

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, Q1, the third quartile, Q3, the minimum, and the maximum for the data set.

For our analysis, let us consider the data set provided in exercise #9 on page 129 of the Sullivan text. To start, let us put the data in numerical order from the smallest to the largest, 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. It is important to note that since this data is for the rainfall, in inches, in Chicago, Illinois, we must remember to include the units with each value of the Five Number Summary.

The minimum is the smallest data value. So, here, the smallest data value is 0.97, and then, the minimum is 0.97 inches.

The maximum is the largest data value. So, here, the largest data value is 7.69, and then, the maximum is 7.69 inches.

In order to determine the quartiles, we must divide the data first 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, 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, 4, 4.02, 4.11, 4.77, 5.22, 5.5, 5.79, 6.14, 6.28, 7.69. Of the quartiles, the first one to find is the second quartile (also known as the median). It is important to note that the median has two representations, Q2 and \tilde{x} ; either of these notations will be acceptable. Then, to calculate the median, since there is an even number of data values, we determine the mean (the average) of the 10th and the 11th data values.

$$\begin{aligned} Q2 &= \frac{3.97+4}{2} \\ &= \frac{7.97}{2} \\ &= 3.985. \end{aligned}$$

Since the original data values were recorded to two decimal places, we can record one more decimal place for each calculated value. So, here, we can record the median to three decimal places. Then, to finally state the median, we must remember to include the units. That is, the median is $\tilde{x} = 3.985$ inches.

Now that we have determined the first half of the data, we can use this first half of the data, 0.97, 1.14, 1.85, 2.34, 2.47, 2.78, 3.41, 3.48, 3.94, 3.97, to determine the first quartile, Q1. Since there is an even number of data values in this first half of the data set (10 values), we divide this first half of the data in half to create the first quarter of the data values, 0.97, 1.14, 1.85, 2.34, 2.47, and the second quarter of the data values, 2.78, 3.41, 3.48, 3.94, 3.97. Since there was an even number of data values in this first half of the data, we must calculate the average of the 5th and the 6th data values in order to determine the first quartile. That is,

$$\begin{aligned} Q1 &= \frac{2.47+2.78}{2} \\ &= \frac{5.25}{2} \\ &= 2.625 \end{aligned}$$

Again, since the original data values were recorded to two decimal places, we can record one more decimal place for each calculated value. So, here we can record the first quartile to three decimal places. Then, to finally state the first quartile, we must remember to include the units. That is the first quartile is Q1=2.625 inches.

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

$$\begin{aligned} Q3 &= \frac{5.22+5.5}{2} \\ &= \frac{10.72}{2} \\ &= 5.36 \end{aligned}$$

Yet again, since the original data values were recorded to two decimal places, we can record one more decimal place for each calculated value. So, here we can record the third quartile to three decimal places. Then, to finally state the third quartile, we must remember to include the units. That is the third quartile is $Q3=5.360$ inches.

(For this extended example, let us take the same units and scenario as above.) Please be careful to make the distinction between the quartiles and the mean (average). The quartiles mark the quarters of the data set: these quarters are the *physical quarters* of the ordered data set. If there are an odd number of data values for any of the halves of the data then the quartiles can turn out to be data values: 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. Then, to determine the value of $Q1$, we must take the average of 4 and 4.02 and to determine the value of $Q3$, 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 3rd and 4th data values since there are an even number of data values;

$$\begin{aligned} Q2 &= \frac{4.11+4.77}{2} \\ &= \frac{8.88}{2} \\ &= 4.44 \text{ inches.} \end{aligned}$$

Then, to determine the first and third quartiles for these six data values (4, 4.02, 4.11, 4.77, 5.22, 5.5), we examine the first half of the data, 4, 4.02, 4.11, and the second half of the data 4.77, 5.22, 5.5. For the first half of the data, 4.02 is the middle value so that $Q1=4.02$ inches, and for the second half of the data, 5.22 is the middle value so that $Q3=5.22$ inches.

The median is the middle value for the *ordered* data while the mean is the physical middle of the data weighted by their values. You can think of the mean as the balance point for the data values where the weights are provided by the distances of the data values from the mean (the middle). So, median can be thought of as the *physical* middle and the mean can be thought of as the *balance* point for the data.