

## Example: SPEARMAN'S RHO ( $\rho$ )

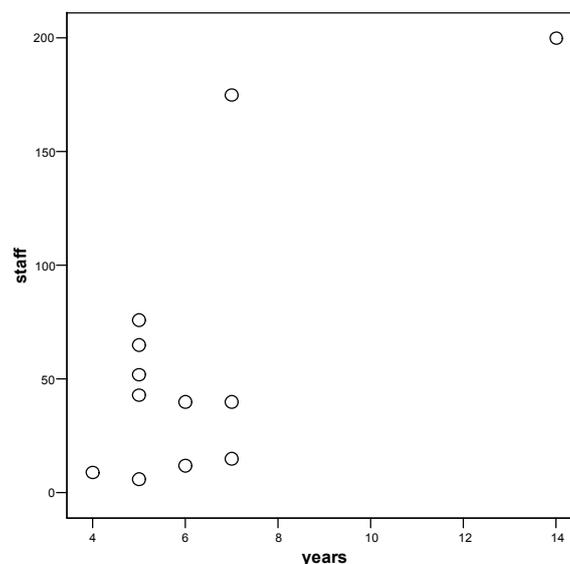
Spearman's  $\rho$  is a statistic for measuring the relationship between two variables. It is a nonparametric measure that avoids assumptions that the variables have a straight-line relationship and can be used when one or both measures is measured on an ordinal scale.

Like the conventional correlation coefficient,  $r$ , Spearman's  $\rho$  can have any value between  $-1$  and  $+1$ . A value of  $0$  indicates no relationship and values of  $+1$  or  $-1$  indicate a one-to-one relationship between the variables or 'perfect correlation'. The difference is that Spearman's rho refers to the *ranked values* rather than the original measurements.

Here are some examples of situations where Spearman's rho provides a useful summary of the degree to which two variables are related:

### Scatter diagram shows outliers: conventional correlation misleading

For a small sample of restaurant chains, the graph below plots the numbers of employees against the age of a business. Both variables are quantitative and it might seem reasonable initially to use the conventional correlation coefficient. The graph shows a small, fairly random cluster of points at the bottom left of the picture and a couple of outliers that have much higher values for one or both variables. Using the conventional correlation coefficient, you might expect to get a fairly high value, purely as the result of the two outliers. This would be misleading, because without the two outliers, there is little evidence that the number of staff depends on the age of the restaurant.



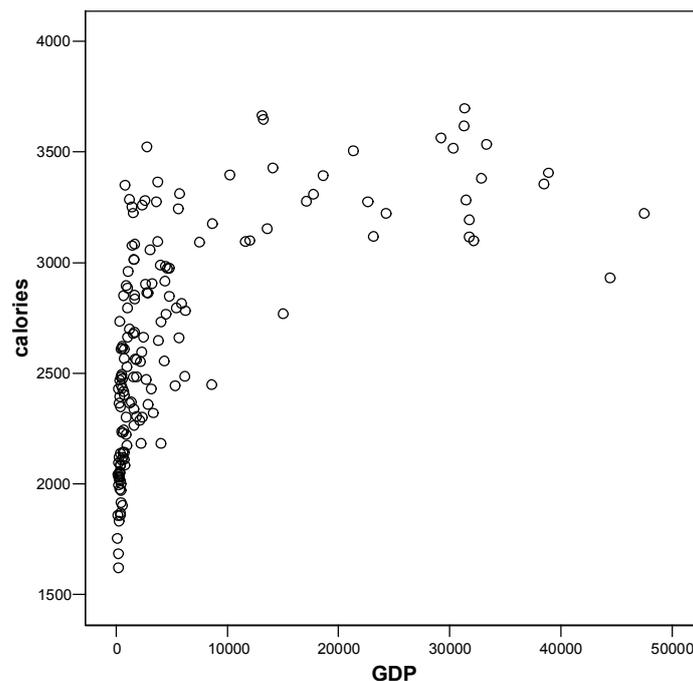
The conventional correlation coefficient produces a value of  $r = 0.731$  which, without the diagram, might be interpreted as showing a moderate/fairly strong

relationship. Spearman's  $\rho = 0.35$  shows a weak relationship and is a more accurate measure of the relationship shown in the graph.

Note: other patterns involving outliers can also produce misleading values of  $r$ .

### Scatter diagram shows non-linear relationship: conventional correlation misleadingly low

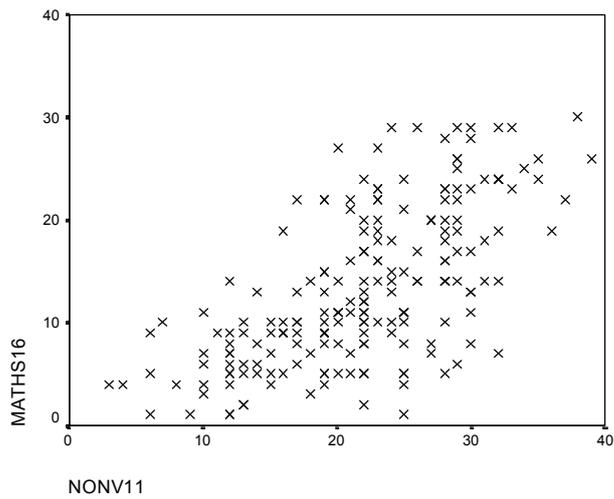
In the next scatter diagram, the mean number of calories per person is plotted against GDP for 160 countries (Source: United Nations Environment Programme: GEO Data Portal). Both variables are quantitative but the shape of the relationship is non-linear. The conventional correlation coefficient,  $r$ , is designed to measure the strength of relationships that are linear in nature.



Here the value of  $r$  is 0.615, but the value of Spearman's  $\rho$  is 0.796. The non-linear pattern shown in the scatter diagram means that  $r$  tends to under represent the strength of the relationship.

### Are nonparametric and parametric correlations always different?

When the scatter diagram shows a linear relationship between two quantitative variables, then  $r$  and  $\rho$  will have similar values. For example, in the next diagram, test scores recorded in maths for a large sample of 16 year olds are plotted against the scores they achieved at an earlier age in a test of non-verbal ability. The scatter diagram shows a positive association between the two sets of scores and the relationship is clearly linear. In this case both  $r$  and  $\rho$  are equal to 0.65 .



**One or both variables is ordinal**

Ordinal data calls for non-parametric measures. Spearman's rho measures the strength of the relationship between two ordinal variables or between an ordinal variable and a quantitative one.