

4.9 Test of Association

Overview

PRIMER 8 offers several options to achieve a non-parametric bivariate test of association. Here, there are two variables sampled from the same set of sampling units and interest lies in examining the extent to which they co-vary. Do values of the two variables tend to go up and down in a similar way across the sampling units (i.e., are they positively associated)? Do we see, instead, that one variable tends to increase while the other one decreases (i.e., are they negatively associated)? Perhaps neither of these patterns occurs and the values of the two variables are essentially unassociated, going up and down independently of one another across the samples.

The null hypothesis

The essential null hypothesis tested here is H_0 : *there is no association between the two variables*. PRIMER allows the end-user to choose the particular measure of association they would like to utilise for the test, and a p -value is then generated empirically using permutations (i.e., random re-orderings of one of the variables, leaving the other fixed) to obtain an exact test under a null hypothesis of 'no relationship' (i.e., the pairing of any specific value of one variable with any specific value of the other variable within each sampling unit is arbitrary across all the pairs).

Description of the test statistics

Let x_i , be the values of a random variable, X , that have been sampled from each of $i = 1, \dots, N$ sampling units, and let y_i be the values of a different random variable, Y , that have also been sampled from the *same set* of N sampling units. There are four different measures of association that can be implemented in PRIMER to examine the potential association between the two variables. These are each described in detail below.

Pearson correlation

Let the mean of the values of x_i be $\bar{x} = \sum_{i=1}^N x_i / N$ and the mean of the values of y_i be $\bar{y} = \sum_{i=1}^N y_i / N$, then the *Pearson correlation* is:
$$\rho_{\{P\}} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}}$$
 Pearson correlation ranges from -1 to $+1$, with the two endpoints corresponding to the cases where there is a perfect linear relationship between the two variables that is either negative or positive, respectively.

Spearman rank correlation

The *Spearman rank correlation* is equivalent to the Pearson correlation calculated on ranks. Thus, let the ranks of x_i values be denoted by $r_{\{xi\}}$, and the ranks of y_i values be denoted by $r_{\{yi\}}$. Furthermore, let the mean of the ranks $r_{\{xi\}}$ be $\bar{r}_x = \sum_{i=1}^N r_{\{xi\}} / N$ and the mean of the ranks $r_{\{yi\}}$ be $\bar{r}_y = \sum_{i=1}^N r_{\{yi\}} / N$, and the Spearman rank correlation is:

$$\rho_{\{S\}} = \frac{\sum_{i=1}^N (r_{\{xi\}} - \bar{r}_x)(r_{\{yi\}} - \bar{r}_y)}{\sqrt{\sum_{i=1}^N (r_{\{xi\}} - \bar{r}_x)^2 \sum_{i=1}^N (r_{\{yi\}} - \bar{r}_y)^2}}$$

In the presence of *ties*, PRIMER first calculates the average of the ordered integers as the rank for any tied values (as described for the [Mann-Whitney](#) or [Kruskal-Wallis](#) tests), then proceeds to calculate the above equation.

In the absence of any ties, the above equation reduces to:

$$\rho_{\{S\}} = 1 - \frac{6}{N(N^2-1)} \sum_{i=1}^N (r_{\{xi\}} - r_{\{yi\}})^2$$

Weighted Spearman rank correlation

The weighted Spearman correlation was described by [Clarke & Ainsworth \(1993\)](#) and was designed primarily for cases where the intention is to obtain an index of association between two whole resemblance matrices (i.e., a *matrix correlation*; see [page 11.4 of Change in Marine Communities](#)). In that case, supposing we have a triangular matrix comprised of ranks of similarities, and that the highest similarity has a rank of 1, the next-highest similarity has a rank of 2, etc., then the calculation of a matrix correlation between this matrix and another of similar size that uses $\rho_{\{S\}}$ may not give sufficient weight to pairs of samples that are more highly similar. To give greater weight to the smaller rank values (i.e., samples having high similarity), then a weighted version is preferable. The following equation (in the absence of ties) achieves such a weighting, yet retains the appropriate scaling from -1 to $+1$. (See [Clarke & Ainsworth \(1993\)](#) for further details).

$$\rho_{\{W\}} = 1 - \frac{6}{N(N-1)} \sum_{i=1}^N \frac{(r_{\{xi\}} - r_{\{yi\}})^2}{(r_{\{xi\}} + r_{\{yi\}})}$$

Note that, importantly, we do *not* consider this to be a desirable measure of association between two *variables*, which is our focus here, but it is provided nevertheless, for completeness.[‡]

Kendall's tau

This statistic was described by [Kendall \(1938\)](#). Consider a pair of values (x_i, y_i) corresponding to the observed values of variables X and Y for sample i , and another pair of values (x_j, y_j) corresponding to those observed for sample j . The two observation pairs (corresponding to two points on a bivariate plot of X and Y) are said to be **concordant** if either one of the following two statements is true:

- $x_i < x_j$ and $y_i < y_j$; or

- $x_i > x_j$ and $y_i > y_j$.

otherwise they are said to be **discordant**. We then define the following:

- n_{con} is the number of concordant pairs,
- n_{dis} is the number of discordant pairs; and
- n_{pairs} is the total number of pairs, i.e., $n_{\text{pairs}} = N(N-1)/2$

and Kendall's tau (τ) is calculated as:

$$\tau = \frac{(n_{\text{con}} - n_{\text{dis}})}{n_{\text{pairs}}}$$

In the case of ties, PRIMER calculates a modification of the τ statistic suggested by [Kendall \(1945\)](#). Specifically, let t_k be the number of tied values for each of $k = 1, \dots, g_X$ groups of ties in the empirical distribution of observed values for X . Similarly, we can let u_{ℓ} be the number of tied values for each of $\ell = 1, \dots, g_Y$ groups of ties in the empirical distribution of observed values for Y . Kendall's tau that has been modified for tied values is then calculated as:

$$\tau_W = \frac{(n_{\text{con}} - n_{\text{dis}})}{\sqrt{(n_{\text{pairs}} - T_X)(n_{\text{pairs}} - U_Y)}} \text{ where } T_X = \sum_k t_k(t_k-1)/2 \text{ and } U_Y = \sum_{\ell} u_{\ell}(u_{\ell}-1)/2$$

Note that the Pearson, Spearman and Kendall coefficients are all scaled to yield values that range from -1 to $+1$, with values close to zero signifying a lack of any association.

Index of Association

The index of association (I_A) was first described by [Whittaker \(1952\)](#), and was subsequently used by [Somerfield & Clarke \(2013\)](#) to identify species that covary in their occurrence and relative abundances across samples. It is equivalent to a Bray-Curtis similarity index calculated between a pair of *variables* (not samples), after values have been standardised by species' totals. Notably, it is a very useful measure of the relationship between two variables that are non-negative, such as *count*, *biomass* or *percentage cover* data. For example, consider cases where x_i and y_i contain counts of the abundances for each of two species, X and Y , respectively, in each of $i = 1, \dots, N$ sampling units. It is calculated as:

$$I_A = 100 \times \left\{ 1 - \frac{1}{2} \left| \frac{x_i}{\sum_i x_i} - \frac{y_i}{\sum_i y_i} \right| \right\}$$

This index ranges from 0 (implying full 'negative' association) to 1 (implying full 'positive' association). For our purpose here in a test of association, we shall re-scale the index so that its values range usefully between -1 and $+1$. Specifically, we simply transform the raw value of I_A to the following, which we shall refer to as the *adjusted index of association*:

$$I_A^* = (2I_A/100) - 1$$

PRIMER runs the test of association on this adjusted index, yielding a more natural interpretation of the extent to which the abundances of two species (each expressed as a proportion of the total number of individuals of that species across all samples, thus accounting for species that have differing ranges, life-history strategies, etc.) either *co-occur* (+1) or are completely *disassociated* with one another (-1) across the set of samples. Making this adjustment permits a natural interpretation for I_{A}^* regarding the *direction* (positive or negative) of any potential relationship between the two species or count variables.[§] The output file also shows the value of the unadjusted index of association (scaled from 0 to 100), which can be interpreted in the usual way.[†]

Please note a couple of practical aspects of using and interpreting this index:

- First, this index ignores the information provided by joint absences. In other words, if a given sample does not house at least one individual of one of the two species being compared, then it is not considered informative, hence does not contribute towards the index measuring those two species' co-occurrence or (dis)association.
- Second, in practice, we can really only measure the association between variables that have a sufficient number of non-zero values to permit an assessment of a potential relationship. For example, if you have a species that only occurs in one sampling unit, then it cannot be sensible to talk about (let alone try to measure) whether it is associated with some other species or not. It occurs too infrequently in our dataset, so there is simply not enough information about its occurrence and abundance values to form a reasonable view.
- Third, it is not possible to construct a two-tailed test for association using this index. The reason is because the distribution of the test-statistic is not necessarily symmetric, so there is no sense in looking at 'the other tail' for any given value obtained. This means the end-user must choose *a priori* the specific alternative hypothesis desired for any particular test of association using the adjusted index - either one expects a positive association or a negative association if the null is false, but one cannot test the null hypothesis against an alternative that 'either' direction could occur.

[†] In passing, we note that the 'Test of Association' tool in PRIMER (i.e., the function `Analyse > Univariate > Association...`) is not to be used as a method of relating two resemblance matrices derived from multivariate data (even if they have each been 'unraveled' into a single long line of numbers, e.g., using `Tools > Unravel`, so that they look univariate). The p-value will be wrong if used in this way, simply because similarities (or dissimilarities) in a triangular matrix are not independent of one another, so permuting these values as if they are randomly exchangeable is incorrect. In PRIMER, if you need to do a permutation test of the degree of association between two triangular resemblance matrices, then you must use `Analyse > RELATE`. See [Chapter 14 in the PRIMER 7 Manual](#) for further details.

[§] An important additional point here is that the adjusted Index of Association, despite being scaled between -1 and +1, is not necessarily expected to have a distribution under permutation (under a true null hypothesis of 'no association') that is actually centred on zero. In fact, its permutation distribution will often be centred on some positive value. It is therefore possible to have a

significant negative association between two variables (i.e., to have their occurrences be disassociated, and for the observed value to be well beyond the left-hand tail of the permutation distribution), even though the value of the index itself (the test statistic) is greater than zero (positive). This need not pose any practical or logical problem for the end-user. Viewing the permutation distribution, and the position of the observed value relative to it, is the essential point and will be extremely informative regarding the appropriate alternative hypothesis for any given pair of variables.

[†] *In the PRIMER software, when running the test of association using the Index of Association as the measure, the test-statistic examined under permutation is the adjusted index, referred to as 'I.adj' in the output file, while the original index of association (also provided in the output, for reference) is referred to as 'IoA'.*

Revision #178

Created 19 November 2024 03:06:23 by Marti

Updated 29 November 2025 23:45:59 by Marti