

# 4c – A Guide to Measures of Dispersion

The range of values over which a data set is found can tell you a lot about the degree of consensus within the data. This is known as **dispersion** and can be measured using some statistical analyses.

## Standard Deviation

The standard deviation (*SD*) of a data set is a measure of the extent to which the range of its values differ from the mean value.

A low standard deviation tells the reader that the values of the data set are close to the mean, and the range of the set is relatively low. This means that the reliability of any conclusions you draw, by using the mean as your indicator, will be higher.

A high standard deviation dictates that the values of the data set are spread out over a wider range and are relatively far from the mean. Therefore, any conclusions that are drawn using the mean as an indicator, will be more tenuous and less reliable.

The size of a single standard deviation on its own is somewhat meaningless, but when one compares it with another data set (such as that of another geographical site) the standard deviation starts to have more meaning.

Calculating the standard deviation for a data set is a little complicated and for very large data sets it is worth considering entering the data first into a spreadsheet which will allow you to carry out each stage of the calculation to all of the values at once. The following shows a step-by-step guide to identifying the standard deviation for a relatively small data set.

For this example, the following data set will be used:

8	12	3	5	23	19	14	8	6	16	10	8	13	9	27	16	2	9	5
---	----	---	---	----	----	----	---	---	----	----	---	----	---	----	----	---	---	---

*Distances (in km) that interviewees travelled to access the High Street of Town A.*

**Step 1.** Calculate the mean value of the data set.

$$2 + 3 + 5 + 5 + 6 + 8 + 8 + 8 + 9 + 9 + 10 + 12 + 13 + 14 + 16 + 16 + 19 + 23 + 27 = 213$$

$$213 / 19 = \mathbf{11.2 \text{ km}}$$

Step 2. For each value in the data set, subtract the mean.

8	12	3	5	23	19	14	8	6	16	10	8	13	9	27	16	2	9	5
---	----	---	---	----	----	----	---	---	----	----	---	----	---	----	----	---	---	---

minus 11.2 =

-3.2	0.8	-8.2	-6.2	11.8	7.8	2.8	-3.2	-5.2	4.8	-1.2	-3.2	1.8	-2.2	15.8	4.8	-9.2	-2.2	-6.2
------	-----	------	------	------	-----	-----	------	------	-----	------	------	-----	------	------	-----	------	------	------

Step 3. Square the resulting figures, thus removing any negative values.

-3.2	0.8	-8.2	-6.2	11.8	7.8	2.8	-3.2	-5.2	4.8	-1.2	-3.2	1.8	-2.2	15.8	4.8	-9.2	-2.2	-6.2
------	-----	------	------	------	-----	-----	------	------	-----	------	------	-----	------	------	-----	------	------	------

squared =

10.24	0.64	67.24	38.44	139.24	60.84	7.84	10.24	27.04	23.04	1.44	10.24	3.24	4.84	249.64	23.04	84.64	4.84	38.44
-------	------	-------	-------	--------	-------	------	-------	-------	-------	------	-------	------	------	--------	-------	-------	------	-------

Step 4. Calculate the mean of the figures produced from stage 3.

$$10.24 + 0.64 + 67.24 + 38.44 + 139.24 + 60.84 + 7.84 + 10.24 + 27.04 + 23.04 + 1.44 + 10.24 + 3.24 + 4.84 + 249.64 + 23.04 + 84.64 + 4.84 + 38.44 = 805.16$$

$$805.16 / 19 = 42.38$$

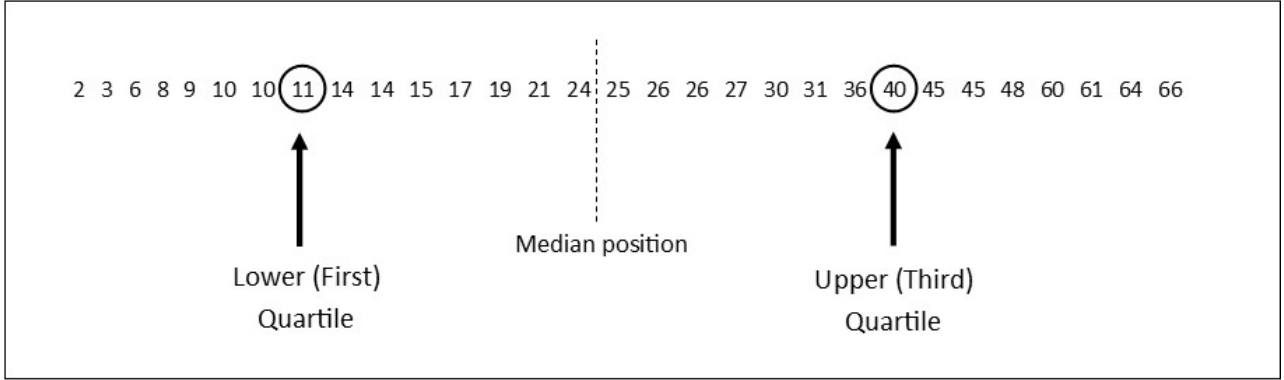
Step 5. Square root this mean. This is the standard deviation for this data set.

$$\sqrt{42.38} = 6.51 \text{ SD}$$

## Interquartile Range

A quartile is a term that refers to 25% of the values in a data set. The first (or lower) quartile is the value that, when the values are placed in numerical order, sits halfway between the median value and the minimum value. The second quartile *is* the median value and, therefore, not generally quoted as such, while the third (or upper) quartile is the value that sits halfway between the median point and the maximum value.

So, if the following set of data represents a sample of the age of visitors to a theme park on a particular day, the lower quartile is 11 years old and the upper quartile is 40 years old. The median value is 24.5 years old (the difference between the two values that sit either side of the median position).



The interquartile range is the difference between the lower and upper quartiles. In the above example, therefore, the interquartile range is 29 years (40 years old minus 11 years old).

As with standard deviation the interquartile range on its own is not wholly meaningful. The larger the interquartile range, the greater the spread of the data across the range of values. The smaller the interquartile range, the more likely the use of the median is going to be for making reliable conclusions as the spread of the values from the median is smaller.