EXERCISE 18

Other measures of association

Before you start

Please read Section 11.4 before proceeding with this Exercise. The **Pearson correlation** was devised to measure a supposed linear association between quantitative variables. There are other kinds of data (ordinal and nominal), to which the Pearson correlation is inapplicable. Moreover, even with data in the form of measurements, there may be considerations which render the use of the Pearson correlation inappropriate. Fortunately, other statistical measures of strength of association have been devised and in this Exercise, we shall consider some statistics that are applicable to ordinal and nominal data.

ORDINAL DATA

The Spearman rank correlation

Suppose that two judges each rank ten paintings, A, B, ..., J. Their decisions are shown in Table 1.

Table 1. The ranks assigned to the same ten objects by two judges

	Best									Worst
First Judge	C	E	F	G	H	J	I	B	D	A
Second Judge	C	E	G	F	J	H	I	A	D	B

It is obvious from this table that the judges generally agree closely in their rankings: at most, the ranks they assign to a painting differ by two ranks. But how can their level of agreement be measured?

Table 2. A numerical representation of the orderings by the two judges in Table 1

Painting	C	E	F	G	H	J	I	B	D	A
First Judge	1	2	3	4	5	6	7	8	9	10
Second Judge	1	2	4	3	6	5	7	10	9	8

The information in this table can be expressed in terms of numerical ranks by assigning the counting numbers from 1 to 10 to the paintings in their order of ranking by one judge, and pairing each of these ranks with the rank that the same painting received from the other judge, as shown in Table 2.

This is not the only way of representing the judgements numerically. It is also possible to list the objects (in any order) and pair the ranks assigned by the two judges to each object, entering two sets of ranks as before. Where the measurement of agreement is concerned, however, the two methods give exactly the same result.

Preparing the SPSS data set

In **Variable View**, name two variables, *Judge1* and *Judge2* (remembering not to put a space before the digit), and set the value in the **Decimals** column to *0*. Switch to **Data View** and enter the ranks assigned by the judges into the two columns. Save the data, because they will be used again later.

Obtaining the Spearman correlation coefficient

Draw the scatterplot using **Simple Scatter** from the **Scatter/Dot** array of diagrams in **Chart Builder**.

- **Does the scatterplot suggest good agreement between the judges?**

Select **Correlate** and then **Bivariate…** from the **Analyze** menu to open the **Bivariate Correlations** dialog box. Transfer the variables to the **Variables** box and select the **Spearman** check box (leave the default **Pearson** check box active). Click **OK** to obtain the **Pearson correlation** and the **Spearman correlation.**

- **How closely do the judges agree (state the value of the Spearman correlation coefficient)?**
- **What do you notice about the values of the two coefficients? Explain.**

Use of the Spearman rank correlation where there is a monotonic, but non-linear, relationship

Table 3. A set of paired interval data showing a monotonic, but non-linear, relationship

Y	1.00	1.58	2.00	2.32	2.58	2.81	3.00
X	2.00	3.00	4.00	5.00	6.00	7.00	8.00

Table 3 shows a set of paired interval data. On inspecting the scatterplot, we see that there is a **monotonic relationship** between the two variables: that is, as *X* increases, so does *Y*. On the other hand, the relationship between *X* and *Y* is clearly non-linear (in fact, $Y = \log_2 X$), and the **Pearson correlation** would belie that perfect association between the two variables.

Save the data from Table 2 (they will be needed later). To prepare a new data set (from Table 3) in a fresh file, enter the **File** drop-down menu, select **New** and then **Data** from the rightmost menu**.** Name the new variables in **Variable View** and enter the values into **Data View**. Obtain the **scatterplot** and compute the **Pearson** and **Spearman** correlation coefficients.

- **Describe the shape of the scatterplot and write down the values of the two correlation coefficients. Since there is a perfect (but non-linear) relationship between *X* and *Y*, the degree of association is understated by the Pearson correlation coefficient.**
- **Which value of *r* is the truer expression of the strength of the relationship between *X* and *Y* ? Explain.**

Kendall's correlation coefficients

The association between variables in paired ordinal data sets (or in paired measurements) can also be investigated by using one of **Kendall's correlation** coefficients, **tau-a**, **tau-b** or **tau-c** (see Section 11.4.2). (When there are no tied observations, **tau-a** and **tau-b** have the same value.) With large data sets, **Kendall's** and **Pearson's** coefficients give rather similar values and tail probabilities. With a given data set, however, their two values will not be identical. This is because the two statistics have quite different rationales and different sampling distributions. Their p-values, however, will usually be very similar. When the data are scarce, however, Kendall's statistics are better behaved, especially when there are tied observations, and more reliance can be placed upon the Kendall tail probability. Kendall's correlations really come into their own when the data are assignments to predetermined ordered categories (rating scales and so on).

There are two ways of obtaining **Kendall's correlations** in SPSS:

1. In the **Bivariate Correlations** dialog box, mark the **Kendall's tau-b** checkbox.
2. Use the **Crosstabs** procedure (see Section 11.5.3).

Use the **Bivariate Correlations** procedure to obtain **Kendall's tau-b** (there are no ties) for the data in Table 3. Now do the same with the data set saved from Table 2.

- **Write down the values of tau-b and compare them with your previously obtained coefficient values.**

With the restored Table 2 data set, use the **Crosstabs** procedure to obtain Kendall's correlations. Note that in this application, there is no variable such as *Count* and hence no need for **Weight Cases….** Enter *Judge1* in **Row(s):** and *Judge2* in **Column(s):**. Click the **Statistics…** button to open the **Crosstabs: Statistics** dialog box and select the checkboxes for **Correlations**, **Kendall's tau-b** and **Kendall's tau-c**. Click **Continue** and **OK** to run the correlations procedure.

- **Write down the values of all the coefficients in the output and comment on any similarities and differences.**

Finishing the session

Close down SPSS and any other windows before logging out of the computer.