

# Y10 STATISTICS UNIT 5: SCATTER GRAPHS AND CORRELATION

## Scattergraphs:

- Used for bivariate data to show if there is a relationship between two variables.
- Explanatory variable (independent – the one that you are changing) is plotted on the x-axis.
- Response variable (dependent – the one you are measuring) is plotted on the y axis.
- Plot the points with crosses. Do not join them up!

## How to calculate SRCC

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

1. Rank both sets of data (largest to smallest)
2. Find the difference between each pair of ranks
3. Square the differences
4. Add the square of differences
5. Find the value of  $n$  – count the number of pairs of data.
6. Substitute into the formula – remember the 1 at the beginning.
7. Interpret your SRCC value in terms of correlation and strength of correlation – make this in context of the question.

## Finding the equation of a regression line:

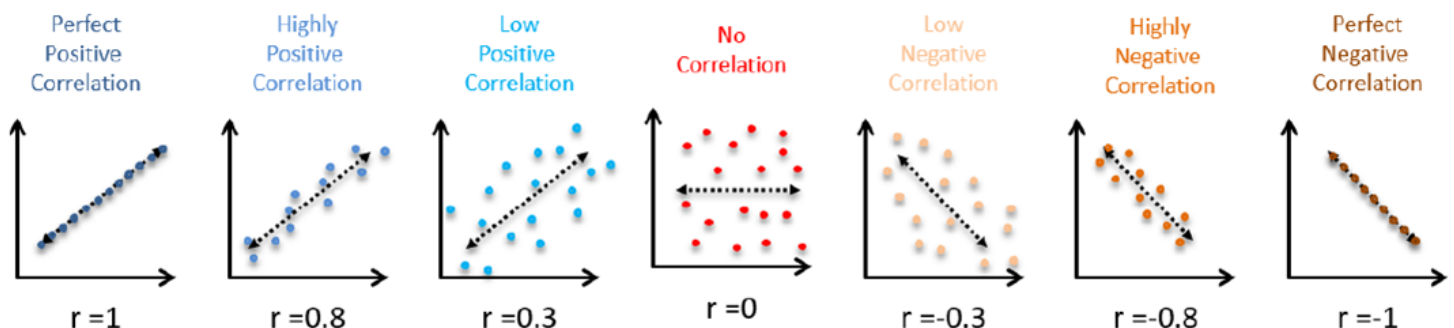
1. Select 2 points on the graph. Call these 2 points  $(x_1, y_1)$  and  $(x_2, y_2)$ .
2. Calculate the gradient,  $m$ . (Change in  $y$  over change in  $x$ )
3. Find the  $y$ -intercept,  $c$ .
4. Write the equation in the form  $y = mx + c$  with the correct values for  $m$  and  $c$ .

## PMCC

Measures the strength of **linear correlation** between two variables.

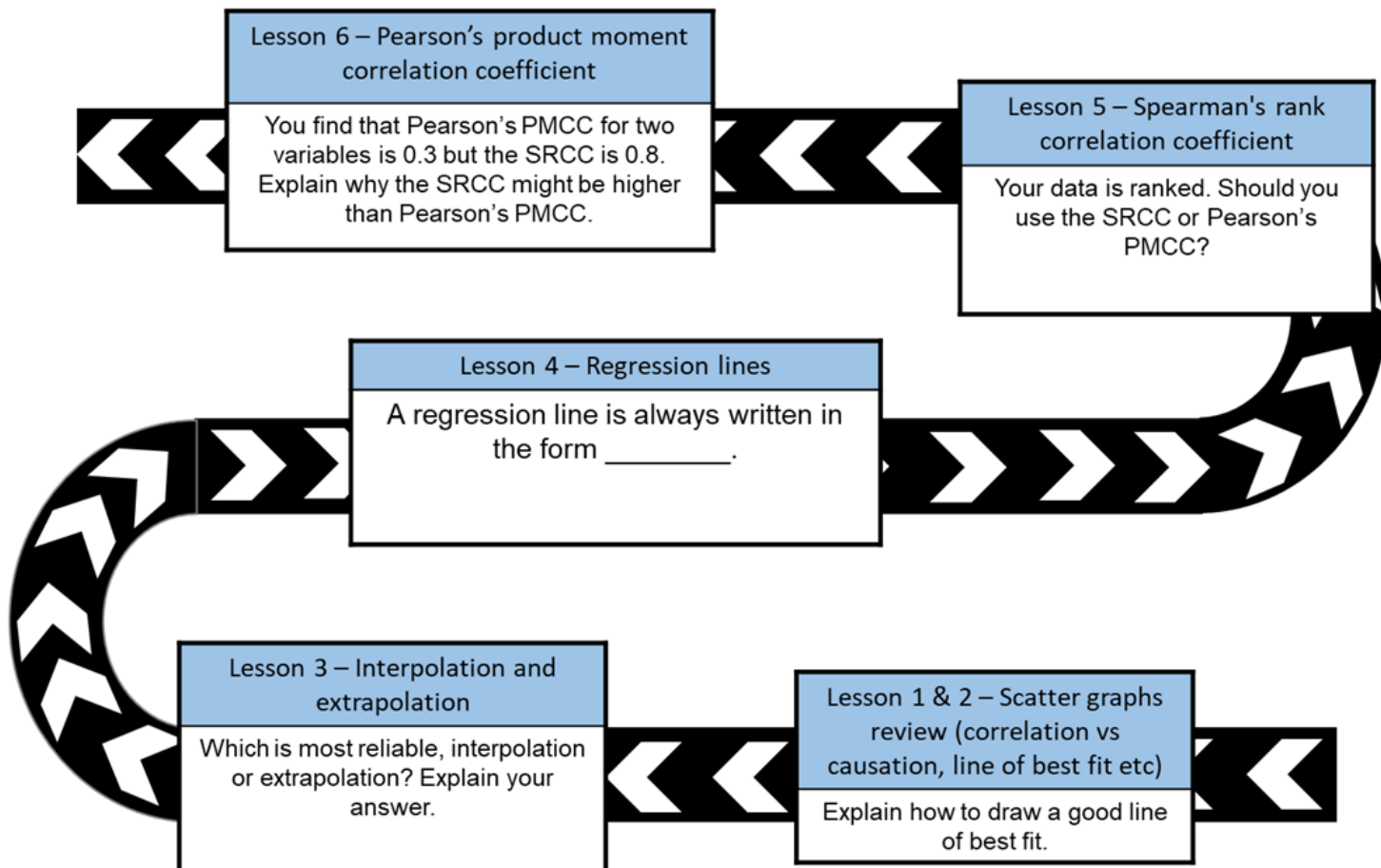
PMCC also between -1 and 1.

It is calculated using actual data values and not ranks so can be used for data that can't be ranked – don't worry you won't have to calculate PMCC.



## SRCC vs PMCC

| SRCC                                                     | PMCC                                              |
|----------------------------------------------------------|---------------------------------------------------|
| Measures the strength of correlation between 2 variables |                                                   |
| Have correlation between -1 and 1                        |                                                   |
| Tests for linear and non-linear correlation              | Tests for linear correlation only                 |
| Best used for data that can be ranked                    | Can be used for data that can't be ranked as well |



|                             |                        |
|-----------------------------|------------------------|
| <b>DEFINITION</b>           | <b>CHARACTERISTICS</b> |
| <b>ETYMOLOGY/MORPHOLOGY</b> |                        |
| <b>EXAMPLES</b>             | <b>NON-EXAMPLES</b>    |