NONPAR CORR

If a WEIGHT variable is specified, it is used to replicate a case as many times as indicated by the weight value rounded to the nearest integer. If the workspace requirements are exceeded and sampling has been selected, a random sample of cases is chosen for analysis using the algorithm described in SAMPLE. For the RUNS test, if sampling is specified, it is ignored. The tests are described in Siegel (1956).

Note: Before SPSS version 11.0, the WEIGHT variable was used to replicate a case as many times as indicated by the integer portion of the weight value. The case then had a probability of additional inclusion equal to the fractional portion of the weight value.

Spearman Correlation Coefficient

For each of the variables *X* and *Y* separately, the observations are sorted into ascending order and replaced by their ranks. In situations where *t* observations are tied, the average rank is assigned. Each time $t > 1$, the quantity $t^3 - t$ is calculated and summed separately for each variable. These sums will be designated ST_x and ST_y.

For each of the *N* observations, the difference between the rank of *X* and rank of *Y* is computed as:

$$d_i = R(X_i) - R(Y_i)$$

Spearman's rho (ρ) is calculated as (Siegel, 1956):

$$\rho_s = \frac{T_x + T_y - \sum_{i=1}^{N} d_i^2}{2\sqrt{T_x T_y}}$$

where

$$T_x = \frac{N^3 - N - ST_x}{12} \quad \text{and} \quad T_y = \frac{N^3 - N - ST_y}{12}$$

If T_x or T_y is 0, the statistic is not computed.

The significance level is calculated assuming that, under the null hypothesis,

$$t = \rho_s \sqrt{\frac{N-2}{1-r_s^2}}$$

is distributed as a t with $N-2$ degrees of freedom. A one- or two-tailed significance level is printed depending on the user-selected option.

Kendall's Tau

For each of the variables X and Y separately, the observations are sorted into ascending order and replaced by their ranks. In situations where t observations are tied, the average rank is assigned.

Each time $t > 1$, the following quantities are computed and summed over all groups of ties for each variable separately.

$$\tau_v = \sum t^2 - t$$

$$\tau'_v = \sum \left(t^2 - t\right)(t-2)$$

$$\tau''_v = \sum \left(t^2 - t\right)(2t+5), \text{ and } v = x \text{ or } y$$

Each of the N cases is compared to the others to determine with how many cases its ranking of X and Y is concordant or discordant. The following procedure is used. For each distinct pair of cases (i, j), $i < j$ the quantity

$$d_{ij} = \left[R\left(X_j\right) - R\left(X_i\right)\right]\left[R\left(Y_j\right) - R\left(Y_i\right)\right]$$

is computed. If the sign of this product is positive, the pair of observations (i, j) is concordant, since both members of observation i are either less than or greater than their respective measurement in observation j. If the sign is negative, the pair is discordant.

The number of concordant pairs minus the number of discordant pairs is

$$S = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \operatorname{sign}\left(d_{ij}\right)$$

where $\operatorname{sign}\left(d_{ij}\right)$ is defined as +1 or –1 depending on the sign of d_{ij}. Pairs in which $d_{ij} = 0$ are ignored in the computation of *S*.

Kendall's tau (τ) is computed as

$$\tau = \frac{S}{\sqrt{\frac{N^2 - N - \tau_x}{2}}\sqrt{\frac{N^2 - N - \tau_y}{2}}}$$

If the denominator is 0, the statistic is not computed.

The variance of *S* is estimated by (Kendall, 1955):

$$d = \frac{1}{18}\left\{K\left(2N+5\right) - \tau''_x - \tau''_y\right\} + \frac{\tau'_x \tau'_y}{9K\left(N-2\right)} + \frac{\tau_x \tau_y}{2K}$$

where

$$K = N^2 - N$$

The significance level is obtained using

$$Z = \frac{S}{\sqrt{d}}$$

which, under the null hypothesis, is approximately normally distributed. The significance level is either one- or two-sided, depending on the user specification.

References

Kendall, M. G. 1955. *Rank correlation methods*. London: Charles Griffin.

Siegel, S. 1956. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.