Measurements of Spread I: Range and Interquartile Range

This guide explains two of the different types of measurement of the spread of a set of data: the range and the interquartile range.

Introduction

Measurements of spread (sometimes called **measures of dispersion** or **variation**) are one of the most fundamental and important concepts in statistics. They are part of the set of **descriptive statistics** and are crucial when you want to compare data sets.

Measurements of spread quantify the amount of scatter in a data set and, in combination with a measure of central tendency, help to give a picture of your data (see study guide: *Measurements of Central Tendency*). If the spread of values in a data set is large then the mean will not be as good a representation of the data as when the spread of data is small. Take care though, as you must consider the magnitude of the numbers in your data set when making a decision about if the scatter of the data is large or small.

Measures of spread can also help to distinguish between data sets with identical means or medians. For example consider the two data sets:

$$\{5, 25, 50, 75, 95\} \quad \text{and} \quad \{35, 40, 50, 60, 75\}$$

both of which have identical means and medians. In order to distinguish between them you need to use other properties of the set, this property is how spread out the data are.

In using and choosing an appropriate measurement of spread careful consideration should be given to what level of data you are analysing (see study guide: *Levels of Data*). For example the idea of spread for nominal level data is nonsensical and as a result you should never quote a measurement of spread as a descriptive statistic for a nominal data set. This guide introduces two simple measures of spread suitable for some ordinal, interval and ratio scales the **range** and the **interquartile range**. (The most common measures of spread, the standard deviation and variance, are explained in the study guide: *Measurements of Spread II: Variance and Standard Deviation*.)

The range

The range is the simplest measurement of spread and is defined as:

Range = highest value in a data set – lowest value in a data set

The range is easily calculated and can be useful when used in conjunction the mean.

Example: Calculate the range for the data sets:

$$A = \{5, 25, 50, 75, 95\}$$
$$B = \{35, 40, 50, 60, 75\}$$

The equation above tells you that to calculate the range of a data set you simply subtract the minimum value in that data set from the maximum value in that data set. For data set A the maximum value is 95 and the minimum value is 5 so:

Range of data set $A = 95 - 5 = 90$

For data set B the maximum value is 75 and the minimum value is 35 so:

Range of data set $B = 75 - 35 = 40$

The range is a useful tool to detect any errors in data entry. For example if the range of ages of pupils in primary school is 20 years you know that there is a data entry mistake. However the range only takes into account two data points (the smallest and the largest) and is easily distorted by extreme values (or **outliers**) in the data set. Therefore you should exercise caution when using it.

The interquartile range

The potential problem of outliers can be overcome by using of a more robust measurement of spread, the **interquartile range** or **IQR**. The IQR follows a similar idea to that of the range but does not use the highest and lowest values in the data set but those at a quarter and three-quarters of the way along when the data is ordered from smallest to largest. Therefore the IQR looks at the central section of your data, either side of the **median**. The interquartile range is defined as:

Interquartile Range (IQR) = Upper Quartile (UQ) value – Lower Quartile (LQ) value where the lower quartile lies a quarter of the way through the data and the upper quartile lies three quarters of the way through the data. The study guide: *Measurements of Central Tendency* explains how the median splits a data set in half, in a similar manner the quartiles split the data set into quarters.

Example: Ten students take a test and the results are as follows: 18, 14, 7, 13, 12, 13, 12, 19, 10 and 11. Find the range and the interquartile range for this data set (this example was used to find a median in study guide: *Measurements of Central Tendency*).

The range is the difference between the highest value and the lowest value, so:

$$\text{Range} = 19 - 7 = 12$$

To find the interquartile range, as with finding the median, the data must be arranged in order first:

7	10	**11**	12	12	13	13	**14**	18	19
		LQ					UQ		

median = 12.5

The median is the middle of the two values in the centre of the data. The lower quartile LQ is a quarter of the way through the data, halfway between the median and the lowest value in the data set, so $LQ = 11$. The upper quartile UQ is three quarters of the way through the data set, halfway between the median and the highest data value, so $UQ = 14$. Therefore:

$$IQR = UQ - LQ = 14 - 11 = 3$$

In other words the interquartile range is 3.

Example: 11 families were asked how many children they had; the replies were 2, 4, 2, 5, 0, 0, 1, 3, 1, 1 and 2. Find the range and the interquartile range for this data set (a similar example was used to find a median in study guide: *Measurements of Central Tendency*).

The range is the difference between the highest value and the lowest value, so:

$$\text{Range} = 5 - 0 = 5$$

As before, the data must be arranged in order first:

0	0	**1**	**1**	1	2	2	**2**	**3**	4	5
		LQ			median		UQ			

The median is the central value, the 6th, of the data set. The lower quartile is a quarter of the way through the data. For the data above, the lower quartile is the mean of the 3rd and the 4th data points, as they are both 1:

$$\mathrm{LQ} = \frac{1+1}{2} = 1$$

The upper quartile is three quarters of the way through the data set. For the data above, this is the mean of 8th and 9th data points and so:

$$\mathrm{UQ} = \frac{2+3}{2} = 2.5$$

Therefore:

$$\mathrm{IQR} = \mathrm{UQ} - \mathrm{LQ} = 2.5 - 1 = 1.5$$

The interquartile range is not affected by outliers in the data set and so in many cases is a better measure of spread than the range. However it has the disadvantage that the first and fourth quarters of the data are not used, so any outliers or data skew within those parts of the data would not be taken into account in this measure of spread.

Want to know more?

If you have any further questions about this topic you can make an appointment to see a **Learning Enhancement Tutor** in the **Student Support Service**, as well as speaking to your lecturer or adviser.

Call: 01603 592761
Ask: ask.let@uea.ac.uk
Click: https://portal.uea.ac.uk/student-support-service/learning-enhancement

There are many other resources to help you with your studies on our website.
For this topic there is a webcast.

Your comments or suggestions about our resources are very welcome.

Scan the QR-code with a smartphone app for a webcast of this study guide.