots+(1+2)}^{1006 \text{ times}}}{2012} = \frac{(1+2)\cdot 1006}{2\cdot 1006} = \frac{1+2}{2} = 1.5$$

(h) We have

$$\frac{(-1006)+(-1005)+(-1004)+\ldots+(-3)+(-2)+(-1)+0+1+2+3+\ldots+1004+1005+1006}{2013}$$

$$= \frac{(-1006+1006)+(-1005+1005)+(-1004+1004)+\ldots+(-3+3)+(-2+2)+(-1+1)+0}{2013}$$

$$= \frac{0+0+0+\ldots+0+0+0+0}{2013} = \frac{0}{2013} = 0$$

(i) We have

$$\frac{(-1006)+(-1005)+(-1004)+\ldots+(-3)+(-2)+(-1)+1+2+3+\ldots+1004+1005+1006}{2012}$$

$$=\frac{(-1006+1006)+(-1005+1005)+(-1004+1004)+\ldots+(-3+3)+(-2+2)+(-1+1)}{2012}$$

$$=\frac{0+0+0+\ldots+0+0+0+0}{2012}=\frac{0}{2012}=0$$

(j) We have

$$\frac{1+2+3+4+5+6}{6}=\frac{3+3+4+5+6}{6}=\frac{6+4+5+6}{6}=\frac{10+5+6}{6}=\frac{15+6}{6}=\frac{21}{6}=\frac{3\cdot 7}{2\cdot 3}=\frac{7}{2}=3.5$$

or

$$\frac{1+2+3+4+5+6}{6}=\frac{(1+6)+(2+5)+(3+4)}{6}=\frac{7+7+7}{6}=\frac{3\cdot 7}{6}=\frac{3\cdot 7}{2\cdot 3}=\frac{7}{2}=3.5$$

(k) We have

$$\frac{1+2+3+\ldots+2012}{2012}$$

$$=\frac{1+2+3+4+5+\ldots+1006+1007+\ldots+2008+2009+2010+2011+2012}{2012}$$

$$=\frac{\overbrace{(1+2012)+(2+2011)+(3+2010)+(4+2009)+(5+2008)+\ldots+(1006+1007)}^{1006\text{ times}}}{2012}$$

$$=\frac{\overbrace{2013+2013+2013+2013+2013+\ldots+2013}^{1006\text{ times}}}{2012}$$

$$=\frac{1006\cdot 2013}{2012}=\frac{1006\cdot 2013}{2\cdot 1006}=\frac{2013}{2}=\left\{\frac{2012+1}{2}=\frac{2012}{2}+\frac{1}{2}=1006+0.5\right\}=1006.5$$

(l) We have

$$\frac{0+1+2+3+\ldots+2012}{2013}=\frac{0}{2013}+\frac{1+2+3+\ldots+2012}{2013}=\frac{1+2+3+\ldots+2012}{2013}$$

$$=\frac{1+2+3+4+5+\ldots+1006+1007+\ldots+2008+2009+2010+2011+2012}{2013}$$

$$=\frac{\overbrace{(1+2012)+(2+2011)+(3+2010)+(4+2009)+(5+2008)+\ldots+(1006+1007)}^{1006\text{ times}}}{2013}$$

$$=\frac{\overbrace{2013+2013+2013+2013+2013+\ldots+2013}^{1006\text{ times}}}{2013}$$

$$=\frac{1006\cdot 2013}{2013}=1006$$

Since we shall have occasion to calculate the means of many different sets of sample data, it will be convenient to have a simple formula that is always applicable. This requires that we represent the figures to be averaged by some general symbol such as x, y, or z; the number of values in a sample, the sample size, is usually denoted by the letter n. Choosing the letter x, we can refer to the n values in a sample as $x_1, x_2, \ldots$, and x_n (which read "x sub-one," "x sub-two", ..., and "x sub-n"), and write

$$\textbf{Sample mean} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

This formula will take care of any set of sample data, but it can be made more compact by assigning the sample mean the symbol $\overline{x}$ (which reads "x bar") and using the $\sum$ notation. The symbol $\sum$ is capital sigma, the Greek letter for S. In this notation we let $\sum x$ stand for "the sum of the x's (that is, $\sum x = x_1 + x_2 + \ldots + x_n$), and we can write

$$\boxed{\overline{x} = \frac{\sum x}{n}}$$

The number of values in a population, the population size, is usually denoted by N. The mean of a population of N items is defined in the same way as the mean of a sample. It is the sum of the N items, $x_1 + x_2 + x_3 + \ldots + x_N$ or $\sum x$ divided by N.

Assigning the population mean the symbol μ (the Greek letter for m) we write

$$\boxed{\mu = \frac{\sum x}{N}}$$

With the reminder that $\sum x$ is now the sum of all N values of x that constitute the population.

Also, to distinguish between descriptions of populations and descriptions of samples, we not only use different symbols such as μ and $\overline{x}$, but we refer to a description of a population as a **parameter** and a description of a sample as a **statistic**. Parameters are usually denoted by Greek letters.

To illustrate the terminology and notation just introduced, suppose that we are interested in the mean lifetime of a production lot of $N = 40,000$ light bulbs. Obviously, we cannot test all of the light bulbs for there would be none left to use or sell, so we take a sample, calculate $\overline{x}$, and use this quantity as an estimate of μ.

EXAMPLE: lf $n = 5$ and the light bulbs in the sample last 967, 949, 952, 940, and 922 hours, what can we conclude about the mean lifetime of the 40,000 light bulbs in the production lot?

Solution: The mean of this sample is

$$\overline{x} = \frac{967 + 949 + 952 + 940 + 922}{5} = 946 \text{ hours}$$

If we can assume that the data constitute a sample in the technical sense (namely, a set of data from which valid generalizations can be made), we estimate the mean of all 40,000 light bulbs as $\mu = 946$ hours.

For nonnegative data, the mean not only describes their middle, but it also puts some limitation on their size. If we multiply by n on both sides of the equation $\overline{x} = \dfrac{\sum x}{n}$, we find that

$$\sum x = n \cdot \overline{x}$$

and, hence, that no part, or subset of the data can exceed $n \cdot \overline{x}$.

EXAMPLE: If the mean salary paid to three NBA players for the 1998-1999 season is \$2,450,000, can:

(a) Anyone of them receive an annual salary of \$4,000,000?

(b) Any two of them receive an annual salary of \$4,000,000?

Solution: The combined salaries of the three players total $3(2,450,000) = \$7,350,000$.

(a) If one of them receives an annual salary of \$4,000,000, this would leave $7,350,000 - 4,000,000 = \$3,350,000$ for the other two players, so this could be the case.

(b) For two of them to receive an annual salary of \$4,000,000 would require $2(4,000,000) = \$8,000,000$, which exceeds the total paid to the three players. Hence, this cannot be the case.

EXAMPLE: If six high school juniors averaged 57 on the verbal part of the PSAT/MSQT test, at most how many of them could have scored 72 or better on the test?

Solution: Since $n = 6$ and $\overline{x} = 57$, it follows that their combined scores total $6(57) = 342$. Since $342 = 4 \times 72 + 54$, we find that at most four of the six students could have scored 72 or more.

EXAMPLE: The editor of a book on nutritional values needs a figure for the calorie count of a slice of a l2-inch pepperoni pizza. Letting a laboratory with a calorimeter do the job, she gets the following figures for the pizza from six different fast-food chains: 265, 332, 340, 225, 238, and 346.

(a) Calculate the mean, which the editor will report in her book.

(b) Suppose that when calculating the mean, the editor makes the mistake of entering 832 instead of 238 in her calculator. How much of an error would this make in the figure that she reports in her book?

EXAMPLE: The editor of a book on nutritional values needs a figure for the calorie count of a slice of a l2-inch pepperoni pizza. Letting a laboratory with a calorimeter do the job, she gets the following figures for the pizza from six different fast-food chains: 265, 332, 340, 225, 238, and 346.

(a) Calculate the mean, which the editor will report in her book.

(b) Suppose that when calculating the mean, the editor makes the mistake of entering 832 instead of 238 in her calculator. How much of an error would this make in the figure that she reports in her book?

Solution:

(a) The correct mean is

$$\overline{x} = \frac{265 + 332 + 340 + 225 + 238 + 346}{6} = 291$$

(b) The incorrect mean is

$$\overline{x} = \frac{265 + 332 + 340 + 225 + 832 + 346}{6} = 390$$

So that her error would be a disastrous $390 - 291 = 99$.

EXAMPLE: The ages of six students who went on a geology field trip are 16, 17, 15, 19, 16, and 17, and the age of the instructor who went with them is 54. Find the mean age of these seven persons.

Solution: The mean is

$$\overline{x} = \frac{16 + 17 + 15 + 19 + 16 + 17 + 54}{7} = 22$$

But any statement to the effect that the average age of the group is 22 could easily be misinterpreted. We might well infer incorrectly that most of the persons who went on the field trip are in their low twenties.

To avoid the possibility of being misled by a mean affected by a very small value or a very large value, it may be advisable to omit such an **outlier** or to describe the middle or center of a set of data with a statistical measure other than the mean; perhaps, with the **median**, which we shall discuss.

The Weighted Mean

When we calculate a mean, we may be making a serious mistake if we overlook the fact that the quantities we are averaging are not all of equal importance with reference to the situation being described. Consider, for example, a cruise line that advertises the following fares for single-occupancy cabins on an 11-day cruise:

Cabin category	Fare
Ultra deluxe(outside)	\$7,870
Deluxe (outside)	\$7,080
Outside	\$5,470
Outside (shower only)	\$4,250
Inside (shower only)	\$3,460

The mean of these five fares is

$$\overline{x} = \frac{7,870 + 7,080 + 5,470 + 4,250 + 3,460}{5} = \$5,626$$

But we cannot very well say that the average fare for one of these single occupancy cabins is \$5,626. To get that figure, we would also have to know how many cabins there are in each of the categories. Referring to the ship's deck plan, where the cabins are color-coded by category, we find that there are, respectively, 6, 4, 8, 13, and 22 cabins available in these five categories. If it can be assumed that these 53 cabins will all be occupied, the cruise line can expect to receive a total of

$$6(7,870) + 4(7,080) + 8(5,470) + 13(4,250) + 22(3,460) = 250,670$$

for the 53 cabins and, hence, on the average $\frac{250,670}{53} \approx \$4,729.62$ per cabin.

To give quantities being averaged their proper degree of importance, it is necessary to assign them **relative importance weights** and then calculate a **weighted mean**. In general, the weighted mean $\overline{x}_w$ of a set of numbers $x_1, x_2, x_3, \ldots$ and x_n, whose relative importance is expressed numerically by a corresponding set of numbers $w_1, w_2, w_3, \ldots$ and w_n is given by:

$$\boxed{\overline{x}_w = \frac{w_1x_1 + w_2x_2 + \ldots + w_nx_n}{w_1 + w_2 + \ldots + w_n} = \frac{\sum w \cdot x}{\sum w}}$$

Here $\sum w$ is the sum of the products obtained by multiplying each x by the corresponding weight, and $\sum w$ is simply the sum of the weights. Note that when the weights are all equal, the formula for the weighted mean reduces to that for the ordinary (arithmetic) mean.

EXAMPLE: The following Table shows the number of households in the five Pacific states in 1990, and the corresponding percentage changes in the number of households 1990-1994:

	Number of households(1,000)	Percentage change
Washington	1,872	9.1
Oregon	1,103	8.3
California	10,381	4.5
Alaska	189	10.3
Hawaii	356	7.1

Calculate the weighted mean of the percentage changes using the 1990 numbers of households as weights.

EXAMPLE: The following Table shows the number of households in the five Pacific states in 1990, and the corresponding percentage changes in the number of households 1990-1994:

	Number of households(1,000)	Percentage change
Washington	1,872	9.1
Oregon	1,103	8.3
California	10,381	4.5
Alaska	189	10.3
Hawaii	356	7.1

Calculate the weighted mean of the percentage changes using the 1990 numbers of households as weights.

Solution: Substituting $x_1 = 9.1,\ x_2 = 8.3, x_3 = 4.5,\ x_4 = 10.3,\ x_5 = 7.1,\ w_1 = 1,872,\ w_2 = 1,103,\ w_3 = 10,381,\ w_4 = 189$, and $w_5 = 356$ into the formula for the weighted mean, we get

$$\frac{9.1(1,872) + 8.3(1,103) + 4.5(10,381) + 10.3(189) + 7.1(356)}{1,872 + 1,103 + 10,381 + 189 + 356} = \frac{77,378.9}{13,901} \approx 5.6\%$$

A special application of the formula for the weighted mean arises when we must find the **overall mean**, or **grand mean**, of k sets of data having the means $\overline{x}_1, \overline{x}_2, \overline{x}_3, \ldots$ and $\overline{x}_k$ and consisting of $n_1, n_2, n_3, \ldots$ and n_k measurements or observations. The result is given by:

$$\overline{\overline{x}} = \frac{n_1\overline{x}_1 + n_2\overline{x}_2 + \ldots + n_k\overline{x}_k}{n_1 + n_2 + n_3 + \ldots + n_k} = \frac{\sum n \cdot \overline{x}}{\sum n}$$

where the weights are the sizes of the samples, the numerator is the total of all the measurements or observations, and the denominator is the number of items in the combined samples.

EXAMPLE: There are three sections of a course in European history, with 19 students in the 1st section meeting MWF at 9 A.M., 27 in the 2nd section meeting MWF at 11 A.M., and 24 in the 3rd section meeting MWF at 1 P.M. If the students in the 9 A.M. section averaged 66 in the midterm examination, those in the 11 A.M. section averaged 71, and those in the 1 P.M. section averaged 63, what is the mean score for all three sections combined?

Solution: Substituting $n_1 = 19, n_2 = 27, n_3 = 24, \overline{x}_1 = 66, \overline{x}_2 = 71$, and $\overline{x}_3 = 63$ into the formula for the grand mean of combined data, we get

$$\overline{\overline{x}} = \frac{19 \cdot 66 + 27 \cdot 71 + 24 \cdot 63}{19 + 27 + 24} = \frac{4,683}{70} = 66.9$$

or 67 rounded to the nearest integer.

The Median

To avoid the possibility of being misled by one or a few very small or very large values, we sometimes describe the "middle" or "center" of a set of data with statistical measures other than the mean. One of these, the **median** of n values requires that we first arrange the data according to size. Then it is defined as follows:

The median is the value of the middle item when n is odd, and the mean of the two middle items when n is even.

EXAMPLE:

(a) The median of 1, 2, and 3 is 2. The mean is $\dfrac{1+2+3}{3} = \dfrac{6}{3} = 2$.

(b) The median of 1, 2, and 100 is 2. The mean is $\dfrac{1+2+100}{3} = \dfrac{103}{3} = 34.\overline{3}$.

(c) The median of 1, 2, 3, and 4 is $\dfrac{2+3}{2} = \dfrac{5}{2} = 2.5$. The mean is $\dfrac{1+2+3+4}{4} = \dfrac{10}{4} = 2.5$.

(d) The median of 1, 2, 3, and 100 is $\dfrac{2+3}{2} = \dfrac{5}{2} = 2.5$. The mean is

$$\frac{1+2+3+100}{4} = \frac{106}{4} = 26.5$$

EXAMPLE: In five recent weeks, a town reported 36, 29, 42, 25, and 29 burglaries. Find the median number of burglaries for these weeks.

Solution: The median is not 42, the third (or middle) item, because the data must first be arranged according to size. Thus, we get:

$$25 \qquad 29 \qquad 29 \qquad 36 \qquad 42$$

and it can be seen that the middle one, the median, is 29.

EXAMPLE: In some cities, persons cited for minor traffic violations can attend a class in defensive driving in lieu of paying a fine. Given that 12 such classes in Phoenix, Arizona, were attended by 37, 32, 28, 40, 35, 38, 40, 24, 30, 37, 32, and 40 persons, find the median of these data.

EXAMPLE: In some cities, persons cited for minor traffic violations can attend a class in defensive driving in lieu of paying a fine. Given that 12 such classes in Phoenix, Arizona, were attended by 37, 32, 28, 40, 35, 38, 40, 24, 30, 37, 32, and 40 persons, find the median of these data.

Solution: Ranking these attendance figures according to size, from low to high, we get

	24	28	30	32	32	35	37	37	38	40	40	40

and we find that the median is the mean of the two values nearest the middle namely, $\frac{35+37}{2} = 36$.

EXAMPLE: On the seventh hole of a certain golf course, a par four, nine golfers scored par, birdie (one below par), par, par, bogey (one above par), eagle (two below par), par, birdie, birdie. Find the median.

Solution: Ranking these figures according to size, from low to high, we get

2 3 3 3 4 4 4 4 5

and it can be seen that the fifth value, the median, is equal to par 4.

The symbol that we use for the median of n sample values $x_1, x_2, x_3, \ldots$, and x_n is $\tilde{x}$ (and hence, $\tilde{y}$ or $\tilde{z}$ if we refer to the values of y's and z's). If a set of data constitutes a population, we denote its median by $\tilde{\mu}$.

Thus, we have a symbol for the median, but no formula; there is only a formula for the **median position**. Referring again to data arranged according to size, usually ranked from low to high, we can write

The median is the value of the $\frac{n+1}{2}$th item.

EXAMPLE: Find the median position for

(a) $n = 17$ (b) $n = 41$

Solution: With the data arranged according to size (and counting from either end)

(a) $\frac{n+1}{2} = \frac{17+1}{2} = 9$ and the median is the value of the 9th item;

(b) $\frac{n+1}{2} = \frac{41+1}{2} = 21$ and the median is the value of the 21th item.

EXAMPLE: Find the median position for

(a) $n = 16$ (b) $n = 50$

Solution: With the data arranged according to size (and counting from either end)

(a) $\frac{n+1}{2} = \frac{16+1}{2} = 8.5$ and the median is the mean of the values of the 8th and 9th items;

(b) $\frac{n+1}{2} = \frac{50+1}{2} = 25.5$ and the median is the mean of the values of the 25th and 26th items.

NOTE: It is important to remember that $\frac{n+1}{2}$ is the formula for the median position and not a formula for the median, itself.

Other Fractiles

The median is but one of many **fractiles** that divide data into two or more parts, as nearly equal as they can be made. Among them we also find **quartiles**, **deciles**, and **percentiles**, which are intended to divide data into four, ten, and a hundred parts. Until recently, fractiles were determined mainly for distributions of large sets of data.

In this section, we shall concern ourselves mainly with a problem that has arisen in **exploratory data analysis** — in the preliminary analysis of relatively small sets of data. It is the problem of dividing such data into four nearly equal parts, where we say "nearly equal" because there is no way in which we can divide a set of data into four equal parts for, say, $n = 27$ or $n = 33$. Statistical measures designed for this purpose have traditionally been referred to as the three quartiles, Q_1, Q_2, and Q_3, and there is no argument about Q_2, which is simply the median. On the other hand, there is some disagreement about the definition of Q_1, and Q_3.

As we shall define them, the quartiles divide a set of data into four parts such that there are as many values less than Q_1 as there are between Q_1 and Q_2 between Q_2 and Q_3, and greater than Q_3. Assuming that no two values are alike, this is accomplished by letting

Q_1 be the median of all the values less than the median of the whole set of data,

and

Q_3 be the median of all the values greater than the median of the whole set of data.

EXAMPLE: Following are the high-temperature readings in twelve European capitals on a recent day in the month of June: 90, 75, 86, 77, 85, 72, 78, 79, 94, 82, 74, and 93. Find Q_1, Q_2 (the median), and Q_3.

Solution: For $n = 12$ the median position is $\dfrac{12+1}{2} = 6.5$ and, after arranging the data according to size, we find that the sixth and seventh values among

	72	74	75	77	78	79	82	85	86	90	93	94

are 79 and 82. Hence the median is $Q_2 = \dfrac{79+82}{2} = 80.5$. For the six values below 80.5 the median position is $\dfrac{6+1}{2} = 3.5$, and since the third and fourth values are 75 and 77, $Q_1 = \dfrac{75+77}{2} = 76$. Counting from the other end, the third and fourth values are 90 and 86, and $Q_3 = \dfrac{90+86}{2} = 88$. As can be seen from the data, there are three values below 76, three values between 76 and 80.5, three values between 80.5 and 88, and three values above 88.

Everything worked nicely in this example, but $n = 12$ happened to be a multiple of 4, which raises the question whether our definition of Q_1 and Q_3 will work also when this is not the case.

EXAMPLE: Suppose that the city where the high temperature was 77 failed to report, so that we are left with the following 11 numbers arranged according to size:

	72	74	75	78	79	82	85	86	90	93	94

Find Q_1, Q_2 (the median), and Q_3.

EXAMPLE: Suppose that the city where the high temperature was 77 failed to report, so that we are left with the following 11 numbers arranged according to size:

	72	74	75	78	79	82	85	86	90	93	94

Find Q_1, Q_2 (the median), and Q_3.

Solution: For $n = 11$ the median position is $\dfrac{11+1}{2} = 6$ and, referring to the preceding data, which are already arranged according to size, we find that the median is $Q_2 = 82$. For the five values below 82 the median position is $\dfrac{5+1}{2} = 3$, and Q_1, the third value, equals 75. Counting from the other end, Q_3, the third value, equals 90. One can see that there are two values below 75, two values between 75 and 82, two values between 82 and 90, and two values above 90. Again, this satisfies the requirement for the three quartiles, Q_1, Q_2, and Q_3.

If some of the values are alike, we modify the definitions of Q_1 and Q_3 by replacing "less than the median" by "to the left of the median position" and "greater than the median" by "to the right of the median position".

EXAMPLE: Consider the following set of data

2 3 3 3 4 4 4 4 5

The median (Q_2), the fifth value, equals 4. Now, the median of the four values to the left of the median position, Q_1, equals 3, and the median of the four values to the right of the median position, Q_3, equals 4.

The Mode

Another measure that is sometimes used to describe the middle or center of a set of data is the **mode**, which is defined simply as the value that occurs with the highest frequency and more than once. Its two main advantages are that it requires no calculations, only counting, and it can be determined for qualitative, or nominal, data.

EXAMPLE: The 20 meetings of a square dance club were attended by 22, 24, 23, 24, 27, 25, 20, 24, 26, 28, 26, 23, 21, 24, 24, 25, 23, 28, 26, and 25 of its members. Find the mode.

Solution: Among these numbers, 20, 21, 22, and 27 each occurs once, 28 occurs twice, 23, 25, and 26 each occurs three times; and 24 occurs 5 times. Thus, the modal attendance is 24.

EXAMPLE: Consider the following set of data

2 3 3 3 4 4 4 4 5

Find the mode.

Solution: Since these data are already arranged according to size, it can easily be seen that 4, which occurs four times, is the modal score.