Spearman's rank correlation coefficient

Scatter diagrams can be described as showing 'strong positive correlation' or weak 'negative correlation'. A correlation coefficient is a number that can measure the correlation between two variables.

An easy correlation coefficient to calculate is called Spearman's Rank Correlation Coefficient. This number varies between -1 and +1.

- A correlation coefficient of +1 means perfect positive correlation
- A correlation coefficient close to 0 means no correlation
- A correlation coefficient of -1 means perfect negative correlation

To calculate Spearman's Correlation Coefficient (ρ) for a set of data, you need to do three things

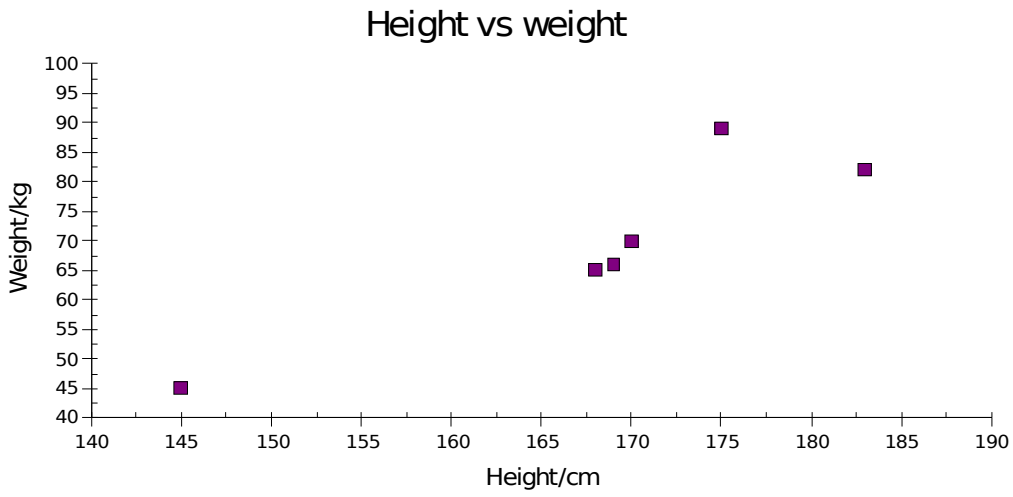
- Rank the data
- Calculate the sum of the squares of the differences of the ranks using a table
- Substitute into the formula for Spearman's $\rho = 1 - \frac{6\sum d^2}{n(n^2-1)}$

Looking at the data

Suppose measure the height and weight of 6 people...

Height/cm	145	183	175	168	169	170
Weight/Kg	45	82	89	65	66	70

And then we plot a scatter diagram of the data...



This data shows positive correlation, but is it 'strong' or 'weak'? The correlation coefficient provides an answer to this question.

Ranking the data

Suppose you had the heights of the six people in centimetres...

Height/cm	145	183	175	168	169	170

The rank for 183 cm is 1 because it is the largest height, and the rank for 175 is 2 because that is the second height in order of size. Try completing the table below – ask if you get stuck.

Height/cm	145	183	175	168	169	170
Rank	6	1	2			

Now for further practice try ranking the weights of the same six people...

Weight/Kg	45	82	89	65	66	70
Rank						

Calculating the sum of the squares of the differences

Once have ranked the data, you can calculate the differences in the ranks, and then the sum of the squares of the differences.

It is easiest to build this calculation up in a series of steps...

Height/cm	145	183	175	168	169	170
Weight/Kg	45	82	89	65	66	70
Rank Height	6	1	2	5	4	3
Rank Weight	6	2	1	5	4	3
d	0	1	1	0	0	0
d^2	0	1	1	0	0	0

The d row in the table is simply the differences between the ranks, and the d^2 row is the square of the differences.

$\sum d^2 = 1 + 1 = 2$ in this case. We will look at some more complex situations including tied ranks later on.

Calculating ρ

The Spearman's rank correlation coefficient is calculated using the formula

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

where $\sum d^2$ is the sum of the squares of the differences we just calculated from the table, and n is the number of data points (in this case six). The number 6 in the top line of the formula never changes – it is always 6.

The table below shows a step by step calculation...

$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)}$	Always state the formula
$\rho = 1 - \frac{6\times 2}{6(6^2-1)}$	Substitute in the values of $\sum d^2$ and n
$\rho = 1 - \frac{12}{6(6^2-1)}$	Work out the top of the fraction
$\rho = 1 - \frac{12}{6\times 35}$	Work out the bracket on the bottom of the fraction
$\rho = 1 - \frac{12}{210}$	Work out the full value of the bottom of the fraction
$\rho = 1 - 0.057$	Work out the decimal form of the fraction
$\rho = 0.943$	Subtract the fraction from 1 to find the correlation coefficient

The correlation coefficient of 0.943 is positive (reflecting the pattern shown by the graph). A correlation coefficient higher than 0.9 (or more negative than -0.9) shows a 'strong' relationship.

Your turn

Plot a scatter diagram of the data below. Comment on the pattern shown by the scatter diagram. Then calculate the Spearman's Rank Correlation Coefficient for this data set. Does the correlation coefficient bear out your original comments on the pattern?

Height/cm	150	165	166	175	179	181	187
Shoe size	36	42	38	44	46	48	50