Tutorials and worked examples for simulation, curve fitting, statistical analysis, and plotting.
http://www.simfit.org.uk

Nonparametric correlation is required when the data are not distributed according to a multivariate normal distribution, so the Kendall-tau or else the Spearman-rank method is to be preferred. As with the Pearson product-moment correlation technique a n by m data matrix (but now with ranked or ordinal scaled data) is supplied, then SIMFIT calculates all possible pairwise correlation coefficients, and all possible two tail probabilities.

From the SIMFIT main menu choose [Statistics], [Multivariate], [Nonparametric correlation] then analyze the test file `npcorr.tf1` which contains the following data set with $n = 9$ and $m = 3$

1.70	1.00	0.50
2.80	4.00	3.00
0.60	6.00	2.50
1.80	9.00	6.00
0.99	4.00	2.50
1.40	2.00	5.50
1.80	9.00	7.50
2.50	7.00	0.00
0.99	5.00	3.00

to obtain these results.

Matrix A: Correlation coefficients
Upper triangle = Spearman's rank
Lower triangle = Kendall's tau

.....	0.2246	0.1186
0.0294	.....	0.3814
0.1176	0.2353	.....

Matrix B: Two tail p-values

.....	0.5613	0.7611
0.9121	.....	0.3112
0.6588	0.3772	.....

To be more precise, matrices A and B in this table are to be interpreted as follows. In the first matrix A, for $j > i$ in the strict upper triangle, then $a_{ij} = c_{ij} = c_{ji}$ are Spearman correlation coefficients (in black), while for $i > j$ in the strict lower triangle $a_{ij} = \tau_{ij} = \tau_{ji}$ are the corresponding Kendall coefficients (in red).

In the second matrix B, for $j > i$ in the strict upper triangle, then $b_{ij} = p_{ij} = p_{ji}$ are two-tail probabilities for the corresponding c_{ij} coefficients (in black), while for $i > j$ in the strict lower triangle $b_{ij} = p_{ij} = p_{ji}$ (in red) are the corresponding two-tail probabilities for the corresponding τ_{ij}.

For instance, because of symmetry,

- $a_{12} = c_{12} = c_{21} = 0.2246$ with $b_{12} = p-\text{Spearman}_{12} = p-\text{Spearman}_{21} = 0.5613$ refer to the Spearman rank correlation and two-tail p-values for analyzing columns 1 and 2, while
- $a_{32} = \tau_{32} = \tau_{23} = 0.2353$ with $b_{32} = p - \text{Kendall}_{32} = p - \text{Kendall}_{23} = 0.3772$ refer to the Kendall τ correlation and two-tail p-values for analyzing columns 2 and 3.

Note that, from these matrices, τ_{jk}, c_{jk} and p_{jk} values are given for all possible correlations j, k. Also, note that these nonparametric correlation tests are tests for monotonicity rather that linear correlation but, as with

the Pearson parametric test, the columns of data must be of the same length and the values must be ordered according to some correlating influence such as multiple responses on the same animals. If the number of categories is small or there are many ties, then Kendall's Tau is to be preferred and conversely. Since you are not testing for linear correlation you should not add regression lines when plotting such correlations.

It should be obvious that SimFIT displays both sets of results for convenience, and so there are just two possible ways to proceed.

1. Decide in advance which correlation coefficients and corresponding p values to accept, or
2. Apply the Bonferroni or similar correction required for two tests on the same data.

Theory

These nonparametric procedures can be used when the data matrix does not consist of columns of normally distributed measurements, but may contain counts or categorical variables, etc. so that the conditions for Pearson product-moment correlation are not satisfied and ranks have to be used. Suppose, for instance, that the data matrix, say X, has n rows (observations) and m columns (variables) with $n > 1$ and $m > 1$, then the x_{ij} are replaced by the corresponding column-wise ranks y_{ij}, where groups of tied values are replaced by the average of the ranks that would have been assigned in the absence of ties. Kendall's tau τ_{jk} for variables j and k is then defined as

$$\tau_{jk} = \frac{\sum_{h=1}^{n}\sum_{i=1}^{n} f(y_{hj} - y_{ij})\, f(y_{hk} - y_{ik})}{\sqrt{[n(n-1) - T_j][n(n-1)T_k]}},$$

$$\text{where } f(u) = 1 \text{ if } u > 0,$$
$$= 0 \text{ if } u = 0,$$
$$= -1 \text{ if } u < 0,$$
$$\text{and } T_j = \sum t_j(t_j - 1).$$

Here t_j is the number of ties at successive tied values of variable j, and the summation is over all tied values. For large samples τ_{jk} is approximately normally distributed with

$$\mu = 0$$
$$\sigma^2 = \frac{4n + 10}{9n(n-1)}$$

which can be used as a test for the absence of correlation.

Another alternative is to calculate Spearman's rank coefficient c_{jk}, defined as

$$c_{jk} = \frac{n(n^2-1) - 6\sum_{i=1}^{n}(y_{ij} - y_{ik})^2 - (T_j + T_k)/2}{\sqrt{[n(n^2-1) - T_j][n(n^2-1)T_k]}}$$

$$\text{where now } T_j = \sum t_j(t_j^2 - 1)$$

and a test can be based on the fact that, for large samples, the statistic

$$t_{jk} = c_{jk}\sqrt{\frac{n-2}{1 - c_{jk}^2}}$$

is approximately t-distributed with $n - 2$ degrees of freedom.