

# community project

encouraging academics to share statistics support resources  
All stcp resources are released under a Creative Commons licence

stcp-gilchristsamuels-6

The following resources are associated:  
Descriptive Statistics – Measures of Middle Value

## Descriptive Statistics – Measures of Spread

**Research question type:** Most

**What kind of variables:** Nominal, ordinal and interval/scale

**Common Applications:** Most quantitative studies

Descriptive statistics are a way of summarising your data into a few values, often for comparative purposes. Commonly they are presented in tables, although appropriate charts are also used. They include measures of middle values (or central tendency), and measures of spread (or dispersion) of your data.

The following is a rough guide to appropriate measures for level of data:

Data level	Middle value (central tendency)	Dispersion (spread)
Nominal	Mode	
Ordinal	Mode, median	Range, interquartile range (IQR)
Interval/scale	Median, mean	Variance, standard deviation

### Range

The range is defined as the difference between the maximum and the minimum values of a set of data.

#### Example

Find the range for: 2, 6, 3, 9, 5, 6, 2, 6.

The maximum value of the data is 9, and the minimum value is 2. Hence the range is  $9 - 2 = 7$ .

### Interquartile range (IQR)

Quartiles split a dataset into four quarters when the values are written in ascending order.

The **lower quartile** is 25% of the way through a data set. This is the  $\left(\frac{n+1}{4}\right)^{\text{th}}$  value, where  $n$  is the number of data values.

(Note: Quarter values are found by starting with the lower value and adding either  $\frac{1}{4}$  or  $\frac{3}{4}$  of the difference. For example, the  $2\frac{1}{4}^{\text{th}}$  value is the  $2^{\text{nd}}$  value plus  $\frac{1}{4}$  of the  $3^{\text{rd}}$  value minus the  $2^{\text{nd}}$ ).



The median is the middle quartile. This is the  $\left(\frac{n+1}{2}\right)^{\text{th}}$  value.

The **upper quartile** is 75% of the way through the data, or the  $3\left(\frac{n+1}{4}\right)^{\text{th}}$  value.

The **interquartile range** (IQR) is then the difference between the upper and lower quartiles.

### Examples

Find the IQR of the following:

- a) 11, 4, 9, 7, 10, 5, 6

Ordering the data gives 4, 5, 6, 7, 9, 10, 11.

There are 7 values so the  $\left(\frac{n+1}{4}\right)^{\text{th}}$  value is the second value which is 5.

The  $3\left(\frac{n+1}{4}\right)^{\text{th}}$  value is the 6<sup>th</sup> value, which is 10.

Hence the IQR is  $10 - 5 = 5$ .

- b) 1, 3, 0.5, 0.6, 2, 2.5, 3.1, 2.9

Ordering the data gives 0.5, 0.6, 1, 2, 2.5, 2.9, 3, 3.1.

There are 8 values so the lower quartile is the  $\left(\frac{n+1}{4}\right)^{\text{th}}$  value, which is the  $2\frac{1}{4}^{\text{th}}$  value.

This is 2<sup>nd</sup> value +  $\frac{1}{4}$  of (3<sup>rd</sup> value - 2<sup>nd</sup> value) =  $0.6 + \frac{1}{4} \times (1 - 0.6) = 0.6 + 0.1 = 0.7$ .

The upper quartile is the  $3\left(\frac{n+1}{4}\right)^{\text{th}}$  value, which is the  $6\frac{3}{4}^{\text{th}}$ .

This is 6<sup>th</sup> value +  $\frac{3}{4}$  of (7<sup>th</sup> value - 6<sup>th</sup> value) =  $2.9 + \frac{3}{4} \times (3 - 2.9) = 2.9 + 0.075 = 2.975$ .

Hence, the IQR is  $2.975 - 0.7 = 2.275$ .

See <http://www.mathsisfun.com/data/quartiles.html> for more detail and an introduction to **box and whisker plots**, which are useful charts to display these values.

### Variance and standard deviation

The range and IQR are quick and easy to calculate measures of spread or variation of the values in a data set. The range is influenced by extreme values, which the IQR compensates for by using just the middle 50% of the data set. However, for data measured at the scale/interval level, a measure using all the data values is more desirable.

The **variance** is such a value. We often use the related measure, its square-root, called the **standard deviation**.

Broadly speaking, the standard deviation is the average of the distances of each value in a data set from the mean of the data. Hence, the smaller the standard deviation, the more closely clustered the values are around the mean; the larger the standard deviation, the more spread-out the values are around the mean.

The variance and standard deviation of a set of numbers are defined by:

$$\text{Variance} = \text{StandardDeviation}^2 = \frac{\sum(x-\bar{x})^2}{n-1}$$



We then take the square-root to find the standard deviation.

Fortunately, the standard deviation can be easily obtained using Excel or SPSS.

**Example**

Find the standard deviation of: 6, 8, 11, 5, 2, 9, 8.

The mean of these numbers is  $\frac{6+8+11+5+2+9+8}{7} = \frac{49}{7} = 7$

$$\begin{aligned} \text{Variance} &= \frac{\sum(x - \bar{x})^2}{n-1} = \frac{(-1)^2+(1)^2+(4)^2+(-2)^2+(-5)^2+(2)^2+(1)^2}{6} \\ &= \frac{1+1+16+4+25+4+1}{6} = \frac{52}{6} = 8.67 \text{ to 2 decimal places.} \end{aligned}$$

Hence the population standard deviation =  $\sqrt{8.67} = 2.94$  to 2 d.p.

**Steps in SPSS**

**Data entry**


Open SPSS, select Type in Data, select the **Data View** tab at the bottom of the screen and enter the values into the first column of the Data Editor window as shown on the right.

Select the **Variable View** tab at the bottom of the screen to name the variable *Values*, and set the number of decimal places to zero.

The image shows two screenshots of the SPSS Data Editor. The top screenshot shows the 'Data View' tab with a grid containing the values 34 and 35 in the first column. The bottom screenshot shows the 'Variable View' tab with a table defining the variable 'Values' as a numeric type with 8 width and 0 decimals. To the right is a data table with 9 rows and 2 columns:

	VAR00001
1	6.00
2	8.00
3	11.00
4	5.00
5	2.00
6	9.00
7	7.00
8	8.00
9	

**Find the range and standard deviation**

- Choose Analyze – Descriptive Statistics – Frequencies... from the menu
- Select the Variable Values, and use  to move it to the Variable(s) list
- Click on the Statistics button
- Check the boxes next to Std. Deviation, Range (and Minimum, Maximum and Variance if desired) and Quartiles.
- Click Continue and OK

The results are given in the Output window or SPSS Statistics Viewer window. The range is 9 and the standard deviation is

The image shows the SPSS 'Frequencies' dialog box and its 'Statistics' sub-dialog box. In the 'Frequencies' dialog, 'Variable(s): Values' is selected. In the 'Statistics' sub-dialog, the following options are checked: 'Display' (under Percentile Values), 'Std. deviation', 'Range', 'Minimum', 'Maximum', and 'Variance' (under Dispersion). Other options like 'Quartiles', 'Percentile(s)', 'Mean', 'Median', 'Mode', 'Sum', 'Skewness', and 'Kurtosis' are unchecked.



2.94 as we already calculated.

The IQR is not given directly, but can be calculated by asking for the quartiles and taking the difference between them. In the above example, the IQR = 75 Percentile – 25 Percentile = 8.75 – 5.25 = 3.

**Note:** The Analyze – Descriptive Statistics – Explore... option also provides most of the descriptive statistics that you will need.

Statistics		
Values		
N	Valid	8
	Missing	0
Std. Deviation		2.726
Range		9
Percentiles	25	5.25
	50	7.50
	75	8.75

## Steps in Excel

When the data is placed in cells A1 to A7, use the formula `=STDEV.S(A1:A7)`.

The range and IQR can be calculated from the MAX, MIN and QUARTILE.EXC functions.

The range is `=MAX(A1:A7)-MIN(A1:A7)`

The IQR is `=QUARTILE.EXC(A1:A7,3)-QUARTILE.EXC(A1:A7,1)`

## Choice between measures of spread

The range is a rough measure of spread as it is only based on the extreme data values. The standard deviation is a good measure of spread when a data set is approximately normally distributed. For irregular distributions use the inter quartile range.

## Exercise

Find the range, IQR and sample standard deviation of the following datasets correct to 2 decimal places:

- 10, 16, 12, 15, 9, 16, 10, 17, 12, 15
- 74, 72, 83, 96, 64, 79, 88, 69
- £326, £438, £375, £366, £419, £424

## Answers

- 8, 6 and 2.94
- 32, 17 and 10.55
- £112, £71.50 and £42.87

