

Box Plots

The *box plot*, or box-whisker plot, gives a good graphical overall impression of the concentration of the data. It also shows how far from most of the data the extreme values are. In principle, the box plot is constructed from five values: the smallest value, the first quartile, the median, the third quartile, and the largest value. The median, the first quartile, and the third quartile will be discussed here, and then once more in the next section.

The *median*, a number, is a way of measuring the “center” of the data. You can think of the median as the “middle value,” although it does not actually have to be one of the observed values. It is a number that separates ordered data into halves. Half the values are the same size or smaller than the median and half the values are the same size or larger than it. For example, consider the following data that contains 14 values:

1, 11.5, 6, 7.2, 4, 8, 9, 10, 6.8, 8.3, 2, 2, 10, 1 .

Ordered, from smallest to largest, we get:

1, 1, 2, 2, 4, 6, 6.8, 7.2, 8, 8.3, 9, 10, 10, 11.5 .

The median is between the 7th value, 6.8, and the 8th value 7.2. To find the median, add the two values together and divide by 2:

$$\frac{6.8 + 7.2}{2} = 7$$

The median is 7. Half of the values are smaller than 7 and half of the values are larger than 7.

Quartiles are numbers that separate the data into quarters. Quartiles may or may not be part of the data. To find the quartiles, first find the median or second quartile. The first quartile is the middle value of the lower half of the data and the third quartile is the middle value of the upper half of the data. For illustration consider the same data set from above:

1, 1, 2, 2, 4, 6, 6.8, 7.2, 8, 8.3, 9, 10, 10, 11.5 .

The median or second quartile is 7. The lower half of the data is:

1, 1, 2, 2, 4, 6, 6.8 .

The middle value of the lower half is 2. The number 2, which is part of the data in this case, is the first quartile which is denoted Q1. One-fourth of the values are the same or less than 2 and three-fourths of the values are more than 2.

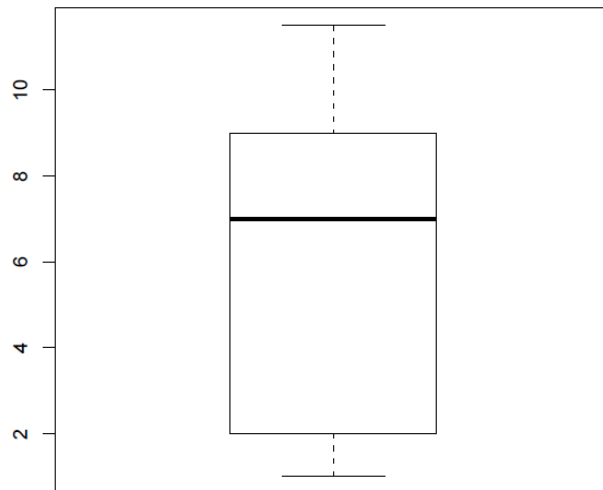


Figure Box Plot of the Example

The upper half of the data is:

7.2, 8, 8.3, 9, 10, 10, 11.5