

S4. MEASURES OF SPREAD

Measuring spread or dispersion in data:

Consider the two sets of values below:

Set A: 4, 4, 5, 5, 5, 6, 6.

Set B: 1, 3, 4, 5, 6, 7, 9.

Both groups A and B have mean = median = 5 but the data sets are quite different. The values in Set A are less spread out than those in Set B.

Range

To compare data sets it is also useful to look at the measure of spread. The most basic measure of spread is the range, the distance from the smallest to the largest value.

$$\text{Range} = \text{Largest Value} - \text{Smallest Value}$$

Set A	Set B
Range = Highest Value - Lowest Value	Range = Highest Value - Lowest Value
Range = 6 - 4	Range = 9 - 1
Range = 2	Range = 8

We can see that Set B has greater spread than Set A.

But the problem with the range is that it uses only two of the values in the data set. One of these may be an odd or unusual value called an **outlier**.

Consider the two sets of values below:

Set Y: 1, 1, 2, 2, 2, 2, 2, 100.

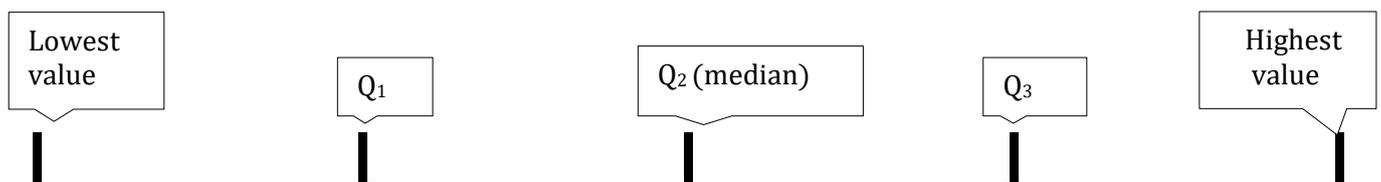
Set Z: 1, 18, 23, 41, 59, 63, 87, 100.

The range for both is 99 because Set Y has one unusual value.

Interquartile Range

The Interquartile Range (IQR) is the distance between the first quartile Q_1 and the third quartile Q_3 .

$$\text{IQR} = Q_3 - Q_1$$



The first and third quartiles are values that are $\frac{1}{4}$ and $\frac{3}{4}$ of the way through the ordered data.

Q_1 is the median of the lower half of the data and Q_3 is the median of the upper half of the data.

(NB: there are other ways to find Q_1 and Q_3 so check with your program).

