Determining the Five-Number Summary

To determine the five-number summary for a data set, we must determine the median (also known as the second quartile), the first quartile, Q_1, the third quartile, Q_3, the minimum, and the maximum for the data set.

For our analysis, let us consider the data representing the amount of rainfall, in inches, from twenty (20) rain storms in Chicago, Illinois. Ordering the data from the smallest to the largest, we have 0.97, 1.14, 1.85, 2.34, 2.47, 2.78, 3.41, 3.48, 3.94, 3.97, 4, 4.02, 4.11, 4.77, 5.22, 5.5, 5.79, 6.14, 6.28, 7.69.

Since the minimum is the smallest data value, the minimum is 0.97 inches. Similarly, since the maximum is the largest data value, the maximum is 7.69 inches. These are the smallest value and the largest value in the five-number summary.

In order to determine the quartiles, we must divide the data in half and, then, into quarters. Examining the twenty ordered data values, 0.97, 1.14, 1.85, 2.34, 2.47, 2.78, 3.41, 3.48, 3.94, 3.97, 4, 4.02, 4.11, 4.77, 5.22, 5.5, 5.79, 6.14, 6.28, 7.69, the first half of the data is made up of the first through the tenth data value, inclusive, 0.97, 1.14, 1.85, 2.34, 2.47, 2.78, 3.41, 3.48, 3.94, 3.97, and the second half of the data is made up of the eleventh through the twentieth data values, inclusive, 4, 4.02, 4.11, 4.77, 5.22, 5.5, 5.79, 6.14, 6.28, 7.69. Of the quartiles, it is best to determine the second quartile, also known as the median, first. It is important to note that the median has two representations, Q_2 and $\tilde{x}$; either of these notations will be acceptable. To determine the median, we calculate the average (i.e. the mean) of the 10th and the 11th data values since there is an even number of values in the distribution.

$$Q_2 = \frac{3.97+4}{2}$$

$$Q_2 = \frac{7.97}{2}$$

$$Q_2 = 3.985.$$

Since the original data values were recorded to two decimal places, we record one more decimal place for each calculated value. So, we record the median to three decimal places. Then, to state the median, we must remember to include the units of measure: the median is $Q_2 = 3.985$ inches.

Now that we have determined the two halves of the distribution, we can use the lower half of the distribution, 0.97, 1.14, 1.85, 2.34, 2.47, 2.78, 3.41, 3.48, 3.94, 3.97, to determine the lower quartile (also known as the first quartile), Q_1. Since there is an even number of data values in this first half of the distribution (that is, there are 10 values), we divide this first half of the distribution in half to create the first quarter of the distribution, 0.97, 1.14, 1.85, 2.34, 2.47, and the second quarter of the distribution, 2.78, 3.41, 3.48, 3.94, 3.97; the lower quartile is the median of the lower half of the distribution. Since there is an even number of data values in the lower half of the distribution, we calculate the average of the 5th and the 6th data values in order to determine the first quartile. That is,

$$Q_1 = \frac{2.47+2.78}{2}$$

$$Q_1 = \frac{5.25}{2}$$

$$Q_1 = 2.625$$

Again, since the original data was recorded to two decimal places, we record one more decimal place for each value calculated from this data. So, we record the first quartile to three decimal places. To state the first quartile, we must remember to include the units of measure: the first quartile is Q_1=2.625 inches.

Using the upper half of the distribution, 4, 4.02, 4.11, 4.77, 5.22, 5.5, 5.79, 6.14, 6.28, 7.69, we determine the upper quartile. Since there is an even number of data values in the upper half of the distribution (that is, there are 10 values), we divide this upper half of the distribution in half to create the third quarter of the distribution, 4, 4.02, 4.11, 4.77, 5.22, and the fourth quarter of the distribution, 5.5, 5.79, 6.14, 6.28, 7.69. Since there is an even number of data values in the upper half of the distribution, we calculate the average of the 15th and 16th data values in order to determine the third quartile.

$$Q_3=\frac{5.22+5.5}{2}$$
$$Q_3=\frac{10.72}{2}$$
$$Q_3=5.36$$

Again, since the original data was recorded to two decimal places, we record one more decimal place for values calculated from the data: we record the third quartile to three decimal places. Then, to state the third quartile, we must include the units of measure: the third quartile is Q_3=5.360 inches.

If we state the five-number summary using set notation, we have $\{0.97, 2.625, 3.985, 5.360, 7.69\}$. Notice that the set notation does not have a title and does not include units of measure. To represent the five-number summary using a table, we may include the units of measure on the individual values or in the title.

Five-Numbers Summary for the Amount of Rainfall, in inches,
from Twenty (20) Rain Storms in Chicago, Illinois

$$\text{Minimum} = 0.97$$
$$Q_1 = 2.625$$
$$Q_2 = 3.985$$
$$Q_3 = 5.360$$
$$\text{Maximum} = 7.69$$

For an extended example, let us take the same units and scenario as above. Be careful to make the distinction between the quartiles and the average. The quartiles mark the quarters of the distribution: these quarters are the *physical quarters* of the ordered data. If there is a odd number of data values in the distribution or the halves of the distribution then the quartiles can be values in the distribution: for an odd number of data values, there is an actual middle value. For example, if the data is 4, 4.02, 4.11, 4.77, 5.22 then the median is 4.11 inches since 4.11 is the middle data value; the third data value if the middle data value for a data set having five data values. To determine the value of Q_1, we must take the average of 4 and 4.02 and to determine the value of Q_3, we must take the average of 4.77 and 5.22; in this case, each of the halves of this data set has two data values.

However, if there were six data values, for example, 4, 4.02, 4.11, 4.77, 5.22, 5.5 then, to determine the median, we must take the average of the 3^{rd} and 4^{th} data values since there is an even number of data values;

$$Q_2=\frac{4.11+4.77}{2}$$
$$Q_2=\frac{8.88}{2}$$
$$Q_2=4.44 \text{ inches.}$$

To determine the lower quartile and upper quartile for these six data values (that is, for 4, 4.02, 4.11, 4.77, 5.22, 5.5), we examine the lower half of the distribution, 4, 4.02, 4.11, and the upper half of the distribution 4.77, 5.22, 5.5. For the lower half of the distribution, 4.02 is the middle value and Q_1=4.02 inches, while for the upper half of the distribution, 5.22 is the middle value so that Q_3=5.22 inches.

It is important to keep in mind that the median is the middle value for the *ordered* data while the mean is the physical middle of the data weighted by their distance or difference from the mean. You can think of the mean as the balance point for the data values where the weights are provided by the distances (or differences) of the values from the mean. The median can be thought of as the *physical* middle of the distribution and the mean can be thought of as the *balance* point for the distribution.