ACADEMI GWELLIANT CYMRU
IMPROVEMENT CYMRU ACADEMY

GIG CYMRU NHS WALES

Improvement Cymru Academy Toolkit Guide

Measures of Central Location and Range

What are Measures of Central Location and Range?

Measures of central location and range can be used to describe data to your audience. These are classed as summary statistics because they are calculated values that aim to give a summary of your data.

The three most common measures of central tendency are the mean, median, and mode. These measures attempt to describe your data by identifying the central position of the values within it.

It can be difficult to tell how a class performed in a test by looking at a long list of student scores. By applying the mean, median or mode, we can use the central tendency of the class's scores to get an idea of how the class did as a whole. This can also be used to understand how individual students did in comparison to the class as a whole.

> **1:1 Mean:** Is the most familiar measure of central tendency. It is the sum of all the data points divided by the number of data points.
>
> **1:2 Median:** Is the middle number in a sorted list of numbers i.e. the value that separates the higher half from the lower half of your data. It may be thought of as the "middle" value.
>
> **1:3 Mode:** Is the number that appears most frequently in your data. A set of numbers may have one mode, more than one mode, or no mode at all.

The **range (1:4)** does not look at the central tendency of the data and instead represents the overall spread of the data. It is the difference between the biggest and smallest value in our data and enables us to understand variation – in our classroom example this is the difference between the best and worst performances in our class for the test.

Rationale

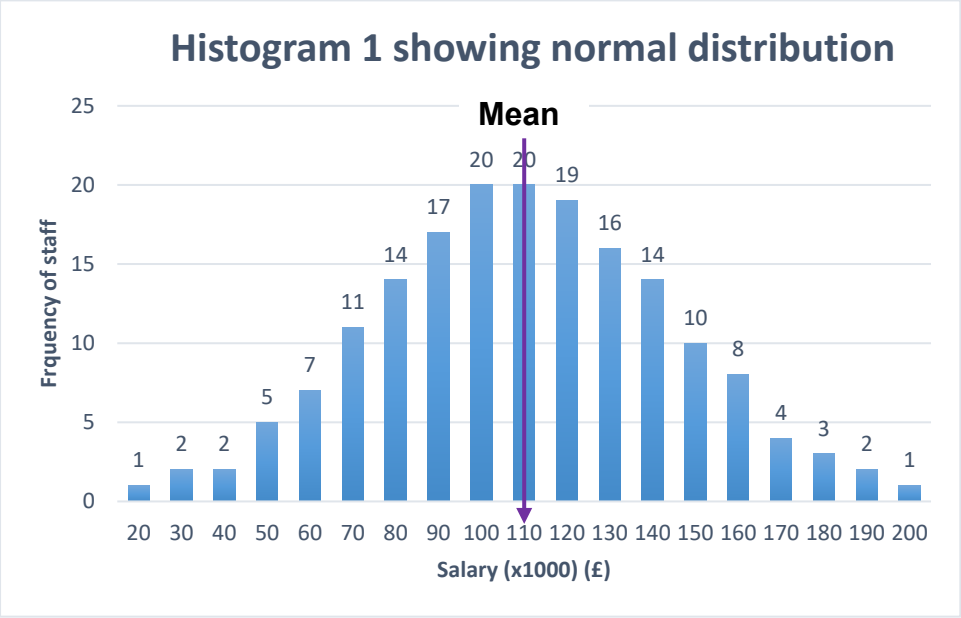
Each of these summary statistics can be used to get an idea of the data contained within our dataset. Used in combination, they can be simple but powerful tools in understanding our data.

1:1 When to use the Mean?

The mean is the most commonly used measure of central tendency, and many will be familiar with it. It is often used because of this, and because it is relatively easy to calculate.

Use of the mean is recommended when the data is normally distributed – this is where there is a small range of most likely values and the rest of the data is evenly dispersed on either side of this central range, and less likely the further we get from the centre, which will be very close to or at the mean, as seen in diagram 1.

Diagram 1



Note: Not all data follows a normal distribution and data that 'leans' to the left or the right are called skewed data. These can affect the mean. If the most likely values are high values and the rest of the data is mostly spread out toward lower values, the mean will be lower than we might expect it to be.

However, because all values are used in calculating the mean, the measure can be misleading when the data has a few very high or very low points in it (these points are often referred to as outliers and can include astronomical points). This is because the inclusion of a very low point can reduce the overall value of the mean, sometimes by a large amount (similarly, the inclusion of a very high point will increase the mean value), which can change our summary of the data as a whole.

How to calculate?

To find the mean of your data, you will need to add all of the values together and then divide by how many values there are.

$$\textbf{Mean} = \frac{\text{Sum of all values}}{\text{Count/ quantity of values}}$$

Example:

1	3	5	6	20

Step 1 - add the values together - 1 + 3 + 5 + 6 + 20 = 35

Step 2 - divide by the count/quantity of number in the set = 5

Mean = 35 ÷ 5 = 7.

1:2 When to use the Median?

Where there are no outliers and the data is not skewed (skewed data 'leans' to higher / lower values, rather than gathering around a centre point), the mean and median will give very similar results. However, whilst the median can be more difficult to calculate than the mean, it is affected less by outliers and skewed data. This means the **median is often a better option than the mean** as a measure of central tendency.

Since the mean and median will give similar results where there are no outliers or skewed data, using the mean and median in combination can reveal more about the data in your dataset.

How to calculate?

The Median is the middle value in a sorted list of values.

It is the value that separates the higher half from the lower half of a data sample and may be thought of as the "middle" value. To determine the median value in a dataset, the values must first be sorted or arranged in order of magnitude, then the middle value can be found.

An easy way to find the location of the median in a larger dataset is to take the number of values in the dataset, add 1 and then divide by 2. This gives you the location of the median when the dataset is ordered by magnitude.

Example

There are 5 values in this dataset.

Step 1 - take the number of values in the dataset and add 1: 5 + 1 = 6

Step 2 - divide by 2: 6 ÷ 2 = 3

The third value in the dataset is 5

The Median is 5.

1	3	5	6	20

When you have an even number of data, the middle value can be unclear, as the midpoint it will be between two values and the value from your location calculation will not be a whole number. In this case, the median is calculated by adding the two middle values together and dividing by two.

Example 2

There are 6 values in this dataset

We take the number of values in the dataset and add 1: 6 + 1 = 7

Then we divide by 2: 7 ÷ 2 = 3.5

The median is the mid-point of the third and fourth value

Step 1 - add the two middle values together - 3 (third value) + 4 (fourth value) = 7

Step 2 – divide the sum of the middle two values by two - 7 ÷ 2 = 3.5

The Median is 3.5

1	2	3	4	5	6

1: 3 When to use the Mode?

The mode is the number that appears most often in a data set.

A set of numbers may have one mode, more than one mode or no mode at all, which means the mode is not always representative. When other measures of central tendency can be calculated from the data, the mode is only recommended for very small datasets, for example when data is divided into small groups of race, sex, age, etc., where the other calculations may not have enough values to draw any meaningful insights.

The mode can also be used when other measures of central tendency cannot be used because we are looking at category data, for instance the colours of scrubs worn in work.

How to calculate?

To find the mode and count how many times each number occurs.

Arranging the numbers in order of magnitude (from least to most) can help avoid miscounting if doing this by hand. The number that occurs the most is the mode.

Example

1	2	2	2	3	4	5	5

Mode = 2

Sometimes you can have more than one mode, for example:

1	2	2	2	3	5	5	5

In this instance the **Mode** = 2 and 5.

1:4 When to use the Range?

The range (also known as the measure of dispersion) describes the spread of data and is simply the difference between the maximum and minimum values in a data.

While it is useful in seeing how large the spread of data is, the range can be distorted by extreme outliers and does not give any information on where most of the values lie.

How to calculate?

To find the range, subtract the smallest value from the largest value in the set. When doing this manually, it helps to first order the data from least to greatest.

Example

Range = max - min

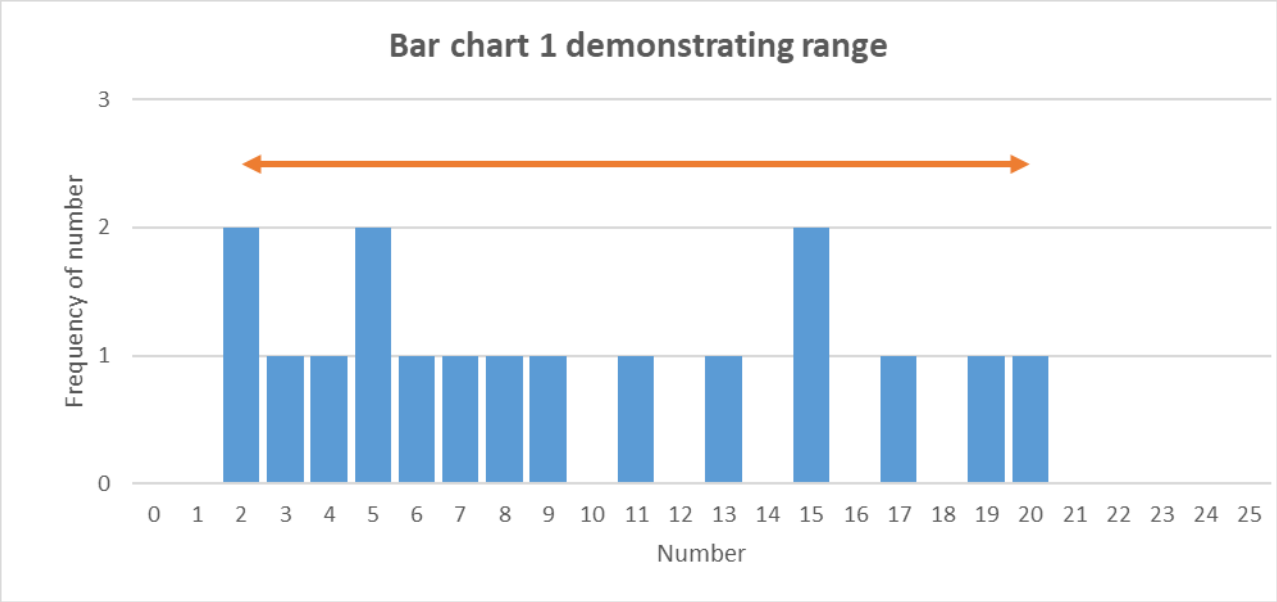
In a dataset of:

2	2	3	4	5	5	6	7	8	9	11	13	15	15	17	19	20

The range is the difference between 2 and 20: 20 – 2 = 18

The Range is 18

Diagram 3



What next?

Once the summary statistics have been generated and interpreted you can share with your team, stakeholders and service users.

Helpful Tips

Run charts use the median, SPC/Shewhart charts use the mean.

Additional resources

If you would like to learn more about making improvement to your workplace take a look at our website for what we offer you https://phw.nhs.wales/services-and-teams/improvement-cymru/improvement-cymru-academy/ or email us improvementcymruacademy@wales.nhs.uk to find about the improvement courses we can offer.

References

BBC (2020). Guides: Analysing data [online] https://www.bbc.co.uk/bitesize/guides/zsxsrwx/revision/3 Accessed 12th May 2022

Anthony McCluskey, Ghaaliq Lalkhen (2007) Statistics II: Central tendency and spread of data [online]

Statistics II: Central tendency and spread of data - ScienceDirect Accessed 17th May 2022

Saylor Academy (2012). Measures of Central Location [online]

Measures of Central Location (saylordotorg.github.io) Accessed 17th May 2022