

# Calculating the correlation coefficient

Tuesday, 26 February 2019 5:56 PM

- ★ By the end of the lesson I would hope that you have an understanding and be able to apply to questions the following concepts:
- Know how to calculate Pearson's correlation coefficient by hand
  - Know how to use a CAS to calculate the correlation coefficient
  - Know how to interpret the results

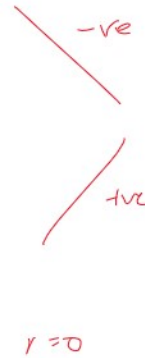
## RECAP:

In a previous lesson we were introduced to the idea of Pearson's correlation coefficient. This is basically a numerical value to give a measure of the linearity of a series of points on a scatter plot.

The values of  $r$ , Pearson's correlation coefficient, fall between **-1 and 1**. The value of  $r$  can only be used to describe linear relationships.

There was a table which we can use to convert between the value of  $r$  and a worded description.

|  |
|--|
| Strong positive association:<br>$r$ between 0.75 and 0.99    |
| Moderate positive association:<br>$r$ between 0.5 and 0.74   |
| Weak positive association:<br>$r$ between 0.25 and 0.49      |
| No association:<br>$r$ between -0.24 and +0.24               |
| Weak negative association:<br>$r$ between -0.25 and -0.49    |
| Moderate negative association:<br>$r$ between -0.5 and -0.74 |
| Strong negative association:<br>$r$ between -0.75 and -0.99  |



In this lesson we're going to use a formula to help us calculate the value of  $r$  and our CAS.

## Finding $r$ by hand

Here is the formula you are going to need to use. It looks pretty disgusting yes?!

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{(n - 1)s_x s_y}$$

$x$ :  $x$  data item  
 $y$ :  $y$  data item  
 $\bar{x}$ : Mean (or average) of the  $x$  data  
 $\bar{y}$ : Mean (or average) of the  $y$  data  
 $s_x$ : Standard Deviation of the  $x$  data  
 $s_y$ : Standard Deviation of the  $y$  data

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{(n - 1) s_x \cdot s_y}$$

Here is an example of how to use it.

Find the value of  $r$  for the following data items

|     |   |   |   |   |   |
|-----|---|---|---|---|---|
| $x$ | 1 | 3 | 5 | 4 | 7 |
| $y$ | 2 | 5 | 7 | 2 | 9 |

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{(n - 1) s_x s_y}$$

We can use the CAS to help us find the values of  $\bar{x}$ ,  $\bar{y}$ ,  $s_x$  and  $s_y$  but that comes in a different lesson. For now, know the following:

$$(n-1) s_x s_y$$

We can use the CAS to help us find the values of  $\bar{x}$ ,  $\bar{y}$ ,  $s_x$  and  $s_y$  but that comes in a different lesson.

For now, know the following:

$$\bar{x}: 4$$

$$\bar{y}: 5$$

$$s_x: 2.236$$

$$s_y: 3.082$$

| $x$ | $y$ | $x - \bar{x}$ | $y - \bar{y}$ | $x$       |
|-----|-----|---------------|---------------|-----------|
| 1   | 2   | -3            | -3            | 9         |
| 3   | 5   | -1            | 0             | 0         |
| 5   | 7   | 1             | 2             | 2         |
| 4   | 2   | 0             | -3            | 0         |
| 7   | 9   | 3             | 4             | 12        |
|     |     |               |               | <u>23</u> |

$$r = \frac{23}{4 \times 2.236 \times 3.082} = \underline{\underline{0.834}}$$

Using the CAS to find the value of  $r$



Using a calculator



VCAA Exam Question on this concept  
2016 Paper 1



10/10

