Animation – Learning about stats

Did you know, that half of women in Britain report between 2 and 10 sexual partners over their lifetime? Here's how we got to those numbers.

The data in the Natsal-4 ‘Sex Stat for Britain tool’ has 6,719 participants, which is an amazing amount of data, but we need to use some statistical tools to help us really understand the data. For example, it can be helpful to know what people do on average, like the average number of sexual partners that people have.But there are different ways to calculate an average.

The average can be calculated by adding up all the numbers, or values that are reported by participants in the survey and dividing this total by the number of participants who responded. This type of average is called the mean.

However, very small or very large values can heavily influence or skew the mean. That is, they distort the average.

So, in this example there is a huge increase in the mean when we include a couple of extra participant’s responses. In particular, one who's an outlier in terms of their number of sexual partners.
The mean then goes from 7 sexual partners to 17.1 sexual partners, even though most values reported by the group here are smaller than this mean.

To get around this problem, we can also calculate the average as just the middle value when all the responses are placed in numerical order. This type of average is called the median.

Unlike the mean, the median isn't distorted by the inclusion of very small or very large values, reflecting how the median is just the middle value in the distribution.
And so the median is five in both of these examples.

Inter-quartile range
It can be helpful to know not just the average, like the median number of sexual partners reported, but also how spread out the data are.

For example, do most people report a number of partners that is fairly close to the median?
Or is there a greater variability in the number of sexual partners reported, with some people reporting hundreds of sexual partners so that the distribution is much more spread out?

We can calculate a statistic called the interquartile range or IQR for short, as a measure of spread that corresponds to the range between which half of response is set.

If we think of the median as the middle value when the values are put in order, then the IQR is calculated as the difference between the value that corresponds to being 3/4 of the way through the distribution (also known as the upper quartile) and the value that corresponds to 1/4 of the way through the distribution (also known as the lower quartile).

In this example, this shows that for half the people, the number of sexual partners reported lies between 2 and 10 sexual partners, resulting in an IQR of eight.

Standard deviation
If the distribution of data is not skewed, then it will be shaped more like a bell and is called a normal distribution when the data have this bell-shaped distribution then you can use either the mean or the median as the average as they will work out to be the same, or very similar.

If you use the mean as your average, then it's usual to measure the spread with a statistic called the standard deviation.

This can be worked out by hand using a formula, or most statistical packages can calculate this for you.
The larger the resulting value for the standard deviation, than the larger the variability or deviation from the mean.

Although the IQR is easier to calculate, the standard deviation is a useful statistic, as 68% of the data are always within one standard deviation of the mean.

And 95% of the data are always within two standard deviations of the mean. Statistics like the standard deviation, are useful for more complex analyses, which you can find out more about below.

www.open.edu/openlearn/britishsexstats