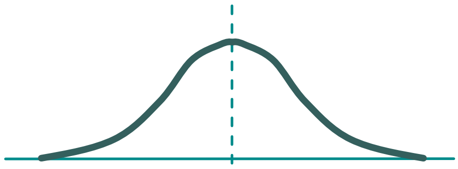
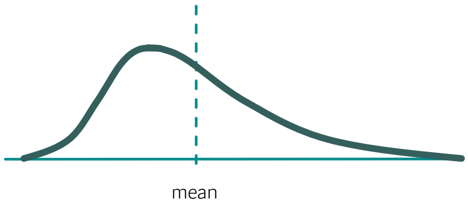
Standard Deviation

Standard Deviation is a calculation that can be used to determine how 'spread out' a sample of data is against the full range of values available. One could determine this by drawing a frequency graph - plotting how frequently certain values are seen in the sample.

A normal distribution is that in which data is evenly spread either side of the mean.



An uneven (or skewed) distribution is that in which the spread of the data favours one side of the mean over the other.



Visualising spread is one form of analysis, but Standard Deviation allows researchers to give this variance in spread a numerical figure, allowing for comparison across multiple samples to be made.

Here are some examples of geographical data one might like to analyse using Standard Deviation:

- Demography - the spread of a **population** among different age categories
- Transport - the range of **prices** people are willing to pay to access an area using public transport
- Weather - the distribution of different **precipitation** values experienced over a month
- Settlement - the spread of **footfall** values experienced in each hour of the day following a pedestrian count
- Rivers - the degree of uniformity observed in the **size of bedload** found in different parts of a stream
- Coasts - variance in **favourability scores** for different types of coastal management schemes
- Development - the spread of **life expectancy** values for a range of different regions or countries
- Tourism - the range of **visitor numbers** to a National Park throughout the summer season

How to calculate the Standard Deviation

In this example, we will look at bedload in a river channel and how it may vary in size on the outside of a meander bend and how it varies in the centre of the channel. The geographical researcher wished to find out in which part of the river bend there was the greatest level of size uniformity. They collected twenty samples in each location and measured the length of the longest axis.

Outside of meander bend:	32	46	20	8	42	24	15	6	14	48
(Bedload size in mm)	22	10	16	5	21	12	7	3	25	7

Centre of channel:	47	55	39	20	12	71	5	10	35	12
(Bedload size in mm)	7	44	16	63	80	8	29	53	50	22

The researcher than calculated the mean ($\overline{x}$) of each data set:

Outside:	$\overline{x}$ = 19.15

Centre:	$\overline{x}$ = 33.9

The following data is then calculated using the mean:

Outside:

x	$(x - \overline{x})$	$(x - \overline{x})^2$
32	12.85	165.1
46	26.85	720.9
20	0.85	0.7
8	-11.15	124.3
42	22.85	522.1
24	4.85	23.5
15	-4.15	17.2
6	-13.15	172.9
14	-5.15	26.5
48	28.85	832.3
22	2.85	8.1
10	-9.15	83.7
16	-3.15	9.9
5	-14.15	200.2
21	1.85	3.4
12	-7.15	51.1
7	-12.15	147.6
3	-16.15	260.8
25	5.85	34.2
7	-12.15	147.6
	Σ =	3552

Centre:

x	$(x - \overline{x})$	$(x - \overline{x})^2$
47	13.1	171.6
55	21.1	445.2
39	5.1	26.0
20	-13.9	193.2
12	-21.9	479.6
71	37.1	1376.4
5	-28.9	835.2
10	-23.9	571.2
35	1.1	1.2
12	-21.9	479.6
7	-26.9	723.6
44	10.1	102.0
16	-17.9	320.4
63	29.1	846.8
80	46.1	2125.2
8	-25.9	670.8
29	-4.9	24.0
53	19.1	364.8
50	16.1	259.2
22	-11.9	141.6
	Σ =	10158

The Standard Deviation (σ) equation then gives a figure for each sample, allowing the researcher to compare the spread of the data for each location in the river channel.

$$\sigma = \sqrt{\frac{\Sigma(x - \overline{x})}{n - 1}}$$

(where n = the number of values in the data set)

Outside:

$$\sigma = \sqrt{\frac{3552}{19}}$$

$$\sigma = \sqrt{186.9}$$

$$\sigma = 13.7$$

Centre:

$$\sigma = \sqrt{\frac{10158}{19}}$$

$$\sigma = \sqrt{534.6}$$

$$\sigma = 23.1$$

The larger the Standard Deviation figure, the greater the spread of the data. Therefore, in this example, the researcher can conclude that the centre of the river channel shows a greater variation in bedload size than the outside of the meander bend, where the bedload is more uniform in size.