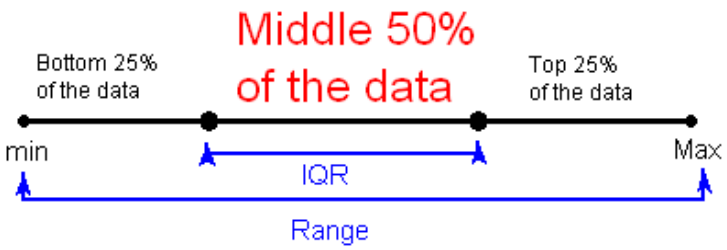
Inter-Quartile Range (IQR)

Learning Objective: Relate measures of center and spread to the shape of the distribution, and choose the appropriate measures in different contexts.

Inter-Quartile Range (IQR)

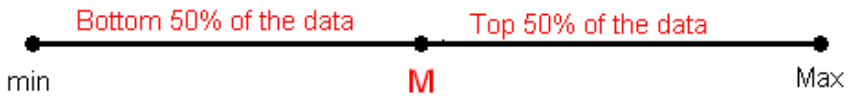
While the range quantifies the variability by looking at the range covered by *ALL* the data, the IQR measures the variability of a distribution by giving us the range covered by the *MIDDLE 50%* of the data.

The following picture illustrates this idea: (Think about the horizontal line as the data ranging from the min to the Max).

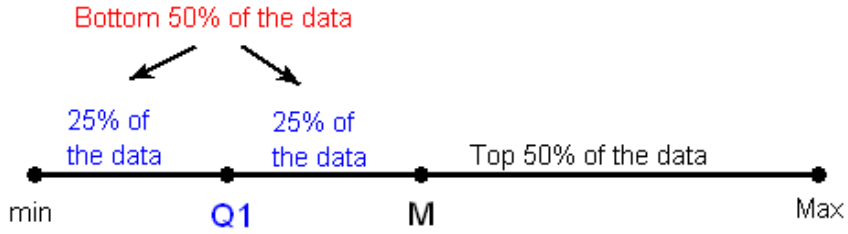


Here is how the IQR is actually found:

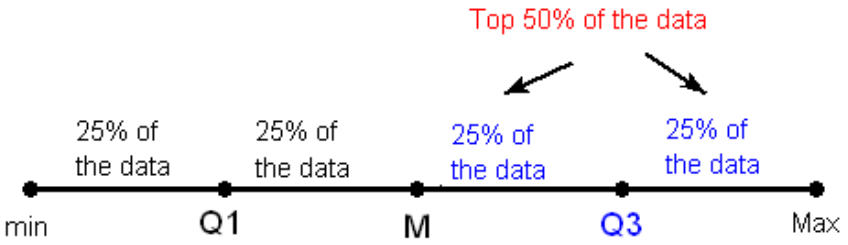
1. Arrange the data in increasing order, and find the median M. Recall that the median divides the data, so that 50% of the data points are below the median, and 50% of the data points are above the median.



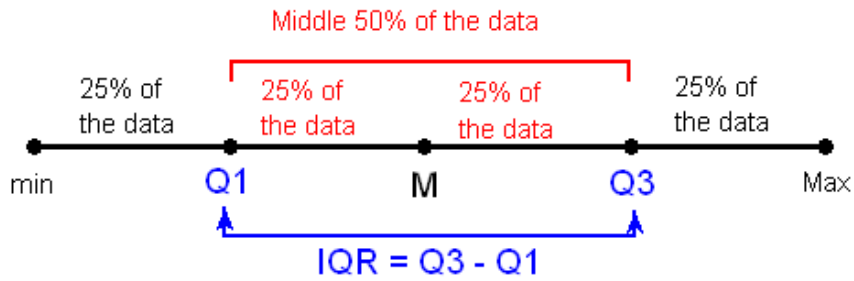
2. Find the median of the lower 50% of the data. This is called the first quartile of the distribution, and the point is denoted by Q1. Note from the picture that Q1 divides the lower 50% of the data into two halves, containing 25% of the data points in each half. Q1 is called the first quartile, since one quarter of the data points fall below it.



3. Repeat this again for the top 50% of the data. Find the median of the top 50% of the data. This point is called the third quartile of the distribution, and is denoted by Q3. Note from the picture that Q3 divides the top 50% of the data into two halves, with 25% of the data points in each. Q3 is called the third quartile, since three quarters of the data points fall below it.



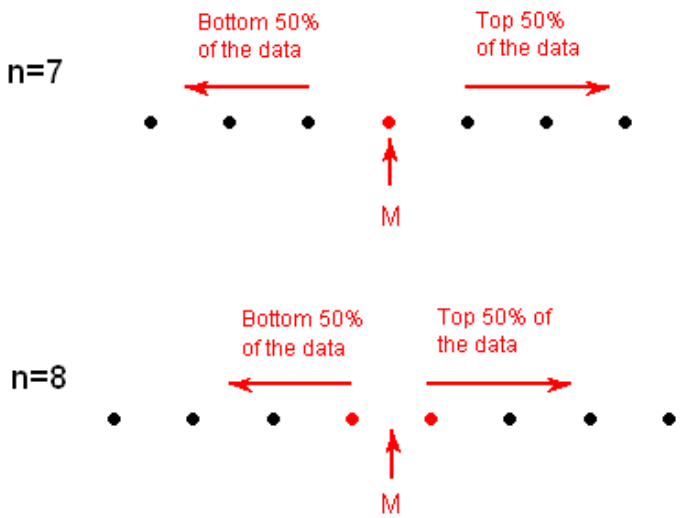
4. The middle 50% of the data falls between Q1 and Q3, and therefore:
IQR = Q3 - Q1



Comments

1. The last picture shows that Q1, M, and Q3 divide the data into four quarters with 25% of the data points in each, where the median is essentially the second quartile. The use of IQR = Q3 - Q1 as a measure of spread is therefore particularly appropriate when the median M is used as a measure of center.

2. We can define a bit more precisely what is considered the bottom or top 50% of the data. The bottom (top) 50% of the data is all the observations whose position in the ordered list is to the left (right) of the location of the overall median M. The following picture will visually illustrate this for the simple cases of n = 7 and n = 8.



Note that when n is odd (as in n = 7 above), the median is **not** included in either the bottom or top half of the data; When n is even (as in n = 8 above), the data are naturally divided into two halves.