4. Univariate non-parametric methods

4.1 Wilcoxon signed-rank test

Overview

The Wilcoxon signed-rank test was described by Wilcoxon (1945) . It is designed for the situation where there are **two groups** of values, and any individual value in one group is **paired** with a specific value in the other group. For example, you might have a treatment and a control value for a given response variable across each of a number of different trials. Interest lies in making a formal comparison of treatment *vs* control values, acknowledging the inherent non-independence of the paired values within each trial. This test is a non-parametric analogue to a classical paired *t*-test, and may be implemented in PRIMER under either a directional (one-tailed) or non-directional (two-tailed) alternative hypothesis.

The null hypothesis

The essential null hypothesis tested here is H_0: *the distribution of paired differences is stochastically symmetric about zero*. In other words, the rank order of the values within each pair is arbitrary, so the two observations within any pair are exchangeable with one another. We might consider writing this as H_0: the median of the distribution of paired differences is equal to zero.

Our alternative hypothesis may be *non-directional* and simply assert that the paired differences are stochastically symmetric about some other value that is *not zero*. This might be phrased as H_A: the median of the distribution of paired differences is not equal to zero. In that case, we have a two-tailed test.

We may, however, assert a *directional* alternative hypothesis, yielding a one-tailed test. For example:

- H_A: The distribution of paired differences is symmetric about some value **greater** than zero (e.g., the median of the distribution of paired differences is positive); or
- H_A: The distribution of paired differences is symmetric about some value **less** than zero (e.g., the median of the distribution of paired differences is negative).

Description of the test

Consider a set of N sampling units. For each sampling unit $i = 1, \ldots N$, two paired (or matched) observation values have been recorded: (x_i,y_i). For any pair i, let the difference in these paired values be $d_i = (x_i - y_i)$. Next, let r_i be the rank of the absolute value of the differences $|d_i|$, with the smallest absolute difference being given a rank of 1 and the largest absolute difference being given a rank of N. Let $\text{sgn}(\cdot)$ be a function to attribute an indicative sign such that $\text{sgn}(d_i) = 1$ if $d_i>0$ and $\text{sgn}(d_i) = -1$ if $d_i<0$. We can obtain the *signed ranks* as: $r_i^{\text{sgn}} = \text{sgn}(d_i) \cdot r_i$.

We define the test statistic, W, as the sum of all the signed ranks, i.e.: $$ W = \sum_{i = 1}^N \text{sgn}(d_i) \cdot r_i $$

Having obtained an observed value of the test statistic from the data, W_obs, then (as in many other PRIMER routines), we can obtain a *p*-value empirically using an appropriate permutation algorithm. Specifically, under the assumption of exchangeability, we can generate a plausible value of the test-statistic W under a true null hypothesis by randomizing the ordering of the paired values (x_i,y_i) separately, within each pair, for each and every sampling unit $i = 1,...,n$. Once this randomization has been done, we can re-calculate the differences, d_i^π ,and their associated (unsigned) ranks, r_i^π, under permutation, to yield:

$$ W^\pi = \sum_{i = 1}^N \text{sgn}(d_i^\pi) \cdot r_i^\pi $$

We repeat the above randomization and re-calculation procedure a large number of times (e.g., say n_perm = 9999) to obtain a large number of values of W^π under a true null hypothesis. The probability (*p*-value) associated with the null hypothesis (and two-tailed alternative hypothesis) is then estimated empirically as the proportion of values of W^π that are equal to or more extreme (in absolute value) than the observed value of the test-statistic, W_obs. Thus, letting W_k^π be the value of W^π obtained for the kth permutation ($k = 1, \ldots, n_\text{perm}$), the *p*--value is:

$$ P = \frac{ \sum_{k=1}^{n_\text{perm} } (\text{I}(|W_k^\pi| \geq |W_\text{obs}|) + 1)}{(n_\text{perm} + 1)} $$

with $\text{I}(\textit{expression}) = 1$ if $\textit{expression}$ is true and zero otherwise. Note that the '$+1$' in the numerator and denominator of this fraction is there to acknowledge the inclusion of the observed value as a member of the distribution of W under a true null hypothesis.

One-tailed alternative hypotheses

As noted above, we may postulate a more specific alternative hypothesis. In such cases, the test-statistic and randomization procedure are all done the same way as described above, but the *p*-value is calculated differently. For example, if our alternative hypothesis is that the median of the distribution of paired differences is *greater than zero*, then the *p*-value is calculated as:

$$ P = \frac{ \sum_{k=1}^{n_\text{perm}} (\text{I}(W_k^\pi \geq W_\text{obs}) + 1)}{(n_\text{perm} + 1)} $$ which tallies only the values of W_k^π that equal or positively exceed W_obs, in the right-hand tail of the permutation distribution.

If, on the other hand, our alternative hypothesis is that the median of the distribution of paired differences is *less than zero*, the *p*-value is calculated as:

$$ P = \frac{ \sum_{k=1}^{n_\text{perm}} (\text{I}(W_k^\pi \leq W_\text{obs}) + 1)}{(n_\text{perm} + 1)} $$ which tallies only the values of W_k^π that equal or negatively exceed W_obs, in the left-hand tail of the permutation distribution.

Treatment of tied ranks

What happens to the rank-order values, r_i, in the event of a *tie*? Let's suppose $|d_1|$ is the smallest absolute difference (so is given a rank of $r_1 = 1$), that $|d_2|$ is the second-smallest absolute difference (hence $r_2 = 2$), but then we find that a third value for the difference $|d_3|$ is precisely (within double precision) equal to $|d_2|$? In other words, suppose $|d_2|$ and $|d_3|$ are tied for second place in the ranking of values from smallest to largest.

In this case PRIMER takes a fairly standard approach and *averages* the ranks of tied values. We simply order the values from smallest to largest and give them 'raw' ranks that correspond to the set of ordered integers, i.e., from 1 to n). Then, we replace the ordered integers for any tied values with the average of those integers. Thus, in this case, the ordered integers are $\{1, 2, 3, \ldots \}$. We have $r_1 = 1$, but then both r_2 and r_3 would be given the average of the ordered integers in their place, i.e., $r_2 = r_3 = (2+3)/2 = 2.5$. So, the set of ranks used for the analysis will be $\{1, 2.5, 2.5, \ldots \}$. The next largest absolute difference would be given the (unsigned) rank value of 4 (presuming it is not tied), and so on. Any other tied values are treated in the same way, by averaging their corresponding ordered integer values, and all subsequent calculations to calculate the test statistic, etc., simply carry on from there precisely as described above.

An important point here about the PRIMER implementation of the Wilcoxon signed-rank test is that, given that *p*-values are calculated using an appropriate randomization algorithm under a true null hypothesis of exchangeability, these non-parametric tests are indeed exact tests, even in the event of there being ties in the ranks. This contrasts with other available software implementations of the Wilcoxon test (e.g., such as 'wilcox.test()' in R), which do not compute exact *p*-values if there are tied ranks.

Treatment of tied pairs (difference = 0)

It is possible to obtain paired values that are identical to one another; i.e., $x_i = y_i$, so $d_i = 0$. This is problematic for the test, not because of the ranking procedure (it would clearly be of very low rank, given its small absolute value), but because, for this sampling unit, there is therefore *no sign to attribute to it*. The simplest solution is to omit differences equal to zero from the calculation (Wilcoxon (1949)). This is the approach that is implemented in PRIMER.

It might be argued, however, that this withdraws certain evidence in favour of the null hypothesis. An alternative approach, suggested by Pratt (1959) , would be to include the zeros when ranking the absolute differences, but subsequently to exclude them from the calculation of the test-statistic; i.e., to thereafter assert that $\text{sgn}(0) = 0$.

From a practical point of view, neither the Pratt (1959) , nor the Wilcoxon (1949) method was found to be universally most efficient (Conover (1972)). Also, as PRIMER does not rely on any formal derivation of the distribution of the W test statistic, but rather uses permutation algorithms to estimate a *p*-value empirically, it would seem that the choice between these two options would have little or no substantive effect on the outcome of PRIMER's implementation of

the Wilcoxon test, although this has not been investigated in detail.

Original test-statistic

We note here, in passing, that the original Wilcoxon test-statistic was defined somewhat differently from the above description, which uses W. First, let R^+ be the sum of the positively signed ranks, i.e. $$ R^+ = \sum_{i=1}^N r_i^{\text{sgn}} \cdot \text{I}(d_i>0) $$ and let R^- be the sum of the negatively signed ranks, i.e., $$ R^- = \sum_{i=1}^N r_i^{\text{sgn}} \cdot \text{I}(d_i<0). $$ Wilcoxon's original test-statistic is then given as the minimum of the absolute values of these two quantities; that is, $$ T = \text{min}(|R^+|, |R^-|). $$

The PRIMER implementation provides both W and T in the output file for this test. Note that the calculation of a *p*-value using an appropriate permutation algorithm under a true null hypothesis of exchangeability ensures that PRIMER achieves an exact test, no matter which test statistic you prefer to report from the output provided.

4.2 Example: Plankton hauls

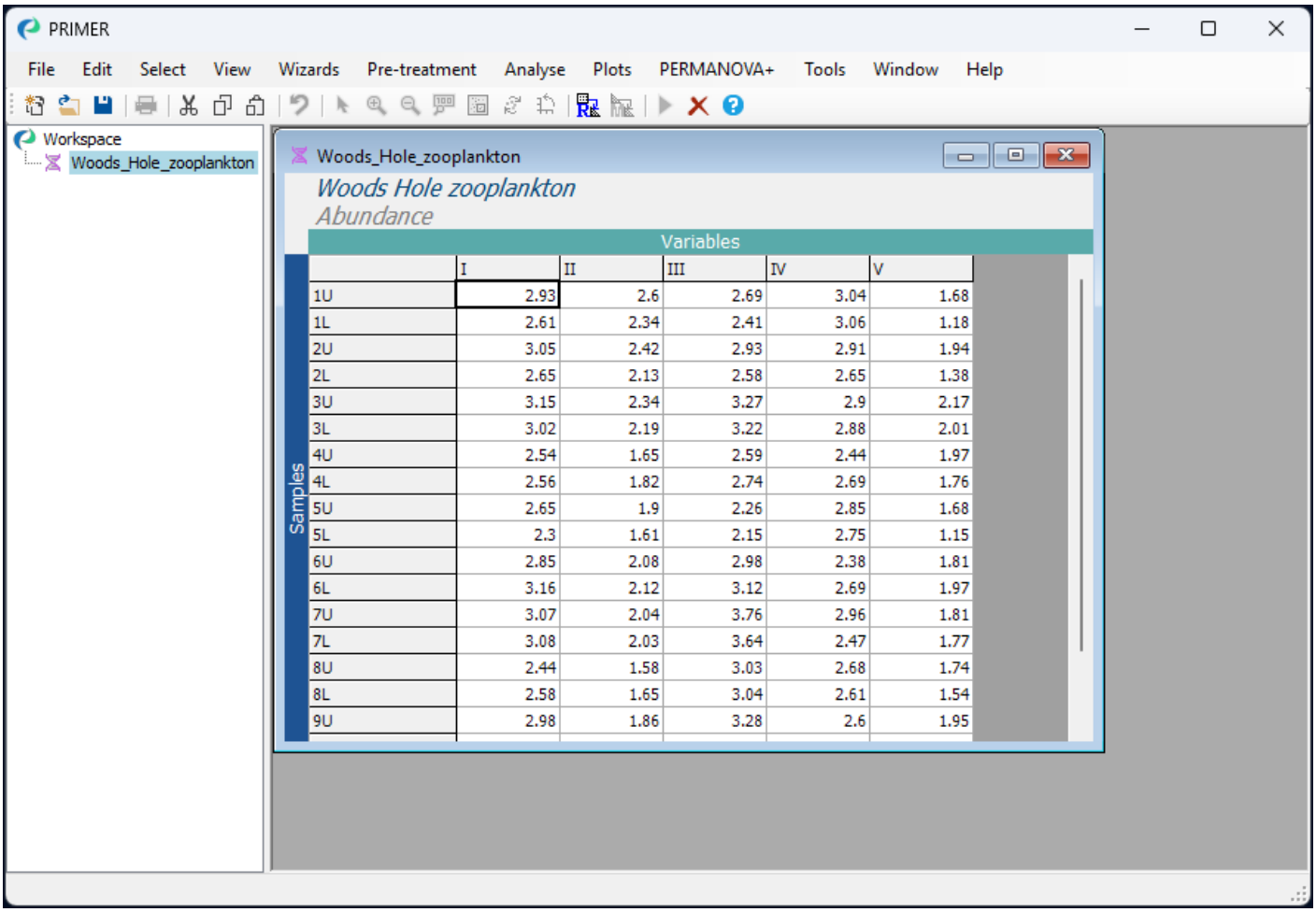
An example of a paired design with two groups is provided by Snedecor (1946) , who described a study by Winsor & Clarke (1940) to investigate the total catch of five different groups of plankton (hence, five variables, named using Roman numerals I, II, III, IV and V) by 2 nets hauled horizontally behind a boat. One net was 2 metres below the other one. More specifically, ten hauls were made with the pair of nets situated at depths of 29 m and 31 m. The factors associated with these data are:

- Position (either the upper (U) or the lower (L) net); and
- Haul (10 hauls, labeled 1-10).

Clearly, the two nets (upper and lower) being hauled at the same time are ‘paired’ with one another, so the ‘Haul’ factor is the factor identifying the different pairs of samples here. These data are located in the file ‘Woods_Hole_zooplankton.pri’, found inside the 'Examples_P8 > Woods_Hole_zooplankton' folder. Here, we are going to analyse a single variable – the **sum** of the log abundances of all plankton types. Specifically, we are going to test the null hypothesis that the paired differences between the upper and lower nets are stochastically symmetric about zero (i.e., there is no effect of ‘Position’).

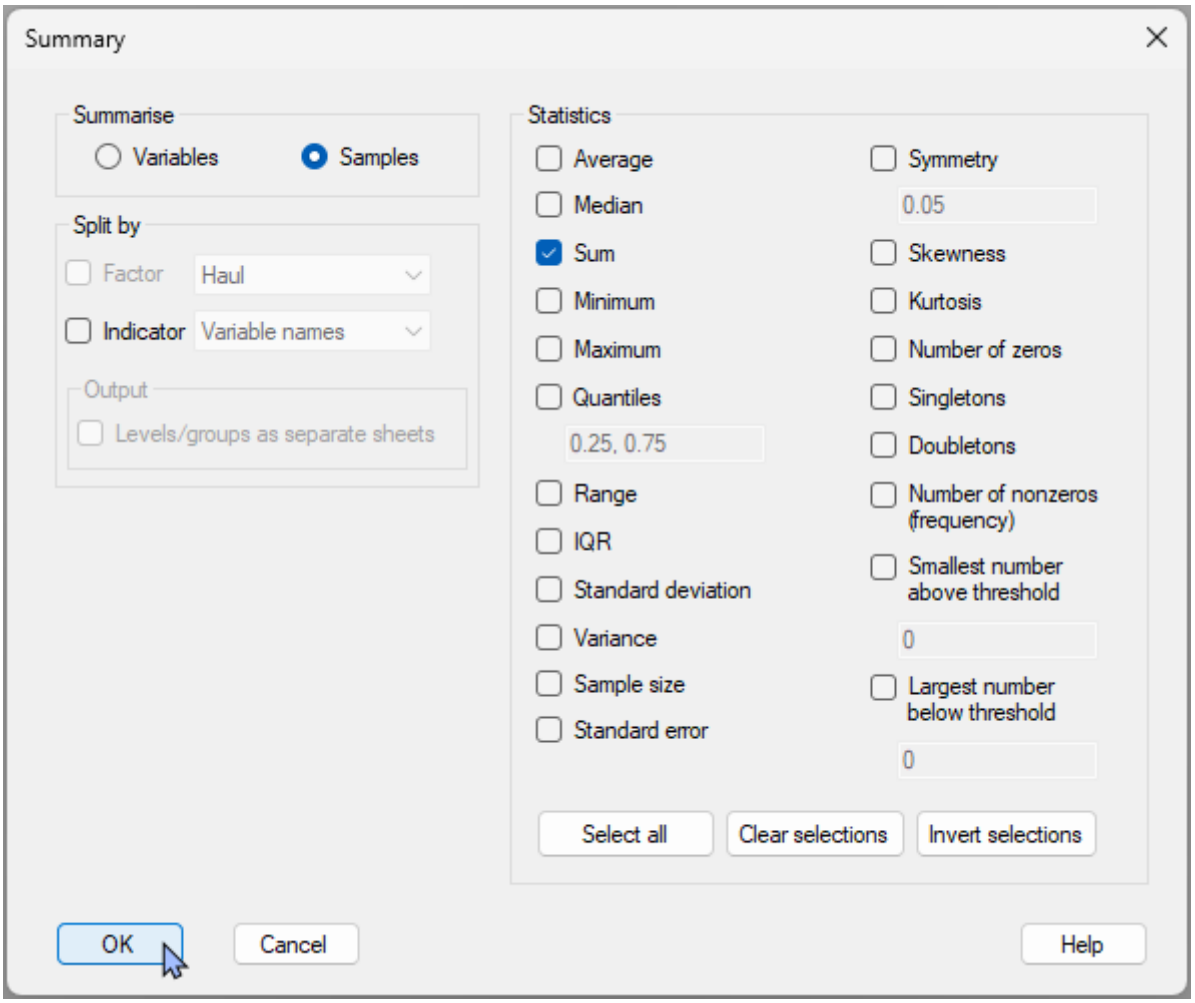
Running the Wilcoxon signed-rank test

1. Open up the file (‘Woods_Hole_zooplankton.pri’) in PRIMER and note that the abundances are already expressed as log abundance values, so there is no need to apply any transformation.

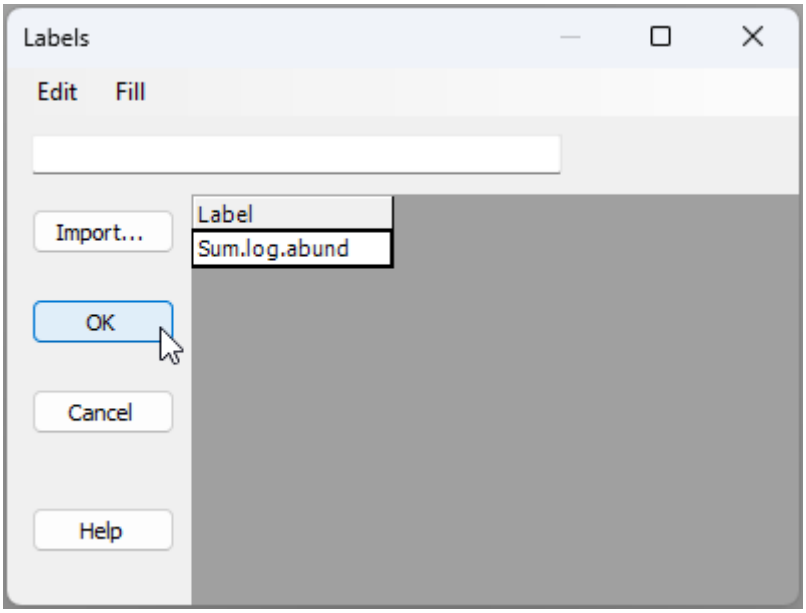


	I	II	III	IV	V
1U	2.93	2.6	2.69	3.04	1.68
1L	2.61	2.34	2.41	3.06	1.18
2U	3.05	2.42	2.93	2.91	1.94
2L	2.65	2.13	2.58	2.65	1.38
3U	3.15	2.34	3.27	2.9	2.17
3L	3.02	2.19	3.22	2.88	2.01
4U	2.54	1.65	2.59	2.44	1.97
4L	2.56	1.82	2.74	2.69	1.76
5U	2.65	1.9	2.26	2.85	1.68
5L	2.3	1.61	2.15	2.75	1.15
6U	2.85	2.08	2.98	2.38	1.81
6L	3.16	2.12	3.12	2.69	1.97
7U	3.07	2.04	3.76	2.96	1.81
7L	3.08	2.03	3.64	2.47	1.77
8U	2.44	1.58	3.03	2.68	1.74
8L	2.58	1.65	3.04	2.61	1.54
9U	2.98	1.86	3.28	2.6	1.95

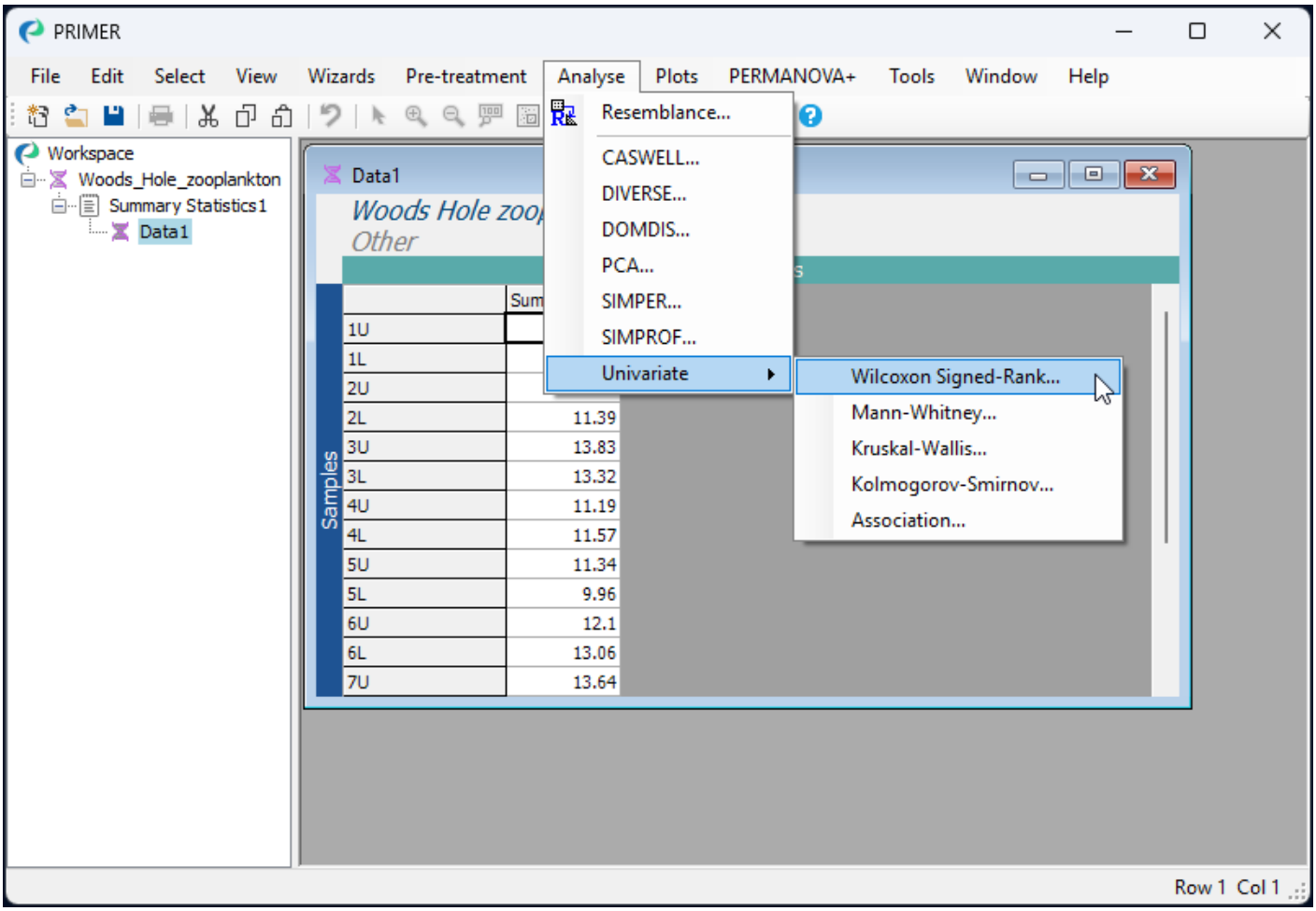
2. Click **Tools** > **Summary Stats…** then (Summarise $\bullet$Samples) & (Statistics $\checkmark$Sum), and click '**OK**'. {*Note:* Be sure to *untick* the default tickbox, ($\square$ Average), as here we only want to get the **sum** across the five variables in order to analyse the sum of log abundances across the 5 variables for each sample, as a univariate variable}.



3. The resulting data sheet will be called 'Data1'. From this data sheet, keep things tidy by re-naming the variable from 'Sum' to 'Sum.log.abund'. Click **Edit** > **Labels** > **Variables…** and make this change, as shown below:



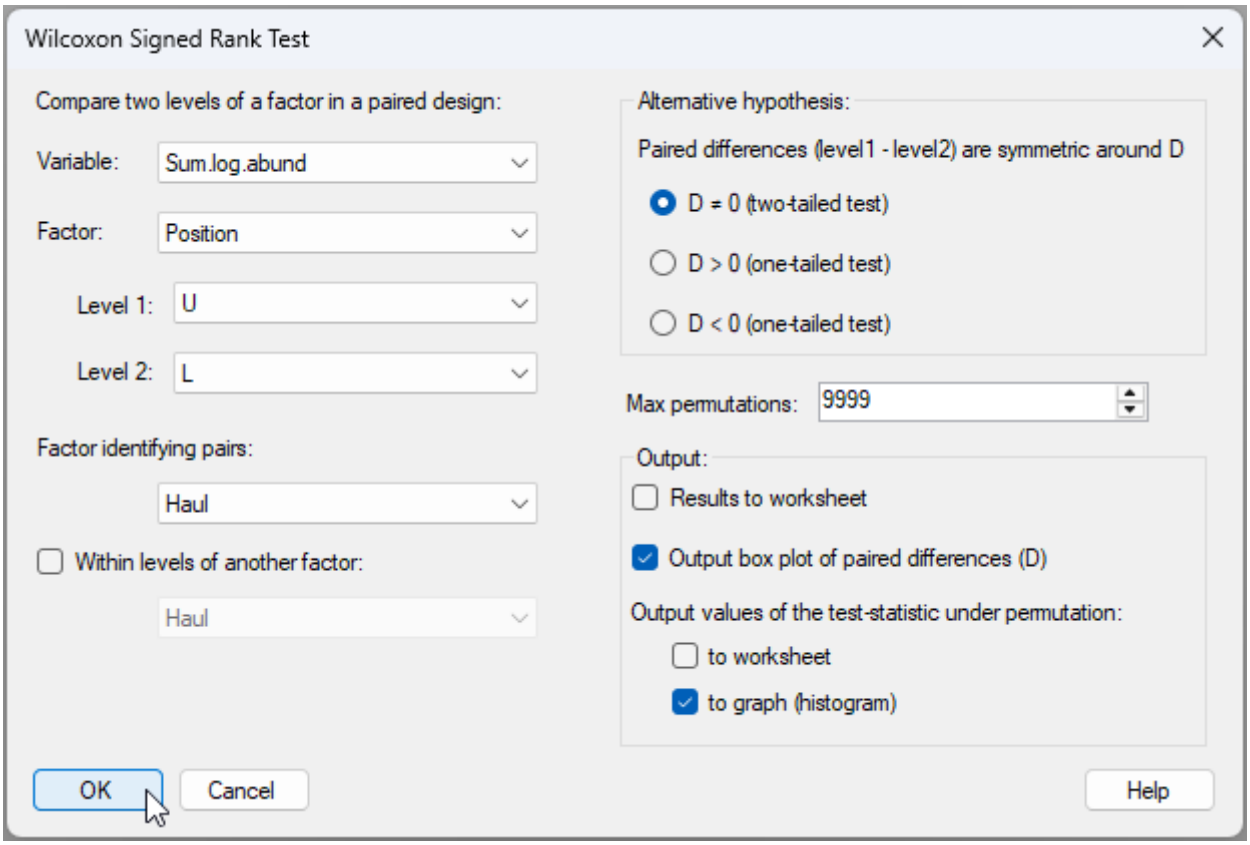
4. Let's now do the test. From 'Data1', click **Analyse** > **Univariate** > **Wilcoxon Signed-Rank…**



5. Choose the following in the Wilcoxon Signed Rank Test dialog:

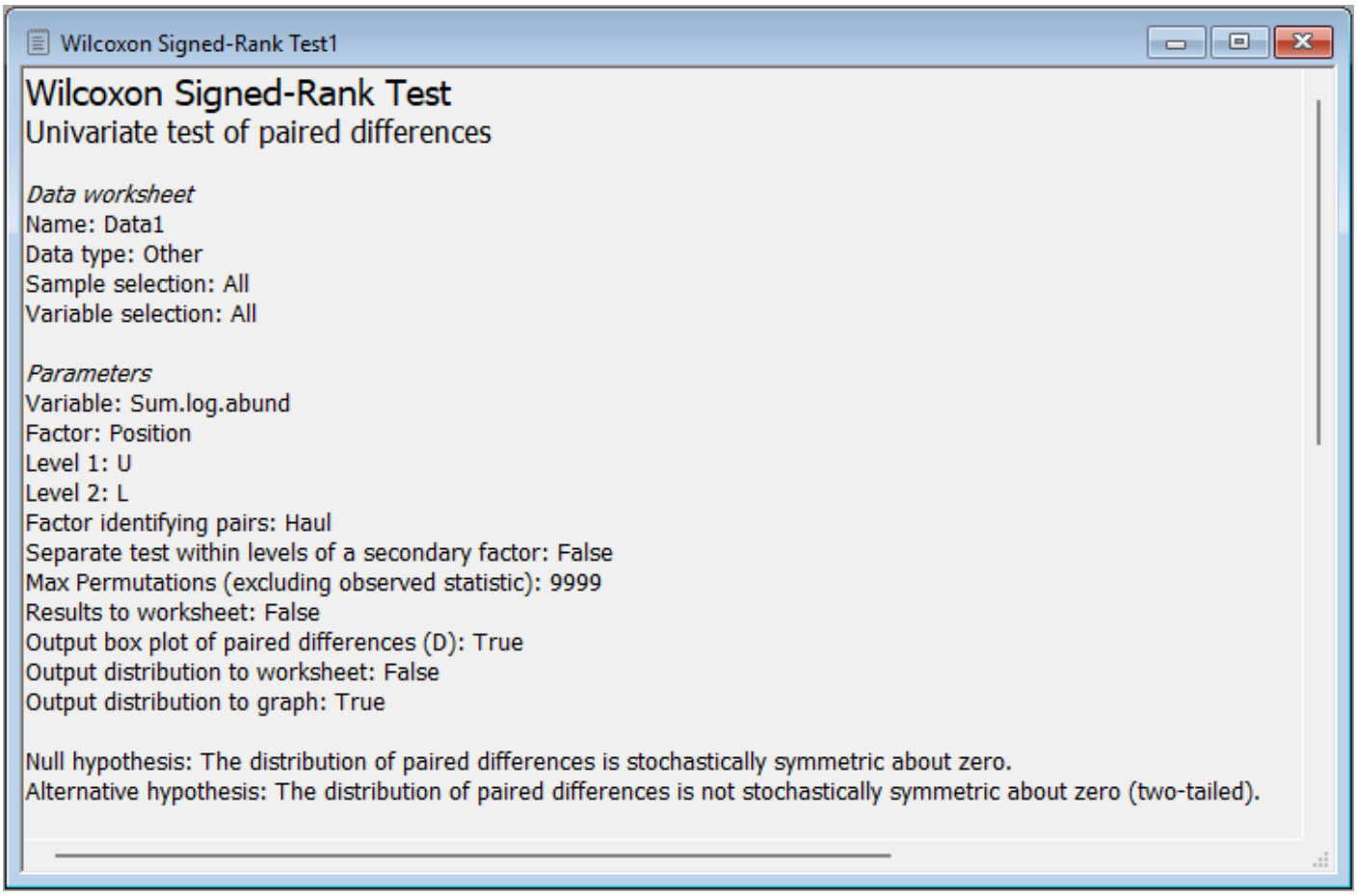
Variable: Sum.log.abund
Factor: Position
    Level 1: U
    Level 2: L
Factor identifying pairs: Haul
Alternative hypothesis: Paired differences (level1 - level2) are symmetric around D
    $\bullet$D ≠ 0 (two-tailed test)
Max permutations: 9999
Output:
$\checkmark$Output boxplot of paired differences (D)
Output values of the test-statistic under permutation:
    $\checkmark$to graph (histogram)

then click '**OK**', as shown below.

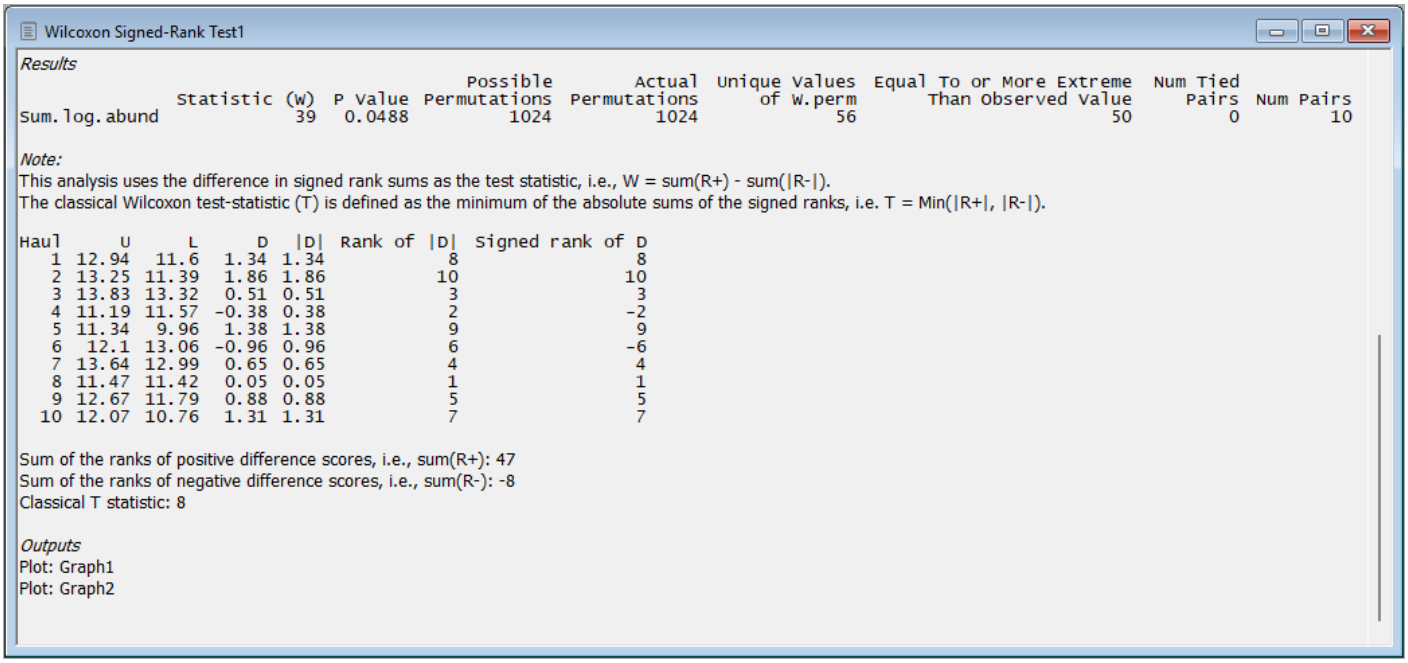


Results of the Wilcoxon test

The resulting notepad , *.rtf) file named 'Wilcoxon Signed-Rank Test1' contains all of the essential elements of this analysis and its results. First are shown the choices that were made by the user ('*Parameters*'):



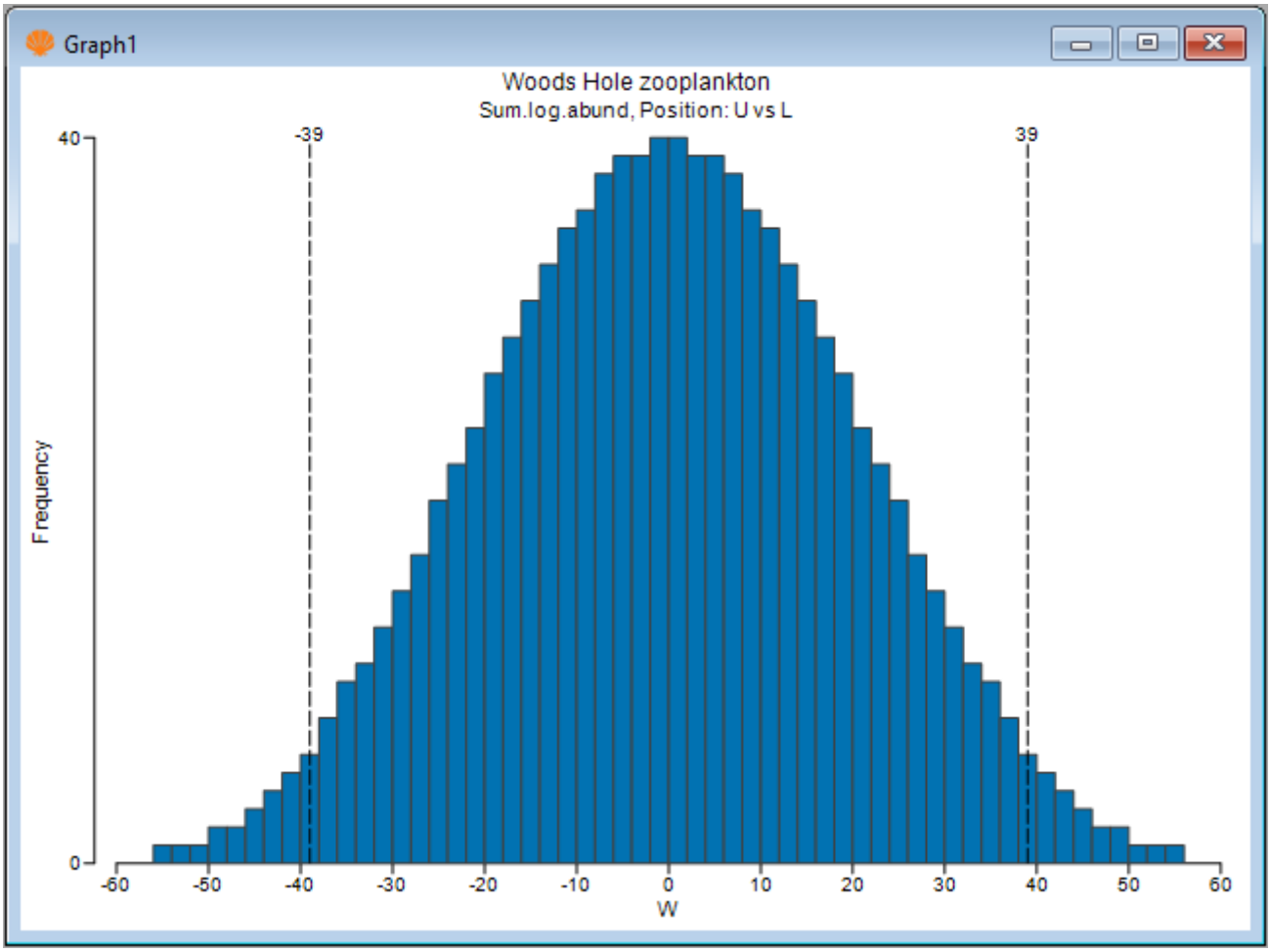
This is followed by the full suite of results, including all of the relevant supporting calculations ('*Results*'), *viz.*:



There was a statistically significant difference between these two nets (upper *vs* lower) in the sum of log abundances of plankton captured (W = 39, P = 0.0488). Furthermore, the positive value of W indicates that 'Level 1' (which in our case was 'U', the upper net, towed at a shallower depth) typically had larger values than 'Level 2' ('L', the lower net, towed at a slightly deeper depth).

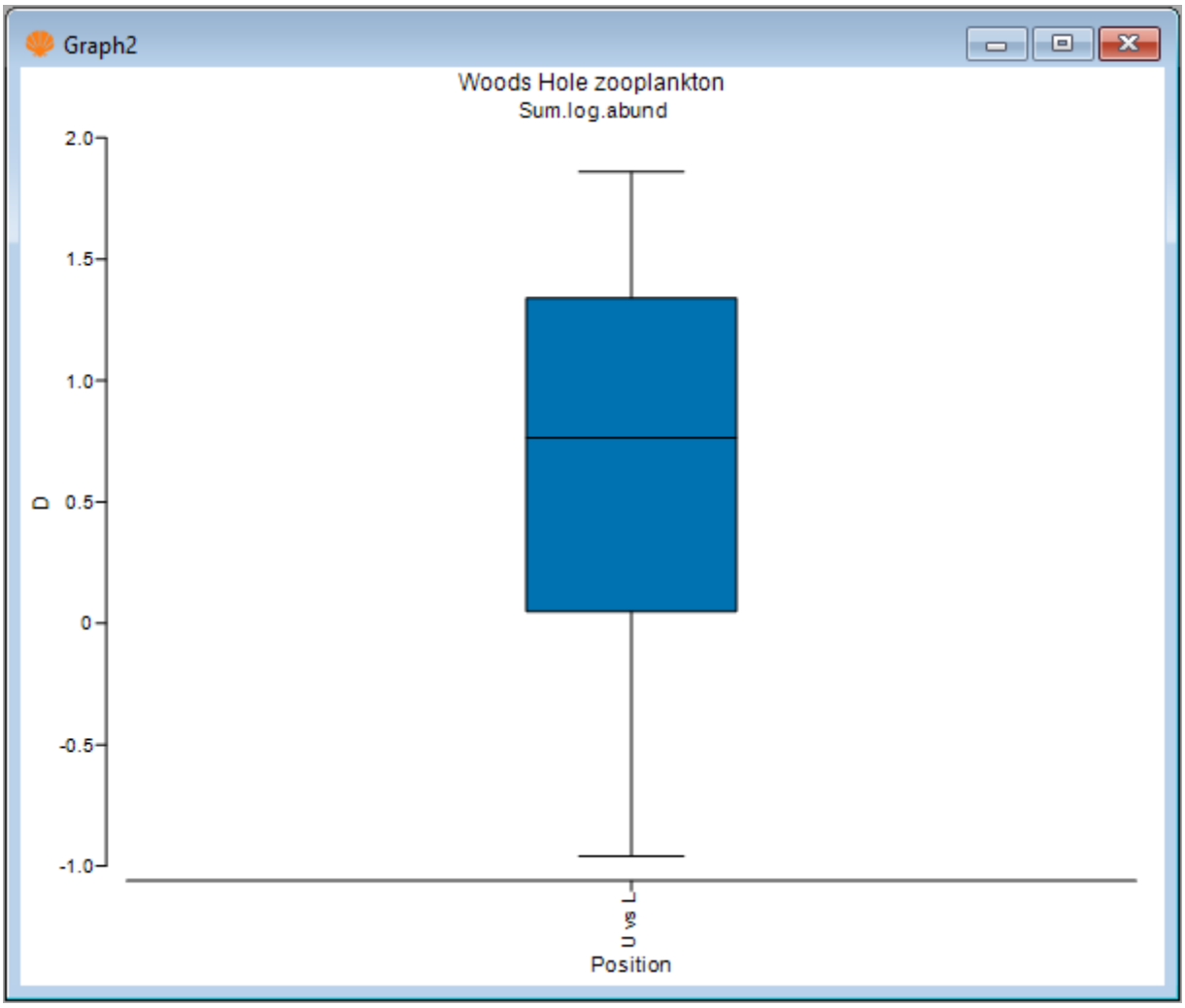
The two graphical outputs from this analysis are:

- 'Graph1 - a histogram of the distribution of values of W^π obtained under permutation, showing also the observed value W_obs (dotted lines):



and

- 'Graph2 - a boxplot of the differences (upper minus lower) in the sum of the log abundance values of these five taxa per haul



The boxplot shows that the median (and indeed the entire inter-quartile range) of the differences in total log abundance is clearly well above zero, consistent with the positive value of W_obs.

Note that this is a two-tailed test. If we wish to postulate a more specific alternative hypothesis, corresponding with the expectation that there will typically be a *greater* sum of log abundance values of plankton captured in the upper net, we can re-run the analysis precisely as above but with the option:

Alternative hypothesis: Paired differences (level1 - level2) are symmetric around D
    $\bullet$D > 0 (one-tailed test)

That will produce a one-tailed test, which will have greater power. The one-tailed test yields the same value of the Wilcoxon test-statistic as the two-tailed test (W = 39 in this example), but naturally yields a *p*-value that is half the size of the corresponding two-tailed test (i.e., P = 0.0244 for the one-tailed test in this example).

4.3 Mann-Whitney U test

Overview

The Mann-Whitney U test was described by Wilcoxon (1945) and Mann & Whitney (1947) . Here, interest lies in comparing two groups of independent samples. This is a non-parametric analogue to a classical two-sample (unpaired) *t*-test.

The null hypothesis

Suppose we have independent response values for a quantity of interest measured from each of two groups. We consider these values in the two groups to be representative observations from each of two random variables, Y_1 and Y_2, respectively. The general null hypothesis tested by the Mann-Whitney U test is:

- H_{01}: the distribution of Y_1 is equivalent to that of Y_2.

Thus, Y_1 and Y_2 are *exchangeable* with one another if H_{01} is true. If we want to assume that the distributions of Y_1 and Y_2 have the same scale, shape, etc., and can only differ in the value of their central location, then we can have a more specific null hypothesis:

- H_{02}: the distribution of Y_1 is *stochastically no larger or smaller than* that of $Y_2;$ or (even more strictly)
- H_{03}: the median of Y_1 = the median of Y_2.

The corresponding (two-sided) alternative hypotheses, in each case, would be phrased as:

- H_{A1}: the distributions of Y_1 and Y_2 differ from one another.
- H_{A2}: the distribution of Y_1 is *stochastically either larger or smaller than* that of $Y_2;$ or (even more strictly)
- H_{A3}: the median of Y_1 $\neq$ the median of Y_2,

respectively.

One may also choose to do a **one-tailed test**, e.g., to assert a more specific alternative hypothesis; namely that Y_1 is stochastically larger than Y_2 (or, the median of Y_1 > the median of Y_2).

Description of the test statistic

Let $\left[y_{1},\ldots,y_{n_1} \right] $ be an independent and identically distributed (i.i.d.) sample of n_1 units from Y_1 ('group 1'), and let $\left[y_{(n_1+1)},\ldots,y_{(n_1+n_2)} \right] $ be an i.i.d. sample of n_2 units from Y_2 ('group 2'). The combined set of sampling units from both groups is therefore given by vector $\mathbf{y} = \left[

y_{1},\ldots,y_{(n_1+n_2)} \right]. $ Let vector $\mathbf{r} = \left[r_1,\ldots,r_{n_1+n_2} \right]$ be the corresponding ranks of all the sample values in $\mathbf{y}$ from both groups combined, such that the smallest value obtains rank = 1 and the largest value obtains rank = (n_1+n_2). Next, define R_1 and R_2 as the sum of the ranks for group 1 and group 2, respectively; i.e.

$$ R_1 = \sum_{i=1}^{n_1} r_i \hspace{1cm} \text{and} \hspace{1cm} R_2 = \sum_{i=(n_1+1)}^{(n_1+n_2)} r_i $$

then calculate U_1 and U_2 as follows:

$$ U_1 = n_1n_2+\frac{n_1(n_1+1)}{2} - R_1 \hspace{1cm} \text{and} \hspace{1cm} U_2 = n_1n_2+\frac{n_2(n_2+1)}{2} - R_2 $$

Now, for any given dataset, the calculated values of U_1 and U_2 are not independent of one another. More specifically, $U_1+U_2 = n_1n_2$, so either U_1 or U_2 can be used as a suitable test statistic here.

PRIMER uses U_2 as the test-statistic, but will calculate and provide values for both U_1 and U_2 in the output file for this test.

Treatment of ties

If any values of y_i are equal to one another, then they are tied with one another in terms of rank. Then, if we order the data from smallest to largest and generate corresponding integers in order from 1 to (n_1+n_2) , PRIMER will simply *average* the corresponding ordered integers for any tied values to obtain an average and equivalent rank value for them. Thus, for example, suppose we have two groups with $n_1 = n_2 = 3$ having the following values: $\mathbf{y} = \{ 2, 4, 5, 5, 10, 15 \}$, then the ordered integers are $\{ 1, 2, 3, 4, 5, 6 \}$ and the ranks would be $\mathbf{r} = \{ 1, 2, 3.5, 3.5, 5, 6 \}$, because the 3rd and 4th-ranked values are tied.

Importantly, the presence of ties poses no problem for calculating *p*-values in PRIMER, as the permutation algorithm simply proceeds in the usual way, and an exact test is achieved directly. This contrasts with other available software implementations of the Mann-Whitney U test (e.g., 'wilcox.test' in R), which do not compute an exact *p*-value if there are ties.

Calculation of the *P*-value

Assuming only *exchangeability* of the observations across the two groups, we shall generate the distribution of U_2 under a true null hypothesis by permuting the observations freely across the two groups (always retaining the original sample sizes of n_1 and n_2), which permits a direct empirical calculation of an appropriate *p*-value.

If there are no tied values, then the distributions of U_1 and U_2 under a true null hypothesis are: (i) symmetric; and (ii) identical to one another. However, if there are any *ties*, then these two permutation distributions are neither symmetric nor identical. They will, nevertheless, be *mirror images* of one another. Specifically, the right-hand tail of U_2 will mirror the left-hand tail of

U_1, and the right-hand tail of U_1 will mirror the left-hand tail of U_2. Because of this mirroring, even in the presence of ties, we still only need one or other of U_1 or U_2 to calculate a correct *p*-value under permutation to achieve an exact test.

One-tailed test

First, we shall consider a one-tailed test. Suppose we have the alternative hypothesis:

- H_A: the distribution of Y_1 is stochastically *larger* than that of $Y_2.$

In this case, we would expect that the observed ranks, r_i, associated with group 1 would typically be larger numbers than those associated with group 2, and therefore (given the above equations), that the value of U_2 would tend to be greater than that of U_1.

Having obtained an observed value of the test-statistic, U_2, for a given dataset, we can then obtain a *p*-value empirically using a permutation algorithm. Specifically, under the simple null hypothesis that the two groups are exchangeable, we can randomly re-order (shuffle) all of the values of y_i in the combined vector $\mathbf{y}$ to yield a vector of permuted data, $\mathbf{y}^\pi$ of same length as the original, (n_1+n_2). The concomitantly re-ordered ranks associated with this permuted vector, $\mathbf{r}^\pi$, are then used to calculate a value of the test-statistic under permutation U_2^π. Repeating this permutation procedure a large number of times (e.g., say n_{perm} = 9999) yields a distribution of values of $U_{2,k}^\pi$ $k = 1,\ldots,n_{\text{perm}}$ under a true null hypothesis.

The *p*-value is then calculated as:

$$ P = \frac{ \sum_{k=1}^{n_\text{perm}} (\text{I}(U_{2,k}^\pi \geq U_2) + 1)}{(n_\text{perm} + 1)} $$ with $\text{I}(\textit{expression}) = 1$ if $\textit{expression}$ is true and zero otherwise. The '$+1$' in the numerator and denominator of this fraction is there to acknowledge the inclusion of the observed value as a member of the permutation distribution. The above expression tallies only the values of $U_{2,k}^\pi$ that equal or positively exceed U_2 in the right-hand tail of the permutation distribution.

Note that the arbitrary ordering of the names of the two groups being compared (i.e., as 'group 1' and 'group 2') can simply be swapped if we wish to perform the test with the alternative hypothesis that Y_1 is stochastically *smaller* than Y_2 (focused on the other tail).

Two-tailed test

For an exact two-tailed test that accommodates ties (hence asymmetry in the permutation distributions), we could examine the permutation distributions for both U_1 and U_2. Specifically, for example, if $U_1 < U_2$, we can calculate the two-tailed *p*-value as the sum of the two relevant individual tail probabilities (i.e., the lower tail of U_1^π and the upper-tail of U_2^π), thusly: $$ P = \text{Pr} (U_1^\pi \leq U_1) + \text{Pr} (U_2^\pi \geq U_2) $$

However, taking advantage of the fact that U_1 and U_2 are not independent of one another, and that the permutation distributions of U_1 and U_2 will be precise mirror images of one another (even if some values are tied and they are not therefore symmetric), we can always

calculate the correct two-tailed *p*-value using only the permutation distribution of U_2 as follows:

- If $U_1 < U_2$, then $$ P = 2 \times \frac{ \sum_{k=1}^{n_\text{perm}} (\text{I}(U_{2,k}^\pi \geq U_2) + 1)}{(n_\text{perm} + 1)} $$
- If $U_2 < U_1$, then $$ P = 2 \times \frac{ \sum_{k=1}^{n_\text{perm}} (\text{I}(U_{2,k}^\pi \leq U_2) + 1)}{(n_\text{perm} + 1)} $$

4.4 Example: Snapper in marine reserves

As an example of the Mann-Whitney U test, we will look at a dataset consisting of counts of the snapper (*Chrysophrys auratus*) sampled using baited remote underwater videos (BRUVs) from multiple areas inside *vs* outside several marine reserves along the north-eastern coast of New Zealand (Smith *et al.* (2014)). These data are located in the file 'NZ_Snapper_counts.pri', found inside the 'Examples_P8 > NZ_snapper_counts' folder and include the following three variables:

- *tot.snapper*: Total count of snapper regardless of size (MaxN)
- *small.snapper*: Count of small snapper $\lt$ 27 cm (MaxN)
- *large.snapper*: Count of snapper $\ge$ 27 cm (MaxN)

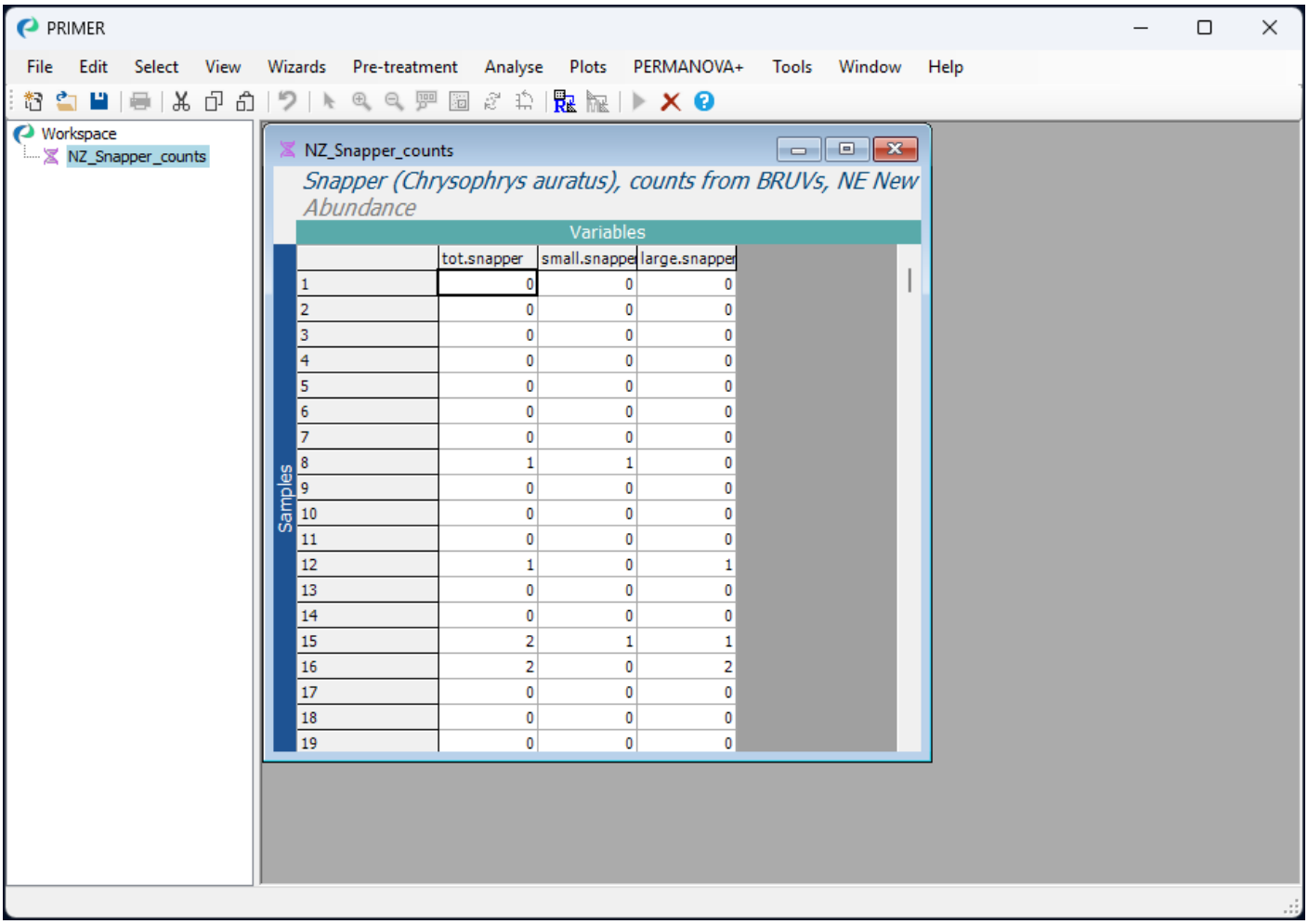
The counts were obtained from video footage as 'MaxN', which is the maximum number of individuals observed in a single frame (during a 60-min period of video, beginning 5 min after the gear makes contact with the sea floor). It is used as an index of relative density. Note also that the minimum legal size for snapper (i.e., the minimum size of fish which may be taken legally by fishers) at the time these data were collected was 27 cm.

There are a number of factors associated with these data. (Click **Edit** > **Factors** to see them). Each sampling unit (MaxN value from a given piece of 60-min BRUV footage) is identified by the factor of 'Status' as having been taken from either inside the marine reserve (reserve = 'R') or outside the marine reserve (non-reserve = 'NR'). These are the two groups we wish to compare using the Mann-Whitney U test. Here, we shall focus on comparing reserve *vs* non-reserve counts of all snapper (*tot.snapper*) that were obtained only in 2003 and only from the locations of Leigh and Hahei.

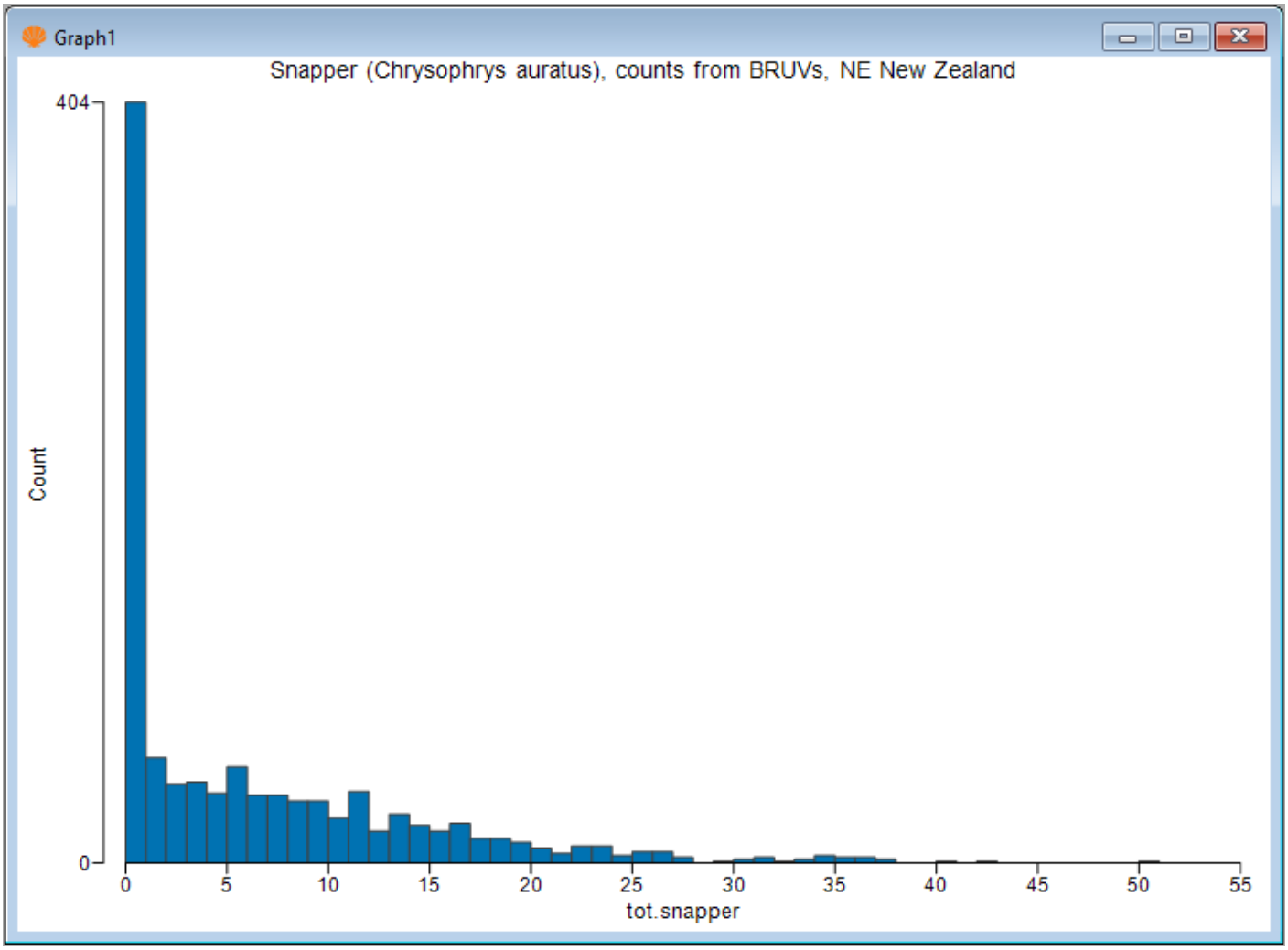
Note that, in PRIMER, it is possible to run the Mann-Whitney U test (or any of the univariate non-parametric tests) separately within levels of another factor. In this example, we shall run the test to compare the 2 levels of 'Status' (R *vs* NR) separately for each 'Location' (Leigh and Hahei). Furthermore, for this example, we shall also (quite naturally) turn to a *one-tailed* Mann-Whitney U test, as we would expect, *a priori*, that there would be greater MaxN values recorded from BRUVs deployed inside *vs* outside any particular reserve at any particular time.

Running the Mann-Whitney U test

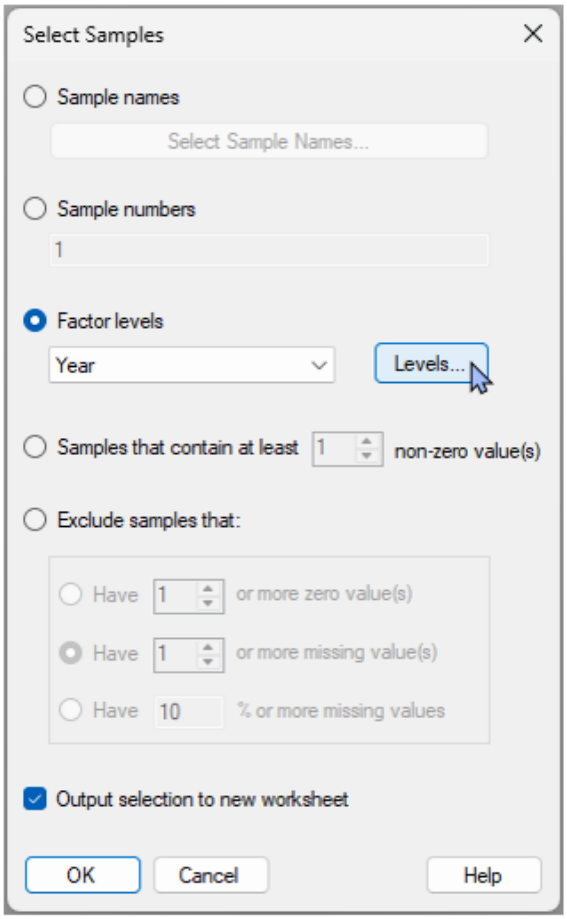
1. Open the file 'NZ_Snapper_counts.pri' in PRIMER.

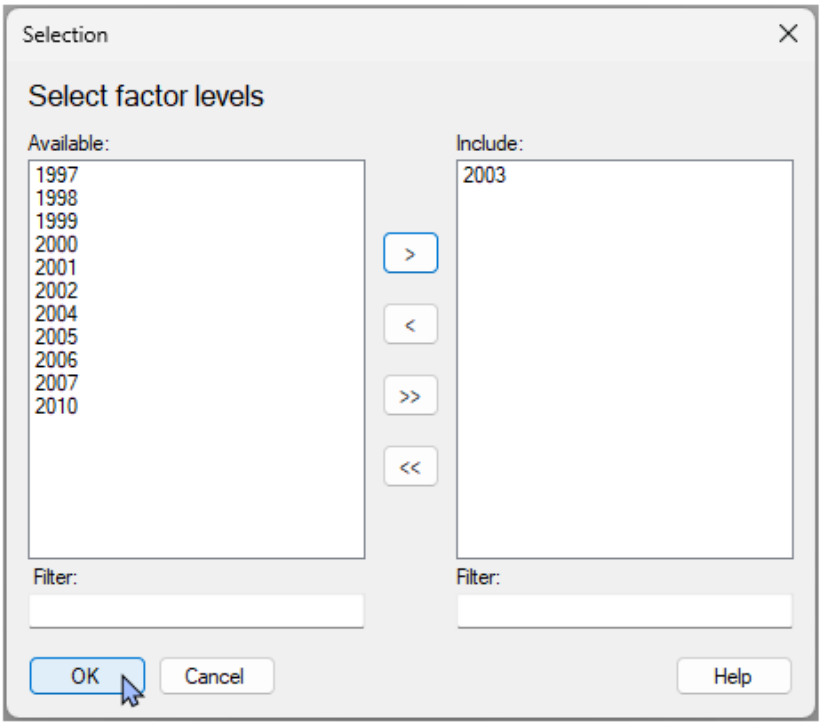


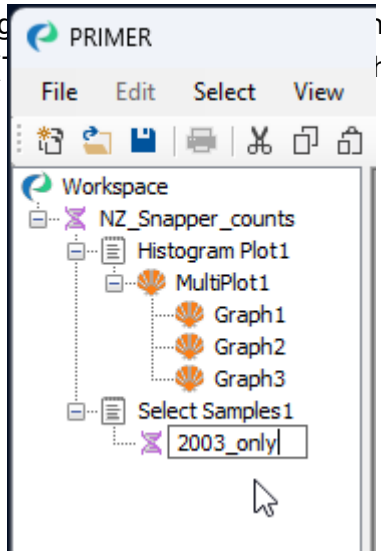
2. We may begin by making some simple histogram plots of these count variables. From the 'NZ_Snapper_counts' data sheet inside PRIMER, click **Plots** > **Histogram Plot...**. Clearly, the distributions of counts have a very large preponderance of zeros, so the distributions of errors from a classical ANOVA model would not be at all 'normal'. For example, consider the histogram for *tot.snapper* (shown below):



3. Select the 2003 data only. From the data sheet, click **Select** > **Samples...** > ($\bullet$ Factor levels: Year > **Levels** > Include: 2003, 'OK') & ($\checkmark$Output selection to new worksheet).



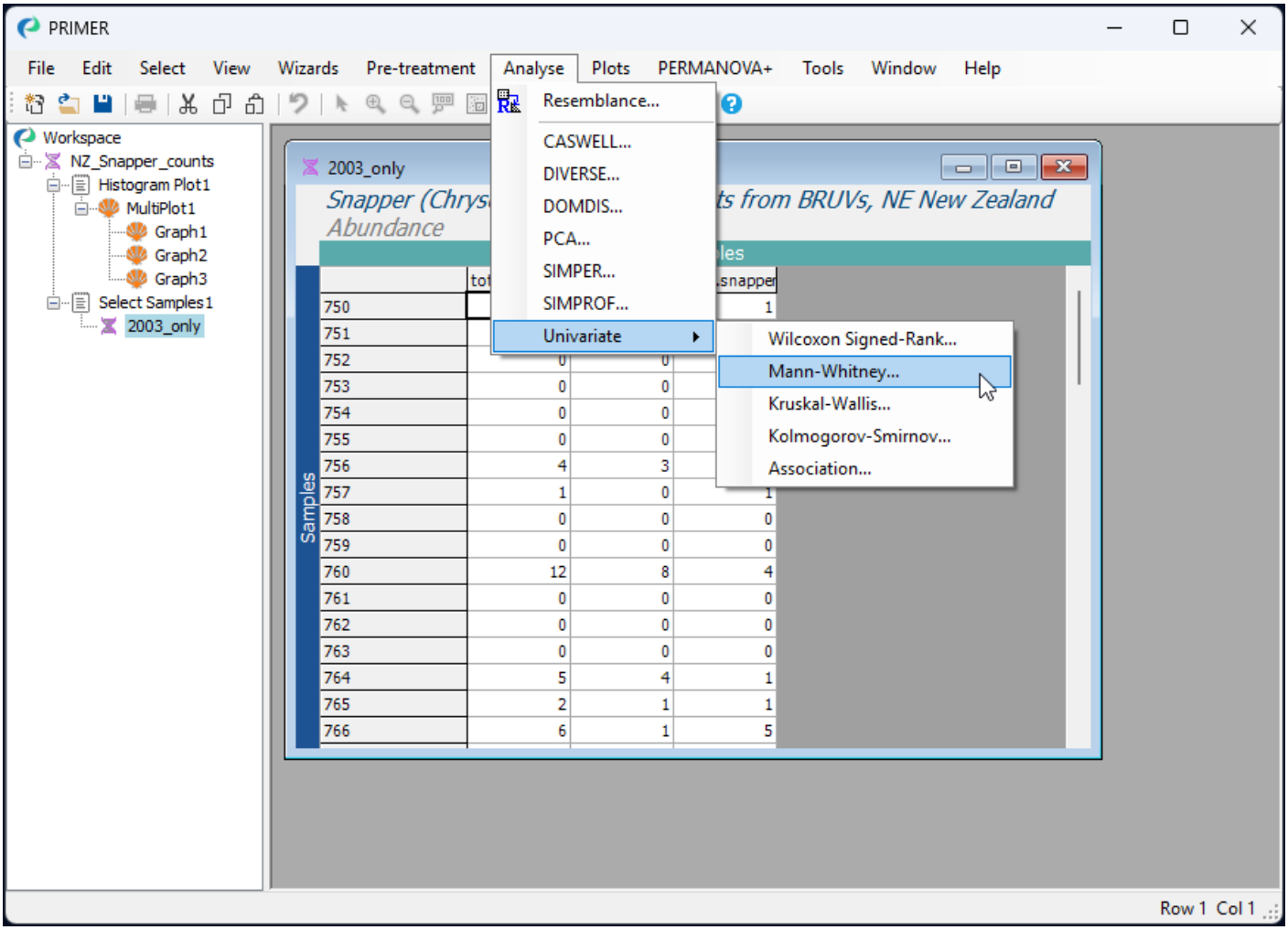


4. To keep thing 

ne resulting sheet (called 'Data1' by default) to '2003_only'. (T he Explorer tree window area, for example):

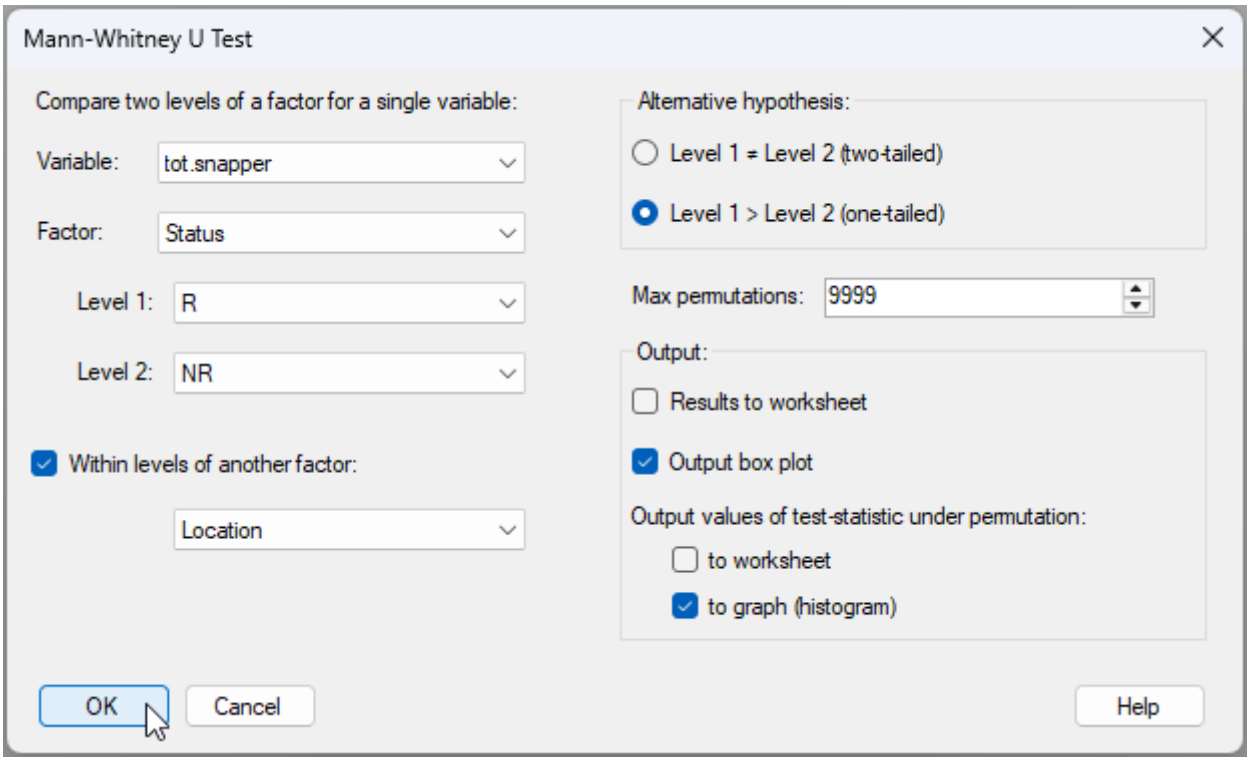
$\hspace{2cm}$

5. Let's now do the test. From '2003_only', click **Analyse** > **Univariate** > **Mann-Whitney…**



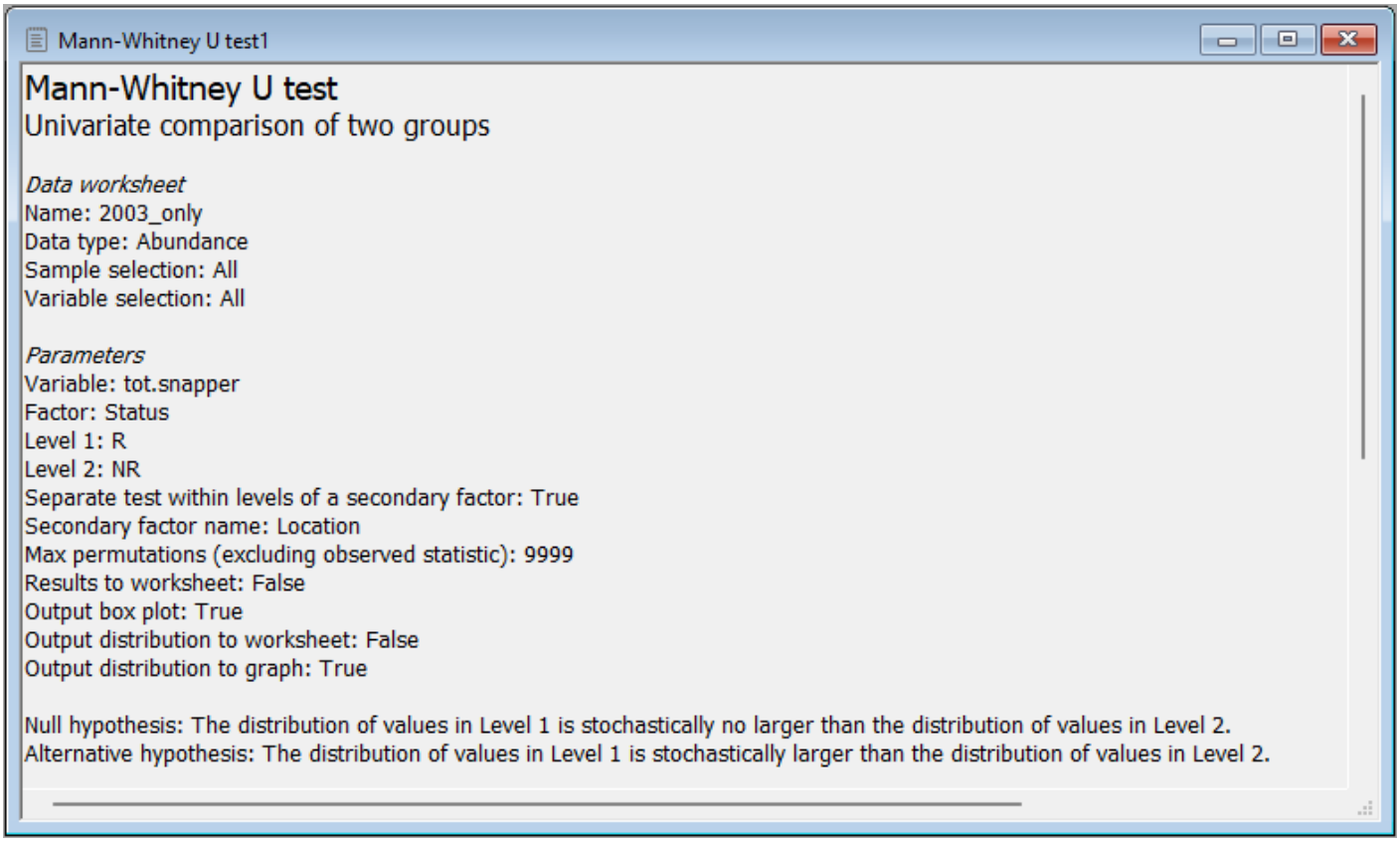
6. In the resulting Mann-Whitney U Test dialog, choose the following:
   - Variable: tot.snapper
   - Factor: Status
     - Level 1: R
     - Level 2: NR
   - $\checkmark$Within levels of another factor: Location
   - Alternative hypothesis: $\bullet$Level 1 $\gt$ Level 2 (one-tailed)
   - Max permutations: 9999
   - $\checkmark$Output box plot
   - Output values of the test-statistic under permutation: $\checkmark$to graph (histogram)

then click '**OK**', as shown below.



Results of the Mann-Whitney U test

The resulting notepad , *.rtf) file named 'Mann-Whitney U test1' contains all of the essential elements of this analysis and its results. First are shown the choices that were made by the user ('*Parameters*'):



This is followed, in the same output file, by the full suite of results, including all of the relevant supporting calculations ('*Results*'), *viz.*:

Mann-Whitney U test1

Null hypothesis: The distribution of values in Level 1 is stochastically no larger than the distribution of values in Level 2.
Alternative hypothesis: The distribution of values in Level 1 is stochastically larger than the distribution of values in Level 2.

Results

Location	Statistic (U1)	Statistic (U2)	P value	Possible Permutations	Actual Permutations	Unique Values of U2.perm	Num >= Observed	Num Tied Data Values	Total N
Hahei	49.5	175.5	0.0029	Large*	9999	285	28	26	30
Leigh	29.5	546.5	0.0001	Large*	9999	552	0	38	48

**greater than 100,000,000*

The test statistic used here under permutation is U2. Note that some researchers use U = Min(U1, U2) as the test statistic.

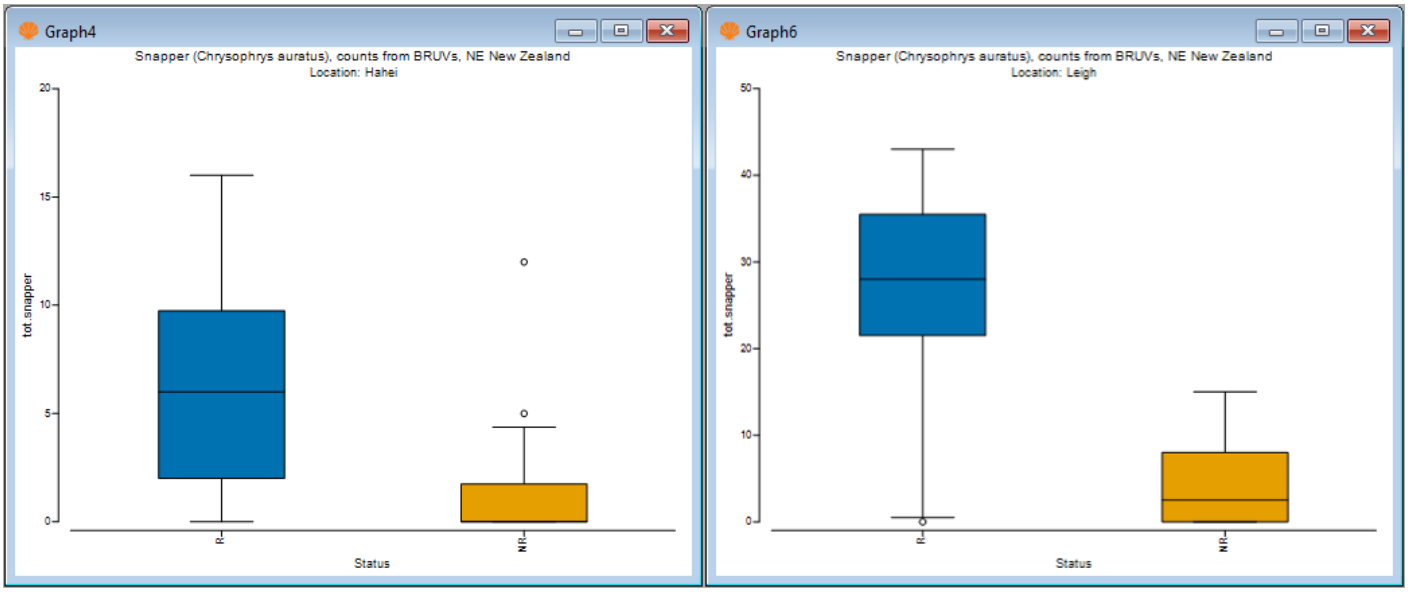
Medians of Status

Location	R	NR
Hahei	6	0
Leigh	28	2.5

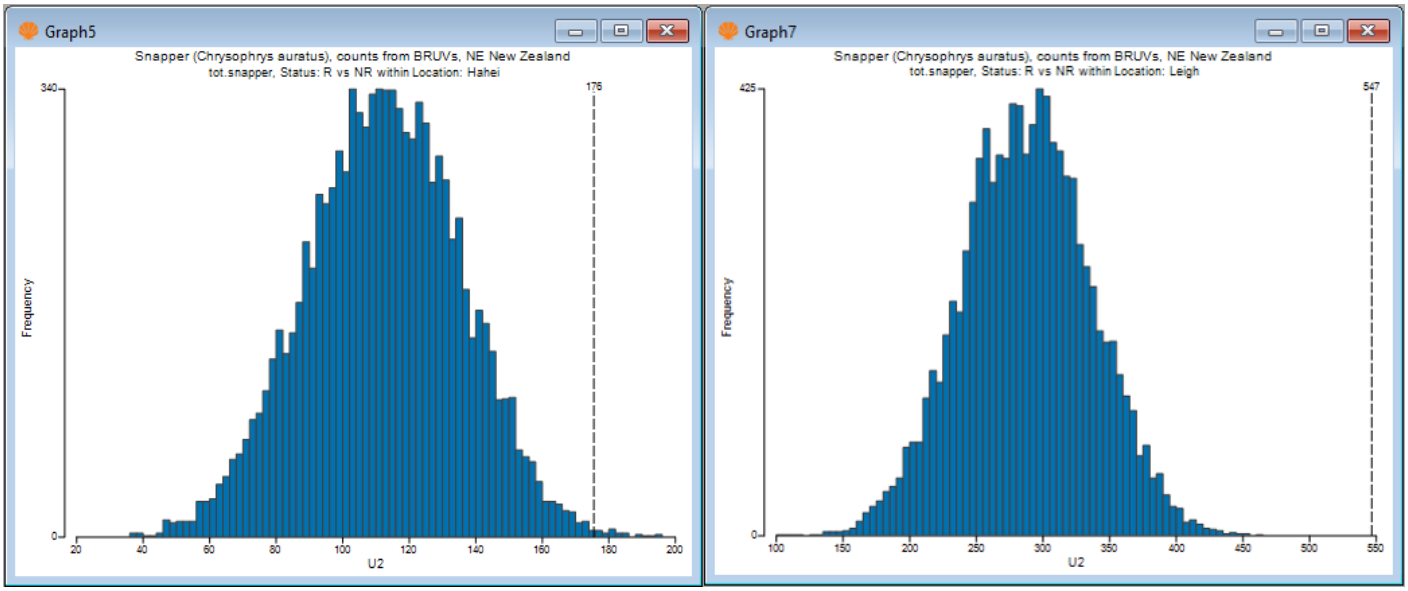
Outputs
Plot: Graph4
Plot: Graph5
Plot: Graph6
Plot: Graph7

We can reject the null hypothesis that there are no differences in the total count of snapper inside *vs* outside the marine reserve at Hahei (U = 175.5, P = 0.0029) and also, even more resoundingly, at Leigh (U = 546.5, P = 0.0001).

Looking next at the graphical output (shown under 'MultiPlot2'), the individual boxplots ('Graph4' and 'Graph6') show these effects and the direction of the differences very clearly: specifically, there was a greater median count of snapper recorded inside the reserve ('R') compared to outside the reserve ('NR') at each of these two locations, *viz*:



The graphical output also shows histograms of the empirical distributions of the U^{π} values (i.e., the values of the test-statistic under permutation) for each of Hahei and Leigh ('Graph5' and 'Graph7', shown below). For this example, we asked for one-tailed tests, so in these distributions, we can see the observed value of U as a dotted vertical line in the upper tail only.



Thus, these marine reserves clearly affected the total counts of snapper obtained from the BRUVs in 2003. Similar comparative tests may be done for different years and/or reserves.

For this example, we might also be interested to compare counts of just the large-sized snapper or just the small snapper, as perhaps the marine reserves only affect distributions of counts for fish that are large enough to be caught (legally) by fishers. We shall leave these ideas for end-users to explore on their own.

4.5 Kruskal-Wallis test

Overview

The Kruskal-Wallis test was described by Kruskal (1952) and Kruskal & Wallis (1952) . Its purpose is to compare two or more independent groups of samples and it is an extenstion of the Mann-Whitney U test. It operates on ranked values and will indeed yield an equivalent result to the Mann-Whitney U test in the case of two groups. It is a nonparametric statistical test whose classical counterpart is a one-way analysis of variance (ANOVA).

The null hypothesis

Like the Mann-Whitney U test, the Kruskal-Wallis test operates on ranks. Suppose we have a factor 'A' with $i=1,...,a$ distinct levels (groups or populations), and that there are n_i values of a given random response variable Y, sampled from each of these groups; so, y_{ij} indicates the jth sample value ($j = 1,...,n_i$) drawn from the ith group, and there are a total of $N = \sum_{i=1}^{a}n_i$ values being evaluated in the test. The general null hypothesis being tested here is:

- H_0: *There are no differences* in the distribution of values in the underlying populations represented by the groups.

The alternative hypothesis is:

- H_A: *At least two of the groups differ from one another* in the distribution of values in the underlying populations represented by the groups.

The Kruskal-Wallis test generally assumes that: (i) the sampled values are independent of one another, (ii) the sampled values in each group are drawn at random from each of their respective populations, and (iii) the response variable is continuous. We avoid any assumption that the underlying population values for each group are normally distributed. The null hypothesis just asserts that the values from different groups come from the same underlying population distribution, with identical shape and scale, whatever that distribution may be.

In practice, the test also can be applied to random variables that are discrete (so not necessarily continuous) and the ranks of any tied values can be replaced with their average rank value. Our use of a permutation algorithm to perform the test means an exact test (where the type I error of the test is equal to the *a priori* chosen significance level) is achieved. If we restrict the alternative hypothesis to a shift in location only, we may assert the null and alternative hypotheses for the Kruskal-Wallis test as follows:

- H_0: *There are no differences in the median values* of the underlying populations represented by the groups.

- H_A: *At least two groups differ in the median values* of the underlying populations represented by the groups.

Description of the test-statistic

The first step is to rank all of the values in the full set of data, combined, regardless of their group membership. Let the combined set of sampling units from all groups be denoted by a vector $\mathbf{y} = \left[y_{11}, y_{12}, \ldots,y_{ij},\ldots,y_{a,n_a} \right]$, of length N. Then, let vector $\mathbf{r} = \left[r_{11}, r_{12}, \ldots,r_{ij},\ldots,r_{a,n_a} \right]$ be the corresponding ranks of all the sample values in $\mathbf{y},$ such that the smallest value obtains rank = 1 and the largest value obtains rank = N.

Next, let R_i be the sum of the ranks for any group i, i.e.

$$ R_i = \sum_{j=1}^{n_i} r_{ij} $$

The Kruskal-Wallis test statistic is then defined as:

$$ H = \frac { 12 } { N(N+1) } \sum_{i=1}^a \frac { R_i^2 } { n_i } - 3(N+1) $$

Treatment of ties

The distribution of H under a true null hypothesis is affected by the existence of ties. Thus, in the event of ties, average ranks are calculated, and the H test-statistic is adjusted as described by Kruskal & Wallis (1952) and outlined below.

Calculating average ranks

If any values of y_{ij} are equal, then they are tied with one another in terms of their rank. For any tied values, PRIMER calculates an average rank for them, just as described for the Mann-Whitney U test. So, for example, if we have the following set of values: $$ \mathbf{y} = \{9, 12, 12, 14, 14, 14, 30\}, $$

the set of ordered integers for these 7 sample values would look like this: $$ \{1, 2, 3, 4, 5, 6, 7\} $$

and the rank values $\mathbf{r}$ that we would actually use for the analysis (obtained by replacing the ordered integers with their averages, calculated separately for each set of tied values) would look like this: $$ \mathbf{r} = \{1, 2.5, 2.5, 5, 5, 5, 7\} $$

As previously noted, the presence of ties poses no problem for calculating *p*-values in PRIMER, as the permutation algorithm simply proceeds in the usual way, and an exact test is achieved directly. Note that, for cases where there is a small number of unique values of the test statistic under permutation, although the *p*-value is *accurate* (there is no bias), its *precision* and hence its utility will depend on the number of unique values of the test-statistic that can be computed for a given

problem.

Adjustment to the test-statistic

In the event of tied values, let the number of sets of tied values be g. Within each set $\ell = 1, \ldots, g$, there are t_ℓ tied values. For every set ℓ, we also calculate $T_\ell = (t_\ell - 1)t_\ell(t_\ell + 1)$. Then the H test-statistic is adjusted by dividing it by quantity D:

$$ H_{\text{adj}} = H/D $$

where D is defined as $$ D = 1 - \frac{ \sum_\ell^g T_\ell }{ N^3 - N} $$

We note in passing that this adjustment is not necessary in our case, as the *p*-value is calculated using a permutation approach (see the following section), rather than relying on tabled values. However, PRIMER does calculate H_{adj} in the event of ties, essentially to maintain consistency with results that would be obtained using other packages and to remain true to the description of the test-statistic as described by Kruskal & Wallis (1952) .

Calculating a *p*-value

Having obtained an observed value of the test-statistic, H_{obs}, for a given dataset, we can then obtain a *p*-value empirically using a permutation algorithm. Specifically, under the null hypothesis that all groups are exchangeable, we can randomly re-order (shuffle) all of the values of y_{ij} in the combined vector $\mathbf{y}$ to yield a vector of permuted data, $\mathbf{y}^\pi$ of same length (N) as the original. The concomitantly re-ordered ranks associated with this permuted vector, $\mathbf{r}^\pi$, are then used to calculate a value of the test-statistic under permutation H^π. Repeating this permutation procedure a large number of times (e.g., say n_{perm} = 9999) yields a large number of values of H^π realised under a true null hypothesis. The probability (*p*-value) associated with the null hypothesis is then estimated empirically as the proportion of values of H^π that are equal to or larger than the observed value of the test-statistic, H_obs. Specifically, if we let H_k^π be the value of H^π obtained for the kth permutation ($k = 1, \ldots, n_\text{perm}$), the *p*-value is calculated as the proportion of H_k^π values that equal or exceed H_obs; i.e., $$ p = \frac{ \sum_{k=1}^{n_\text{perm}} (\text{I}(H_k^\pi \geq H_\text{obs}) + 1)}{(n_\text{perm} + 1)} $$

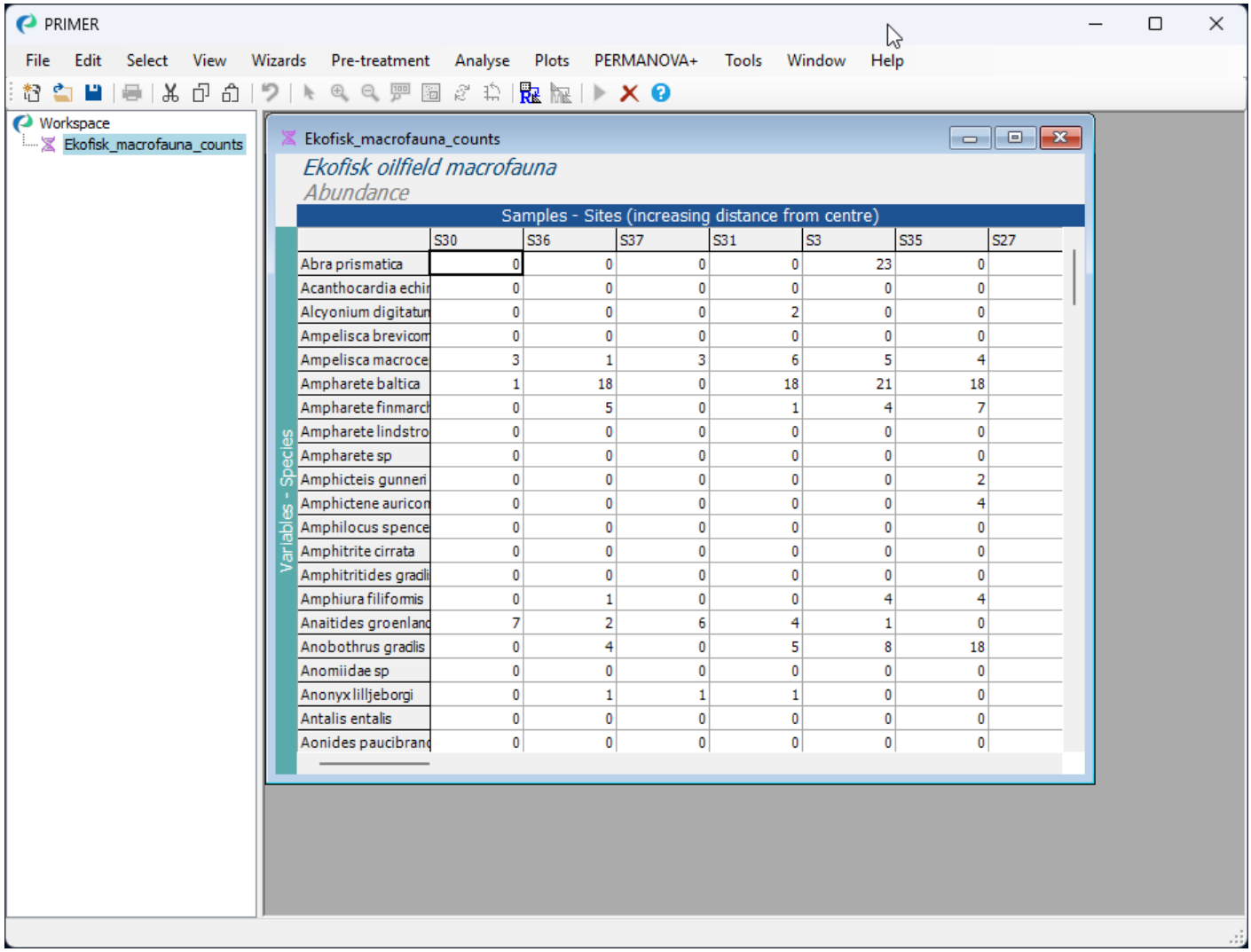
with $\text{I}(\textit{expression}) = 1$ if $\textit{expression}$ is true and zero otherwise. Note that the '$+1$' in the numerator and denominator of this fraction is there to acknowledge the inclusion of the observed value as a member of the distribution of H^π under a true null hypothesis.

4.6 Example: A bivalve species from Ekofisk

We will use the Kruskal-Wallis test to compare counts of a bivalve species, *Abra prismatica*, occurring at sites classified into groups according to their proximity to the Ekofisk oilfield in the North Sea (Gray *et al.* (1990)). Macrofauna were sampled from each of 29 sites that were laid out roughly along five transects that radiated out from the centre of the oil platform. The sites were classified into 4 strata, based on their relative distance from the oil platform. These 4 strata (groups) of sites were defined and labeled as: A (> 3.5 km), B (1 km - 3.5 km), C (250 m - 1 km) and D (< 250 m). There were three day-grab sampling units taken from each site; data from these three units were combined to yield abundance values for each of p = 173 soft-sediment macrofaunal taxa at every site.

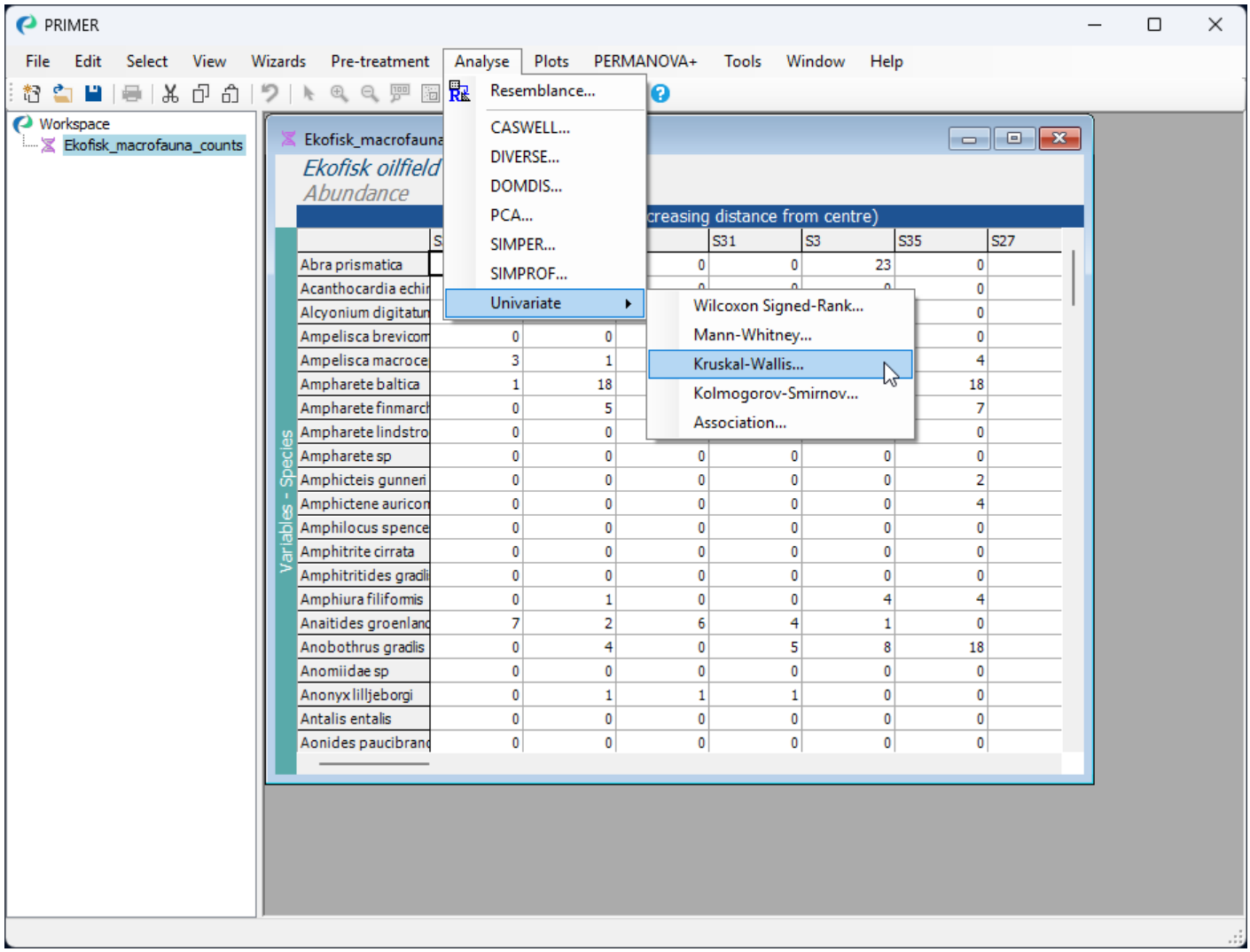
Running the Kruskal-Wallis test

1. Open up the example data file in PRIMER. These data are located in the file named ‘ Ekofisk_macrofauna_counts.pri’, found inside the 'Examples_P8 > Ekofisk_macrofauna' folder.

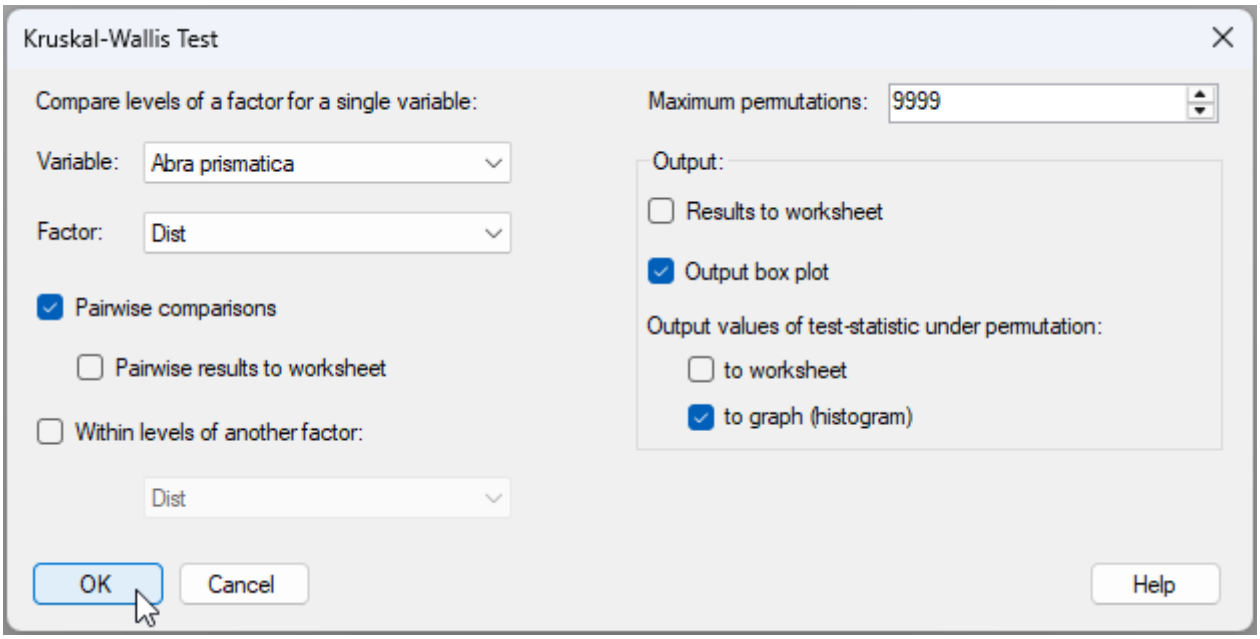


Note that there is no need to do any transformations on these data, as there would be no effect of a monotonic transformation on this univariate non-parametric test. This is simply because the ranks of values, upon which the test depends, are naturally not in any way changed by such a transformation.

2. Now run the analysis. From the 'Ekofisk macrofauna counts' data sheet, click **Analyse** > **Univariate** > **Kruskal-Wallis...**



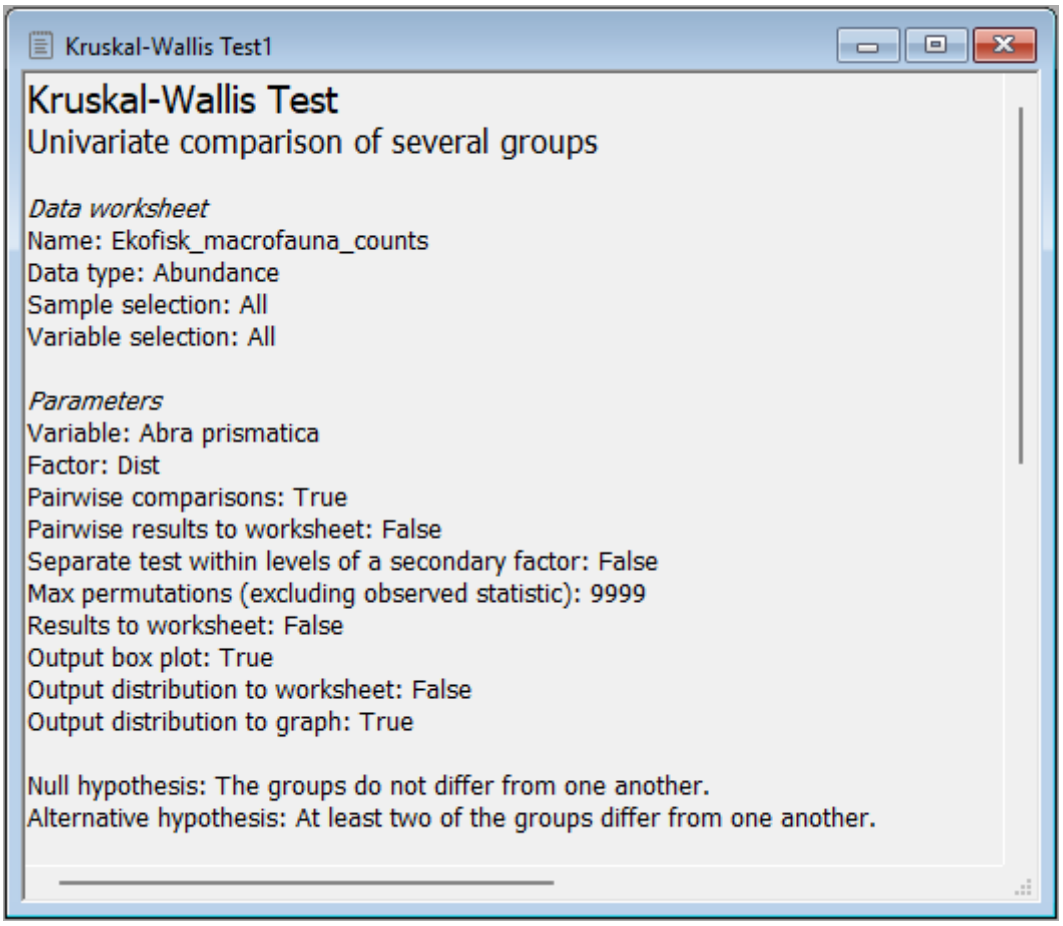
3. Choose the following options in the 'Kruskal-Wallis Test' dialog window:
   - Variable: Abra prismatica
   - Factor: Dist
   - $\checkmark$Pairwise comparisons
   - Max permutations: 9999
   - $\checkmark$Output box plot
   - Output values of test-statistic under permutation: $\checkmark$to graph, then click '**OK**'.



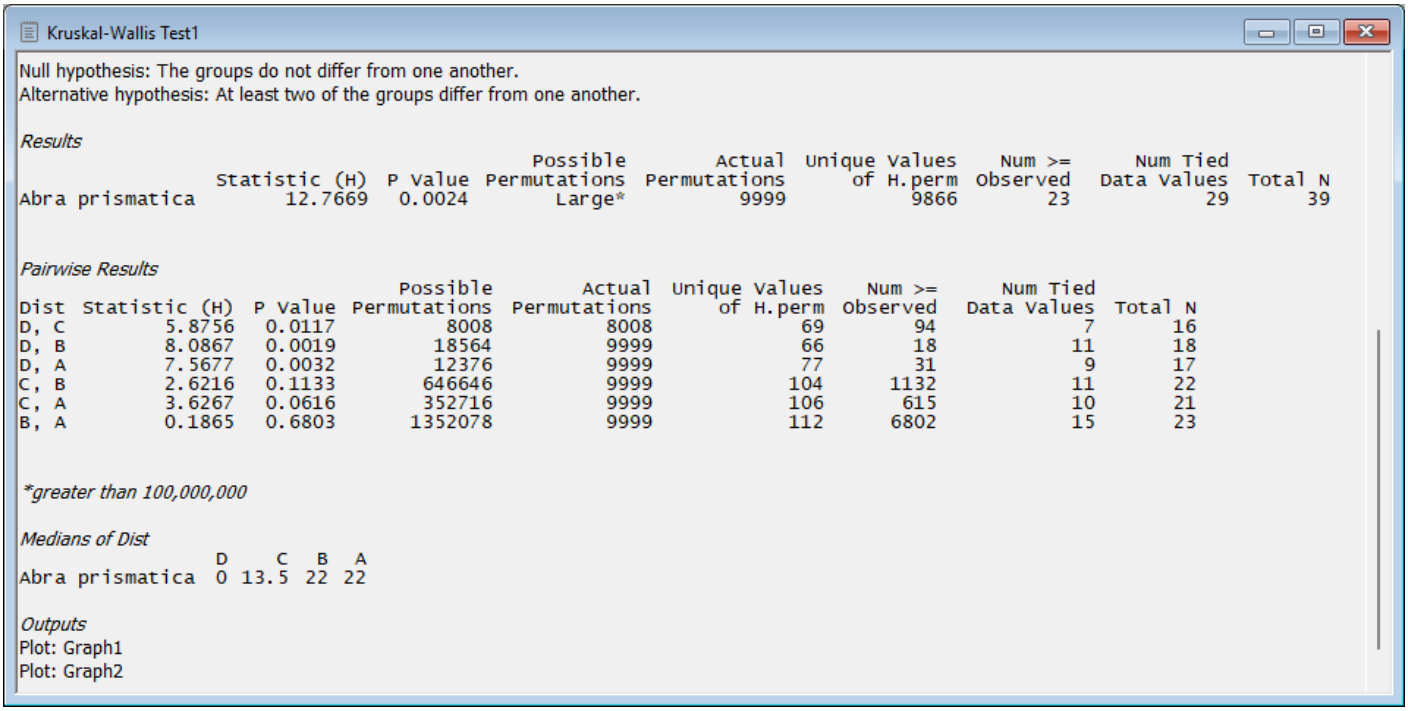
You may note in passing that it is also possible *via* this dialog to run the Kruskal-Wallis test separately within levels of another factor, in cases where there may be greater complexity in the design (e.g., crossed factors).

Results of the Kruskal-Wallis test

The resulting notepad (, *.rtf) file named 'Kruskal-Wallis Test1' contains all of the essential elements of this analysis and its results. First the choices that were made by the user ('*Parameters* ') are shown:



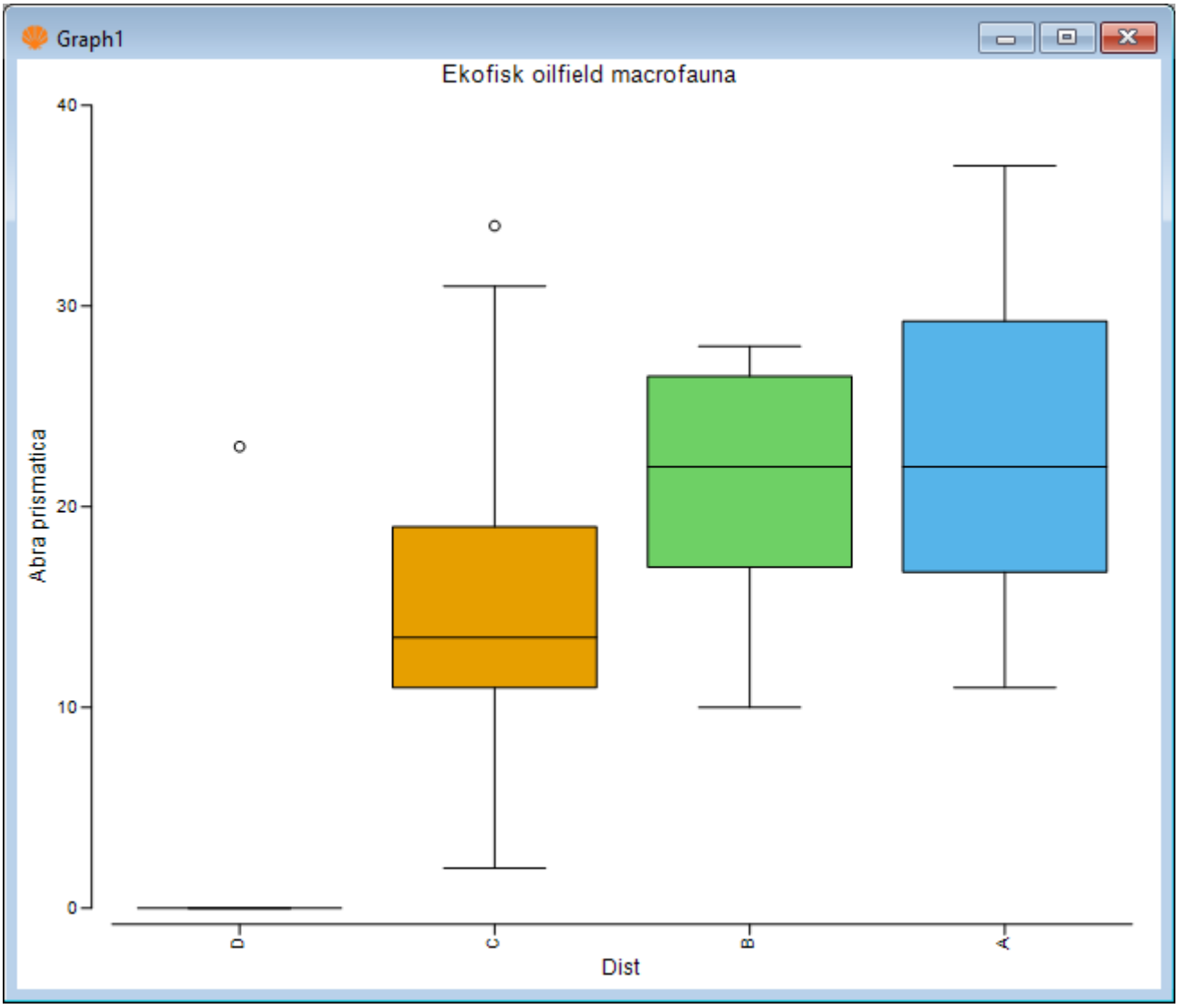
This information is followed by the full suite of results, including all of the relevant supporting calculations ('*Results*'), *viz.*:



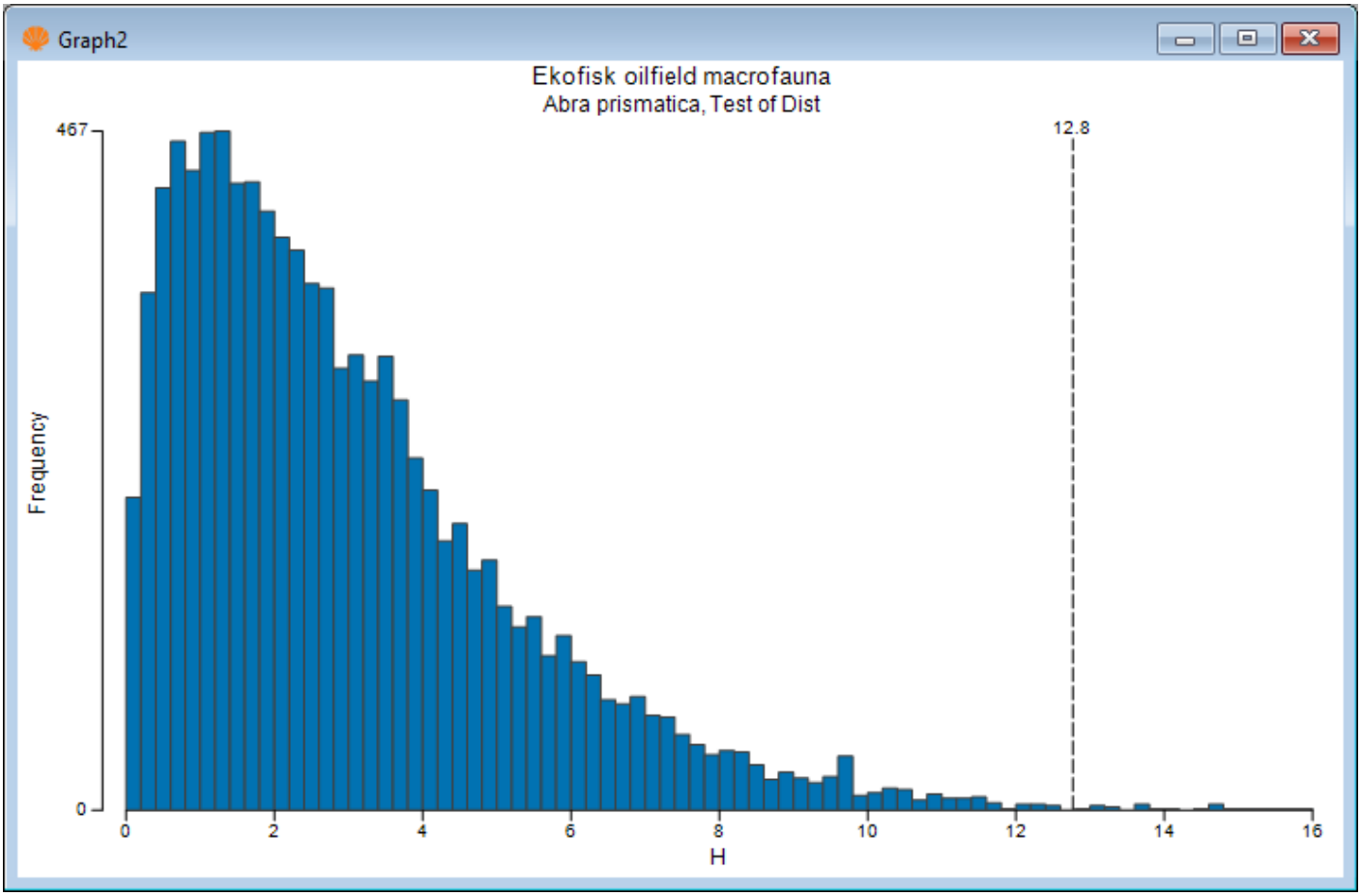
We can reject the null hypothesis that the four groups of sites are equal with respect to counts of *Abra prismatica* (H = 12.77, P = 0.0024) and conclude that some of these groups differ from one another. The pairwise comparisons show statistically significant differences between group D (the sites closest to the oil platform) and all other groups (P < 0.015 for all of those tests). Groups A and B do not differ significantly from one another (P > 0.68), nor do groups B and C

(P > 0.10), while the comparison between groups A and C approaches significance at the 0.05-level (P = 0.06). The median of *Abra prismatica* abundance values in each group are also given directly in this output file.

The results are clarified by the graphical outputs (shown under 'MultiPlot1'). The boxplot ('Graph1') shows how the distributions change across the sites, with greater median abundances per site being observed far away fom the oil platform (groups A and B); almost no individuals of this species were observed near the platform (group D).



We also see a histogram of the empirical distribution of the H^{π} values under permutation ('Graph2', shown below). Clearly the observed value (H_obs, the veritcal dotted line) is quite unusual (far out in the right-hand tail) by comparison with this empirical distribution for our test-statistic that was generated under a true null hypothesis.



Taken all together, these results indicate that the bivalve, *Abra prismatica*, is strongly and negatively affected by oil drilling at sites occurring within 250 m (in any direction) of the Ekofisk oil platform (group D), where their median abundance was effectively reduced to zero. Median abundance was also reduced (compared to background levels) at sites occurring between 250 m and 1 km away from the oil platform (group C). At distances greater than 1 km, however, no further effects of the Ekofisk oil platform on median abundances were detected (groups A and B).

One may wish to apply the test on some other individual species of interest from the Ekofisk dataset (e.g., *Montacuta substriata*, *Eteona longa*, *Goniada maculata*, etc.) or perhaps on some univariate metric of diversity or abundance (such as richness, even-ness or total abundance), derived from the multivariate data. Note that you can use PRIMER's **Analyse** > **DIVERSE...** routine to obtain a wide variety of diversity metrics from multivariate data for subsequent analysis.

4.7 Kolmogorov-Smirnov test

Overview

The Kolmogorov-Smirnov test is a non-parametric test for comparing two distributions of a continuous variable. Rejection of the null hypothesis indicates that the two distributions differ from one another in some way (location, dispersion, skewness, etc.). The evolution of the test can be traced in the work of Kolmogorov (1933), Kolmogorov (1941), Smirnov (1939a) and Smirnov (1939b). See also Darling (1957) for a detailed synopsis.

The null hypothesis

There are essentially two versions of the test in common usage: one is a goodness-of-fit test, designed to compare the distribution of a sampled random variable with some known distribution. The other is to compare the distributions of two sampled random variables ('the two-sample problem' *sensu* Darling (1957)). The Kolmogorov-Smirnov test in PRIMER implements this latter (two-sample) test.

- Let $X_1, X_2, \ldots, X_{n_1}$ be a set of n_1 observations of independent random variables ('sample 1' or 'group 1') that each have the same continuous distribution function, $U(x)= \text{Pr} \lbrace{ X_i < x \rbrace}$.
- Similarly, we let $Y_1, Y_2, \ldots, Y_{n_2}$ be a set of n_2 observations of independent random variables ('sample 2' or 'group 2') that each have the same continuous distribution function, $V(x)= \text{Pr} \lbrace{ Y_i < x \rbrace}$.

The null hypothesis for the Kolmogorov-Smirnov test is that these two groups of samples come from the same distribution, i.e. $$ \text{H}_ 0 \text{:} \hspace{0.2cm} U(x) = V(x) $$

Description of the test statistic

Let $\hat{F}_ {n_1}(x)$ be the empirical distribution function for group 1. Specifically, $\hat{F}_ {n_1}(x)$ is the proportion of the X_i, $i = 1, \ldots, n_1$, that are less than x. Thus, if X_i = 20, then $\hat{F}_ {n_1}(X_i)$ is the proportion of the n_1 values in group 1 that are less than 20. This will be equal to one minus the proportion of n_1 values in group 1 that are greater than or equal to 20. Similarly, we can let $\hat{G}_ {n_2}(x)$ be the empirical distribution function for group 2, defined in the same way but for that group.

Once these two empirical distributions have been calculated, the Kolmogorov-Smirnov test-statistic is defined as

$$ D = \sup_{-\infty < x < \infty} |\hat{F}_ {n_1}(x) - \hat{G}_ {n_2}(x)| $$

Thus, D is the supremum of the absolute values of differences calculated between the two empirical distribution functions. In essence, the test-statistic here captures the largest possible difference that is observable between the two empirical distributions for any value of x.

Calculating a *p*-value

There are tabled values for D that can be calculated under certain conditions, but in PRIMER we simply generate a distribution for D under the null hypothesis directly and empirically by assuming only *exchangeability* between the two groups. For the permutation test, the $(n_1 + n_2)$ values are permuted randomly across the two groups (preserving each of the individual group's original sample sizes, n_1 and n_2, respectively), and the test statistic is calculated for the permuted data as D^π. We can repeat this randomisation procedure a large number of times (e.g., n_perm = 9999), to obtain a permutation distribution of D^π under a true null hypothesis.

By comparing our observed value (obtained with the original ordering of the data, D_obs) with the distribution of D^π, we obtain a direct empirical estimate of the probability associated with the null hypothesis. Specifically, if we let D_k^π be the value of D^π obtained for the kth permutation ($k = 1, \ldots, n_\text{perm}$), the *p*-value is calculated as the proportion of D_k^π values that equal or exceed D_obs:

$$ p = \frac{ \sum_{k=1}^{n_\text{perm}} (\text{I}(D_k^\pi \geq D_\text{obs}) + 1)}{(n_\text{perm} + 1)} $$

with $\text{I}(\textit{expression}) = 1$ if $\textit{expression}$ is true and zero otherwise.(Note that the "+1" in the numerator and denominator here acknowledge and include D_obs as a member of the permutation distribution; the original order is one possible ordering of the data, after all!)

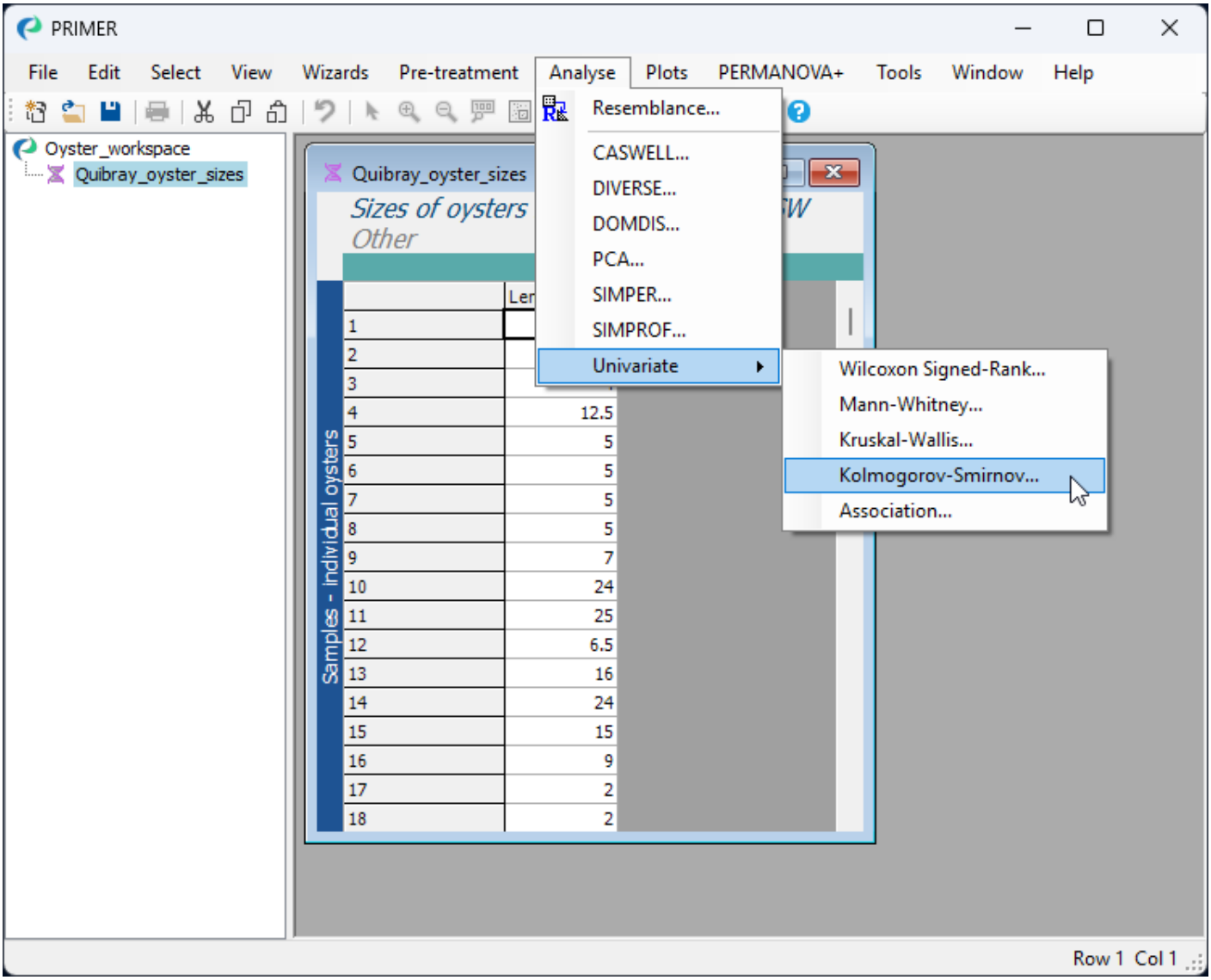
4.8 Example: Sizes of oysters

To demonstrate the Kolmogorov-Smirnov test in PRIMER, we shall return to the dataset consisting of length measurements (in mm) of the Sydney rock oyster (*Saccostrea commercialis*) settling on four different types of surfaces in intertidal estuarine environments (deployed as 10 cm x 10 cm settlement panels at an oyster farm) in Quibray Bay, New South Wales, Australia (Anderson (1992) , Anderson & Underwood (1994)). Rock oysters comprised one of the dominant species colonising the settlement panels in this study (in terms of area), and interest lies in examining the distributions of sizes of these oysters on different types of surfaces.

The data are contained in the file 'Quibray_oyster_sizes.pri', found in the 'Quibray_oysters' folder in 'Examples_P8'. These data were also examined in sections 2.2 and 3.2 above. Each row of the data file contains the length measurement for an individual oyster (in mm), and the factor 'Substratum' identifies the type of surface (concrete, marine plywood, fibreglass or aluminium) to which each measured oyster was attached.

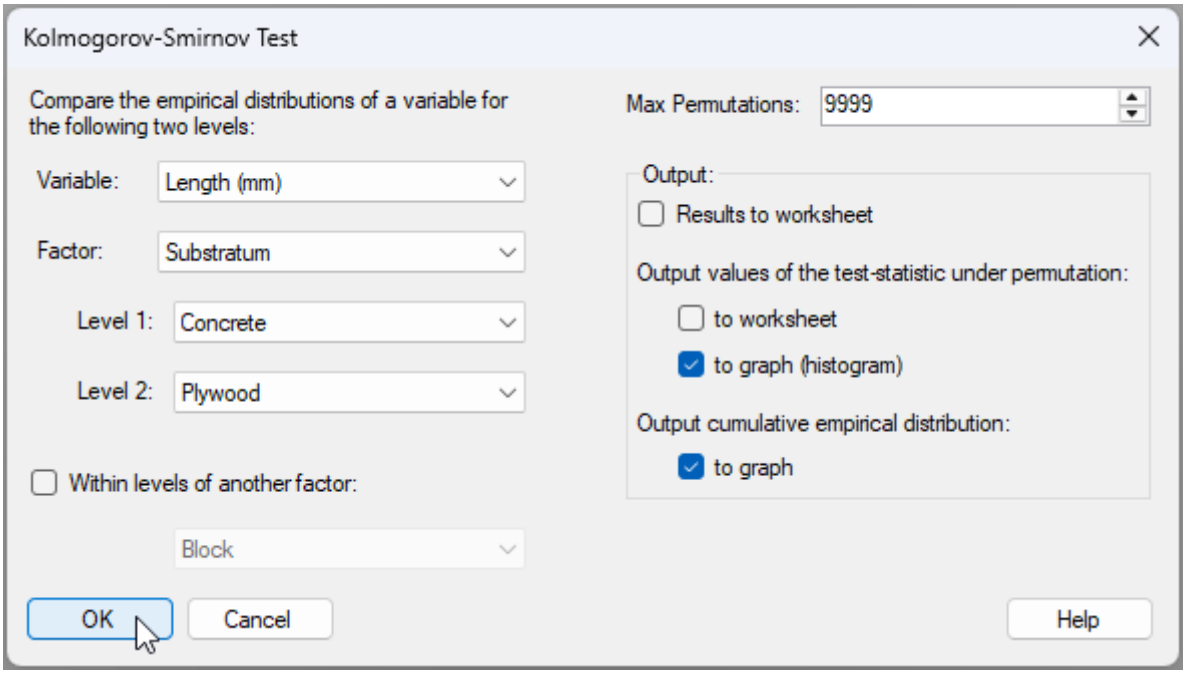
Here, we shall test the null hypothesis of no difference in the distribution of sizes of oysters colonising the concrete surfaces *vs* those colonising marine plywood surfaces.

1. Open the file named 'Quibray_oyster_sizes.pri' in PRIMER 8. Run the Kolmogorov-Smirnov test by clicking **Analyse** > **Univariate** > **Kolmogorov-Smirnov...**, as shown below.

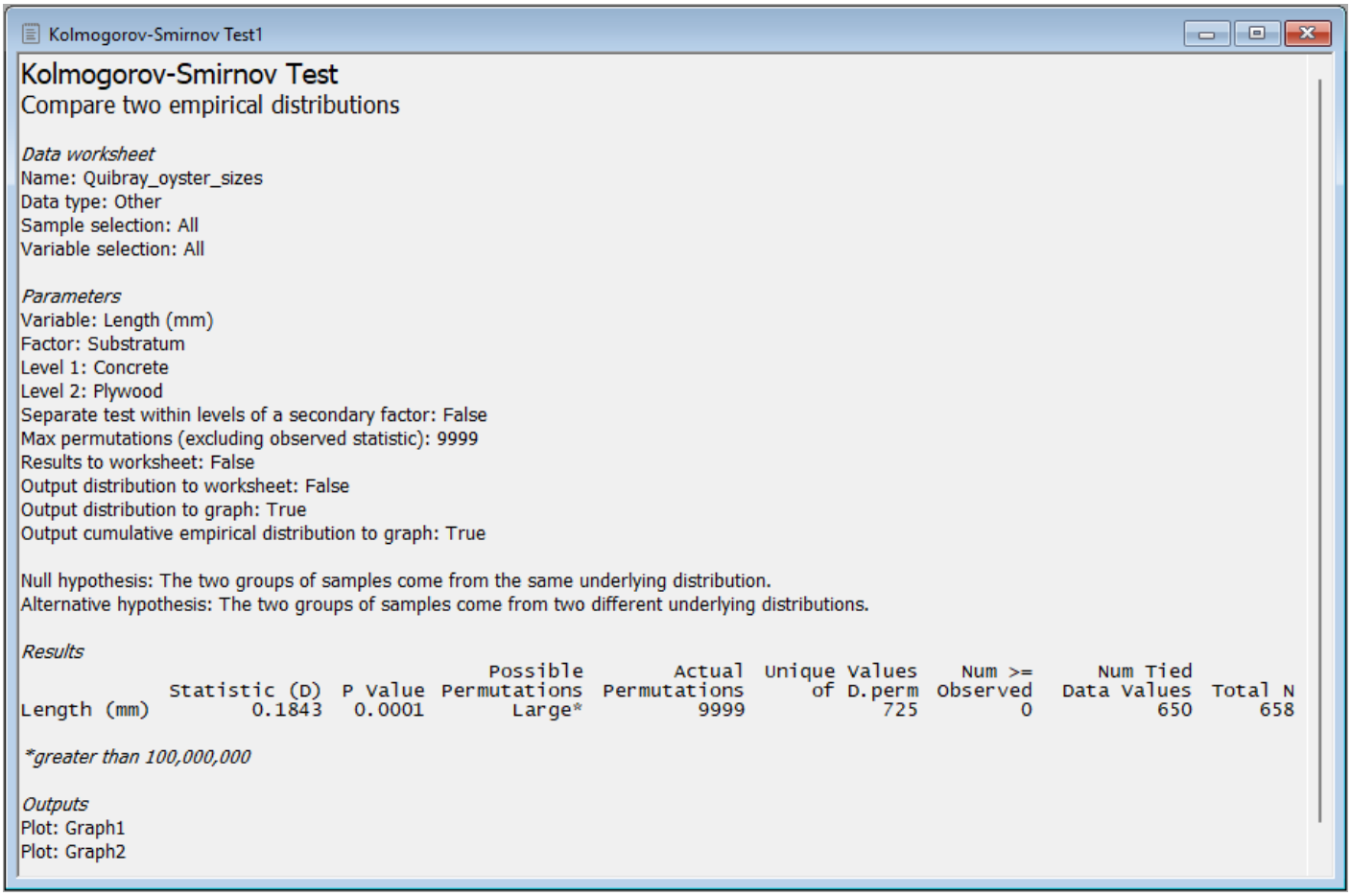


2. In the resulting Kolmogorov-Smirnov Test dialog, choose the following:
   - Variable: Length (mm)
   - Factor: Substratum
     - Level 1: Concrete
     - Level 2: Plywood
   - Max permutations: 9999
   - Output values of the test-statistic under permutation: $\checkmark$to graph (histogram)
   - Output cumulative empirical distribution: $\checkmark$to graph

then click '**OK**', as shown below.



3. Running the test yields a file that shows all of the choices made by the end-user ('*Parameters*'), as well as the results of the test ('*Results*'), viz:



Kolmogorov-Smirnov Test1

Kolmogorov-Smirnov Test
Compare two empirical distributions

Data worksheet
Name: Quibray_oyster_sizes
Data type: Other
Sample selection: All
Variable selection: All

Parameters
Variable: Length (mm)
Factor: Substratum
Level 1: Concrete
Level 2: Plywood
Separate test within levels of a secondary factor: False
Max permutations (excluding observed statistic): 9999
Results to worksheet: False
Output distribution to worksheet: False
Output distribution to graph: True
Output cumulative empirical distribution to graph: True

Null hypothesis: The two groups of samples come from the same underlying distribution.
Alternative hypothesis: The two groups of samples come from two different underlying distributions.

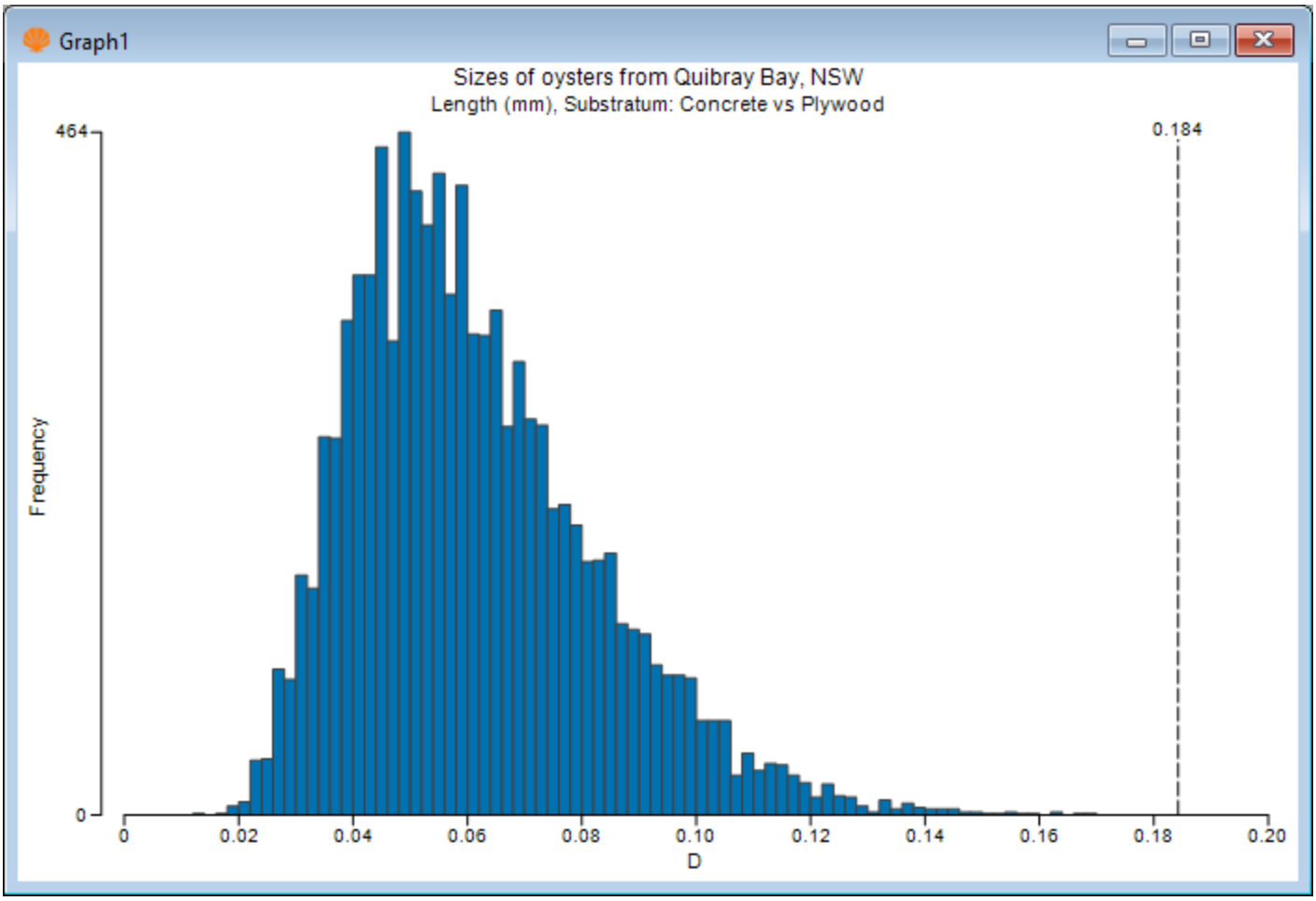
Results

	Statistic (D)	P Value	Possible Permutations	Actual Permutations	Unique Values of D.perm	Num >= Observed	Num Tied Data Values	Total N
Length (mm)	0.1843	0.0001	Large*	9999	725	0	650	658

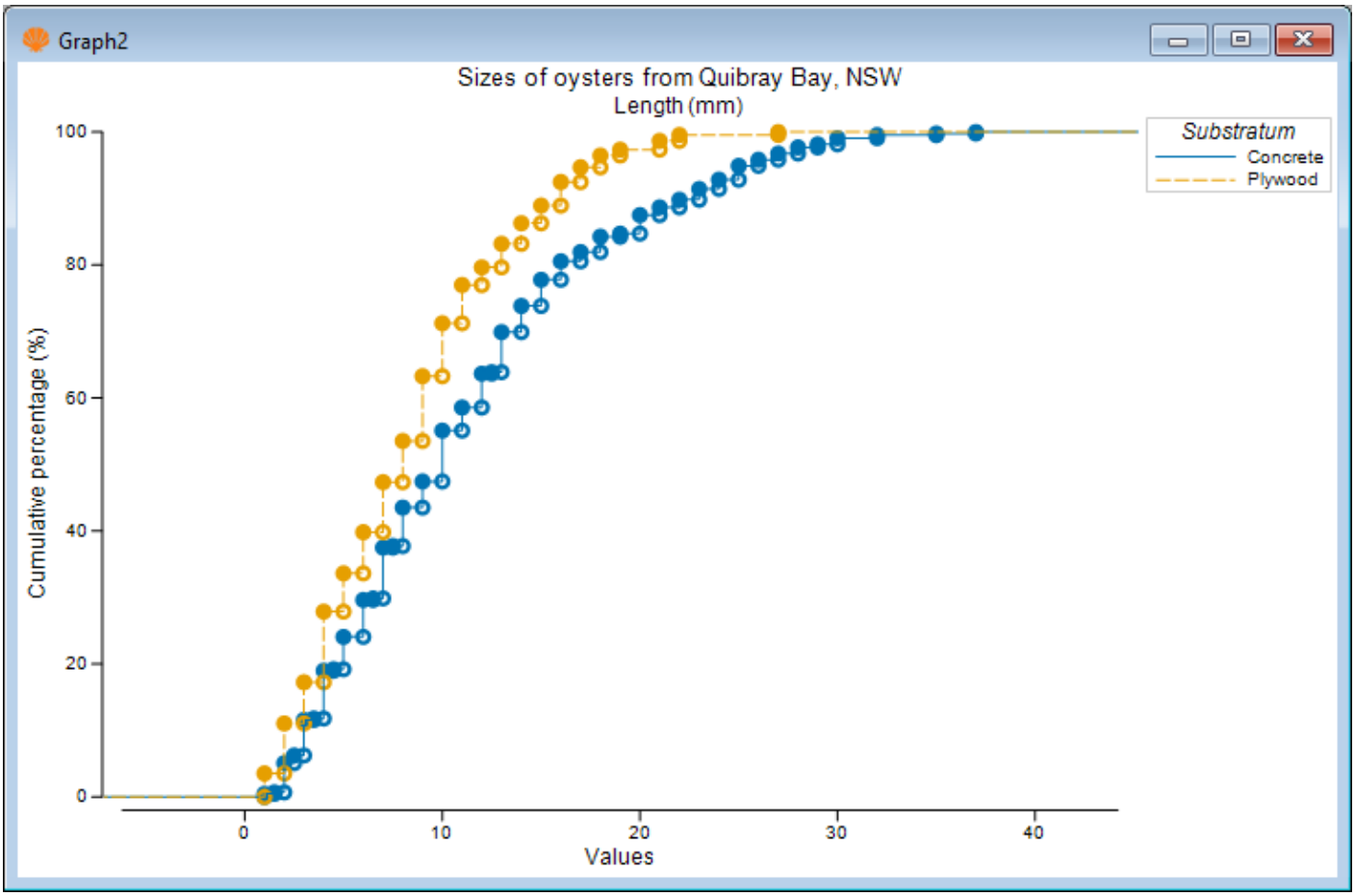
*greater than 100,000,000

Outputs
Plot: Graph1
Plot: Graph2

A histogram of the distribution of the test-statistic (*D*) is also provided, as shown below ('Graph1'):



A natural graphic to look at here, to accompany the test and to help characterise the two distributions of oyster sizes, is an Empirical Distribution plot (provided as 'Graph2' in the Explorer tree, and shown below).



There was a significant difference between concrete and marine plywood surfaces with respect to the distribution of sizes of oysters on them (D = 0.184, P = 0.0001). Clearly, there were proportionately many more oysters of larger sizes on concrete surfaces, compared to marine plywood. This is highly likely to have been caused by more rapid initial colonisation of oysters on the alkaline surfaces of concrete in the early stages of deployment, and faster growth of oysters on concrete over time (Anderson (1996)).

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 $i = 1, \ldots, N$ 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: $$ \rho_{\tiny{P}} = \frac { \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) } {\sqrt{ \sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2 } } $$ 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:

$$ \rho_{\tiny{S}} = \frac { \sum_{i=1}^N (r_{xi} - \bar{r}_ {x})(r_{yi} - \bar{r}_ {y}) } {\sqrt{ \sum_{i=1}^N (r_{xi} - \bar{r}_ {x})^2 \sum_{i=1}^N (r_{yi} - \bar{r}_ {y})^2 } } $$

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:

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

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 $\rho_{\tiny{S}}$ 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).

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

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 $k = 1, \ldots, g_{\tiny{X}}$ 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 $\ell = 1, \ldots, g_{\tiny{Y}}$ 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:

$$ \tau_{\tiny{W}} = \frac{ (n_{\text{con}} - n_{\text{dis}}) } { \sqrt{ (n_{\text{pairs}} - T_{\tiny{X}}) (n_{\text{pairs}} - U_{\tiny{Y}}) } } $$ where $$ T_{\tiny{X}} = \sum_k t_k(t_k-1)/2 \hspace{1cm} \text{and} \hspace{1cm} U_{\tiny{Y}} = \sum_{\ell} u_{\ell}(u_{\ell}-1)/2 $$

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 ($I_{\tiny{A}}$) 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 $i = 1, \ldots, N$ sampling units. It is calculated as:

$$ I_{\tiny{A}} = 100 \times {\Bigg \lbrace} 1 - \frac{1}{2} \left| { \frac{x_i}{ \sum_i x_i } - \frac{y_i}{ \sum_i y_i } } \right| {\Bigg \rbrace} $$

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 $I_{\tiny{A}}$ to the following, which we shall refer to as the *adjusted index of association*:

$$ I_{\tiny{A}}^\star = (2I_{\tiny{A}}/100) - 1 $$

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 $I_{\tiny{A}}^\star$ 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*

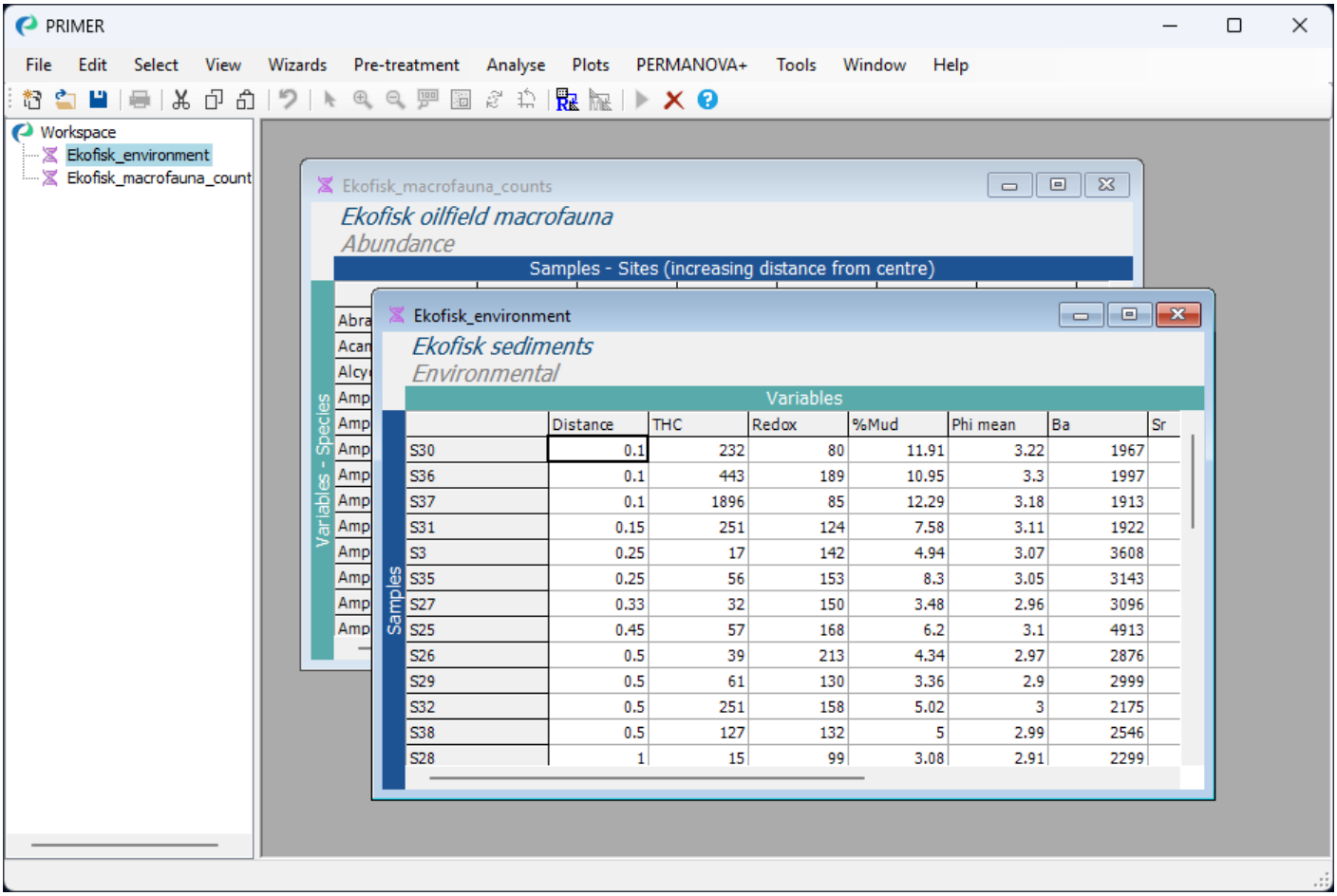
distribution), even though the value of the index itself (the test statistic) is greater than zero (positive). This need not pose any practical or logical problem for the end-user. Viewing the permutation distribution, and the position of the observed value relative to it, is the essential point and will be extremely informative regarding the appropriate alternative hypothesis for any given pair of variables.

---

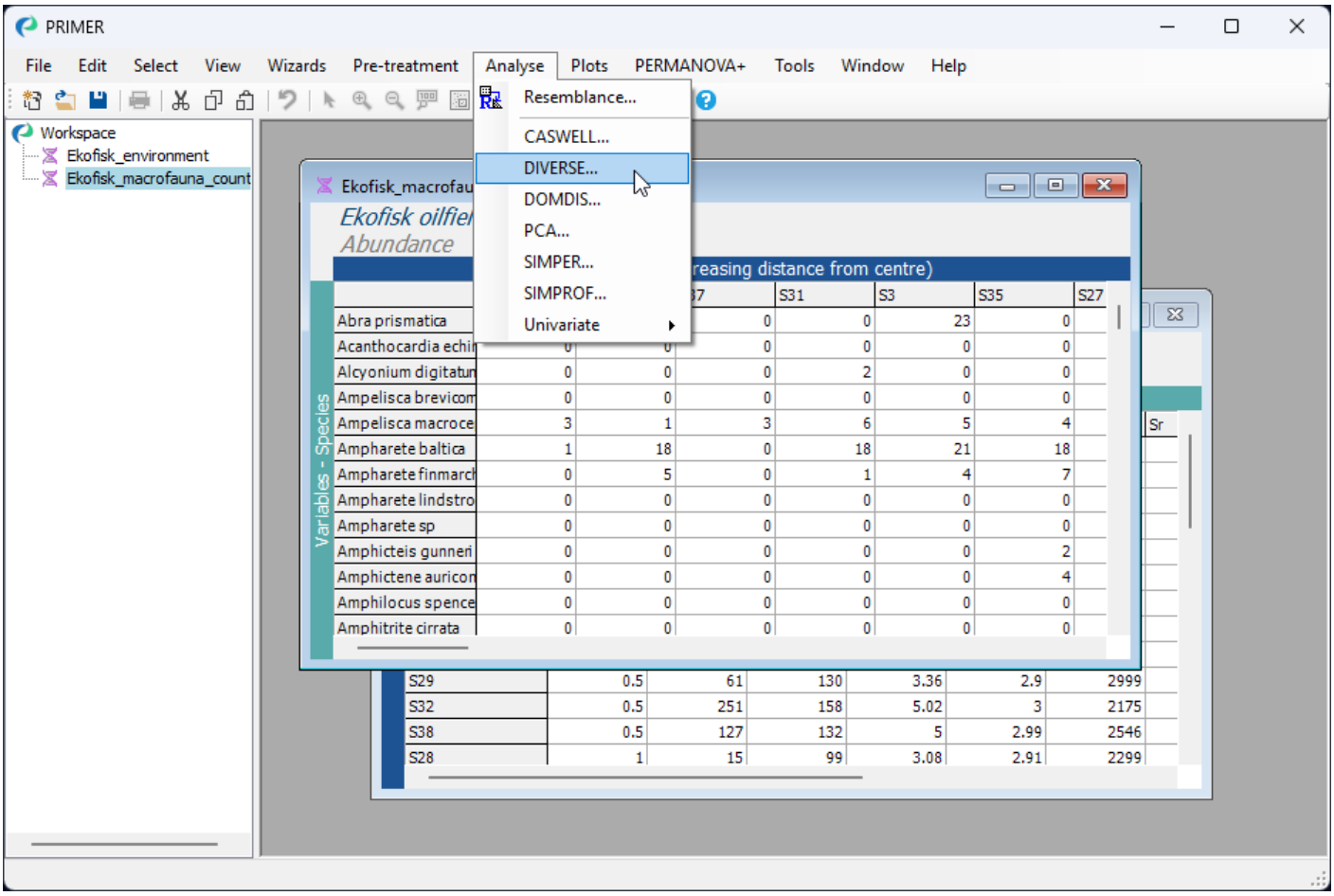
† *In the PRIMER software, when running the test of association using the Index of Association as the measure, the test-statistic examined under permutation is the adjusted index, referred to as 'I.adj' in the output file, while the original index of association (also provided in the output, for reference)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

4.10 Example: Ekofisk diversity

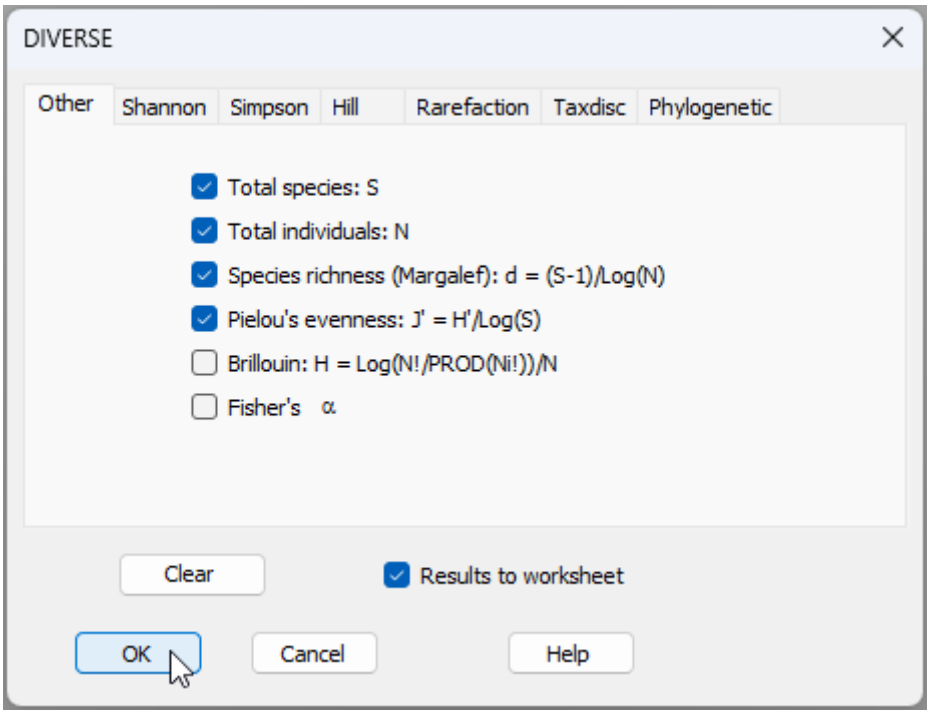
To demonstrate the test of association, we shall re-visit a dataset of macrofauna assemblages collected from sites near an oilfield in the North Sea. In this study by Gray *et al.* (1990) , macrofauna were sampled from 39 sites in an approximately 5-spoke radial design, at increasing distances from undersea drilling activities at the Ekofisk oilfield. There were three day-grab samples taken from each site; information from these were combined to yield abundance values for each of p = 173 soft-sediment macrofaunal taxa at every site. The macrofauna data are given in the file 'Ekofisk_macrofauna_counts.pri', found in the 'Examples_P8 > Ekofisk_macrofauna' folder. Environmental variables were also obtained at each site, and these, along with the actual distance from the oilfield centre, are given in the file 'Ekofisk_environment.pri'. Open up both of these files in PRIMER, as shown below:



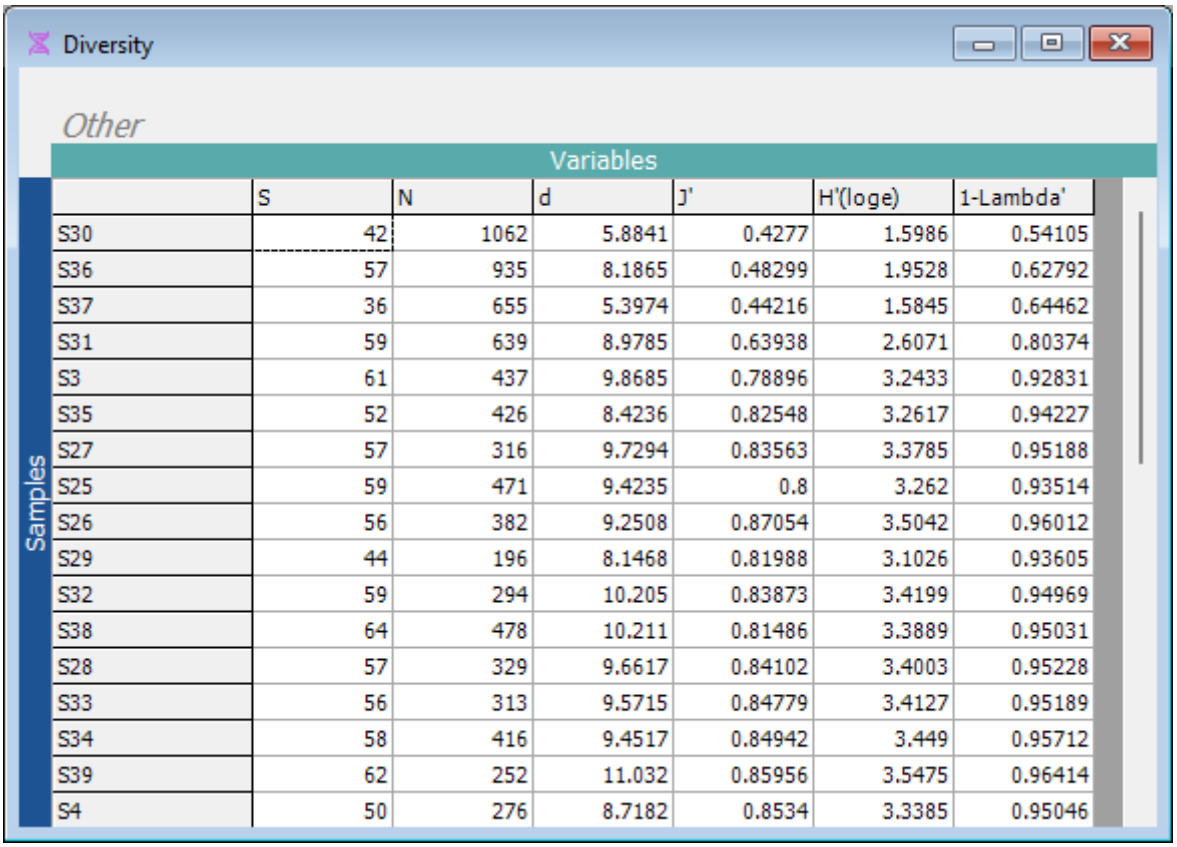
Does species richness increase or decrease with increasing distance from the oil platform? Let's start by calculating some simple univariate diversity metrics, such as S = total richness (the number of taxa), along with some others. From the Ekofisk_macrofauna_counts data sheet inside PRIMER, click **Analyse** > **DIVERSE...**.



We can take all of the default options here, but also tick the option to output ($\checkmark$Results to worksheet), then click '**OK**'.



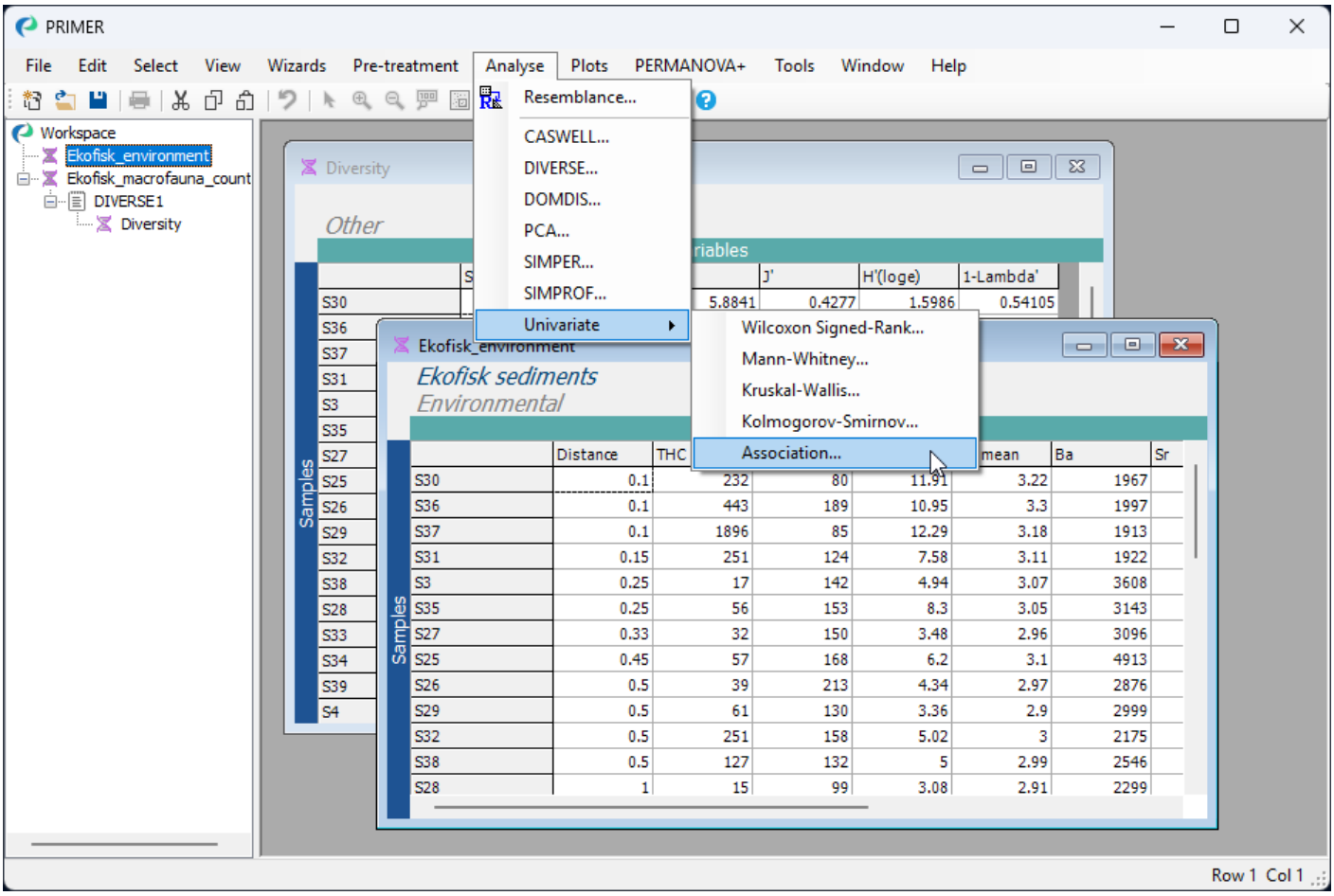
You can (optionally) re-name the resulting data sheet of diversity metrics (called 'Data1') by clicking on that sheet, then click **File** > **Rename Data** and type a new name 'Diversity', then click '**OK**'. The suite of univariate diversity metrics, calculated on each sampling unit, are now evident in this renamed data sheet, as follows:



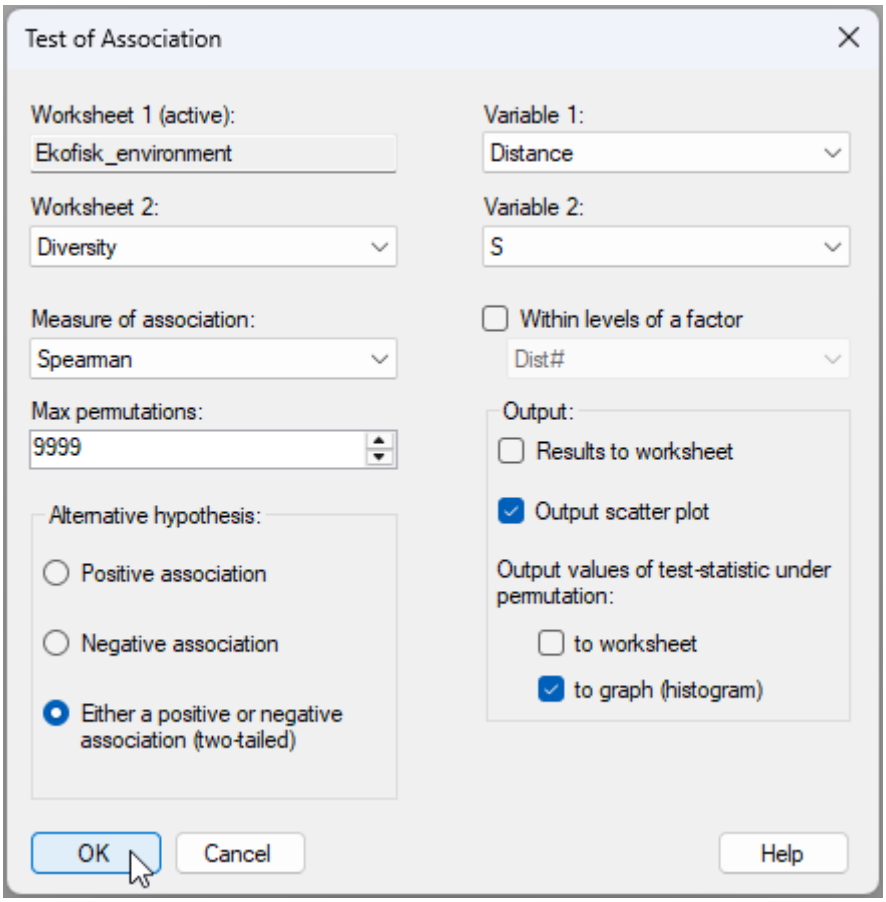
	S	N	d	J'	H'(loge)	1-Lambda'
S30	42	1062	5.8841	0.4277	1.5986	0.54105
S36	57	935	8.1865	0.48299	1.9528	0.62792
S37	36	655	5.3974	0.44216	1.5845	0.64462
S31	59	639	8.9785	0.63938	2.6071	0.80374
S3	61	437	9.8685	0.78896	3.2433	0.92831
S35	52	426	8.4236	0.82548	3.2617	0.94227
S27	57	316	9.7294	0.83563	3.3785	0.95188
S25	59	471	9.4235	0.8	3.262	0.93514
S26	56	382	9.2508	0.87054	3.5042	0.96012
S29	44	196	8.1468	0.81988	3.1026	0.93605
S32	59	294	10.205	0.83873	3.4199	0.94969
S38	64	478	10.211	0.81486	3.3889	0.95031
S28	57	329	9.6617	0.84102	3.4003	0.95228
S33	56	313	9.5715	0.84779	3.4127	0.95189
S34	58	416	9.4517	0.84942	3.449	0.95712
S39	62	252	11.032	0.85956	3.5475	0.96414
S4	50	276	8.7182	0.8534	3.3385	0.95046

Run the test of association

Now, let's run the test of association. We wish to test the null hypothesis of 'no association' between species richness (the variable called 'S' in the 'Diversity' data sheet) and distance from the oil platform's drilling activity (the variable called 'Distance' in the 'Ekofisk_environment' data sheet). From the 'Ekofisk_environment' data sheet, click **Analyse** > **Univariate** > **Association...**.

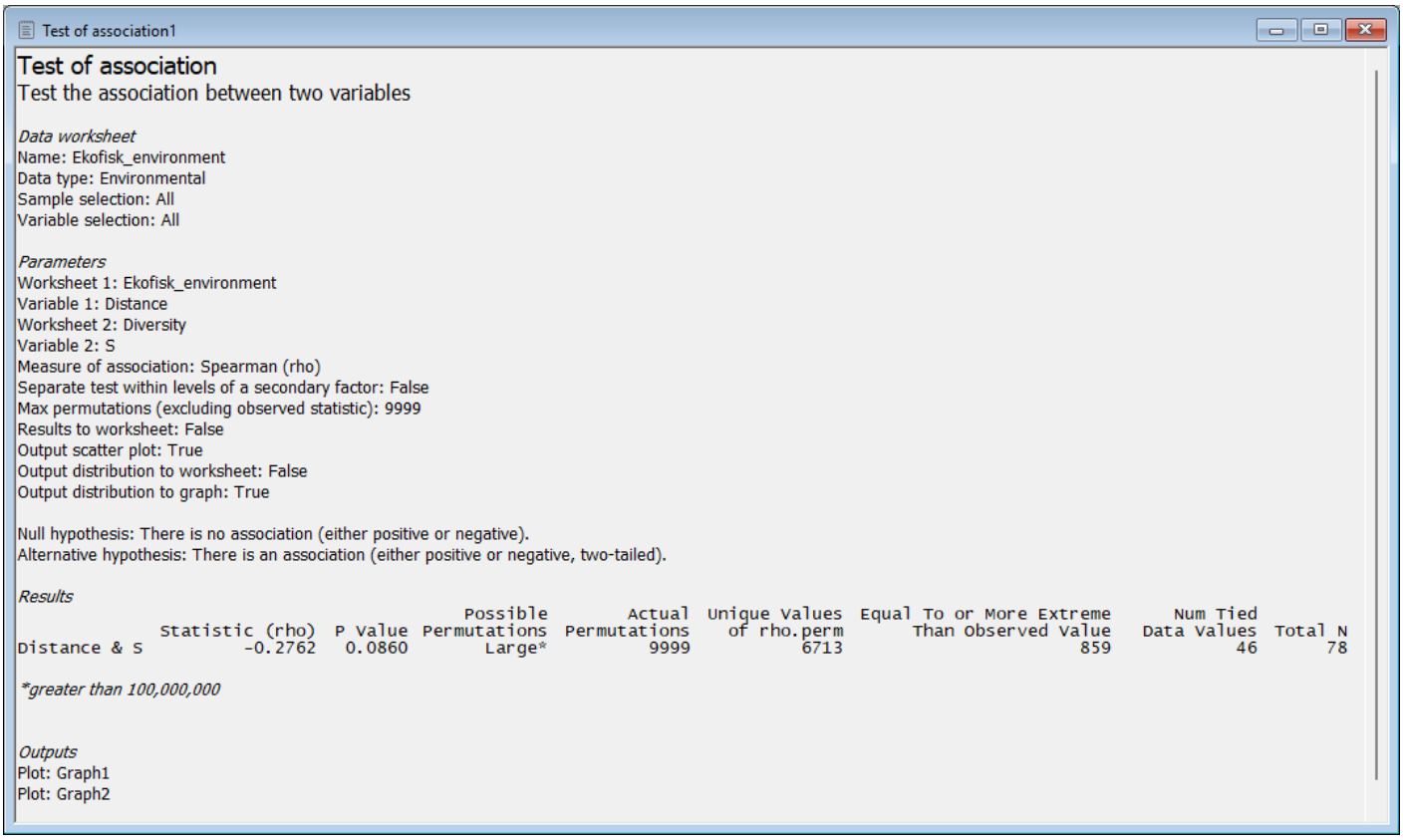


Complete the relevant information required for this test into the resulting dialog window, as shown below, then click '**OK**'.

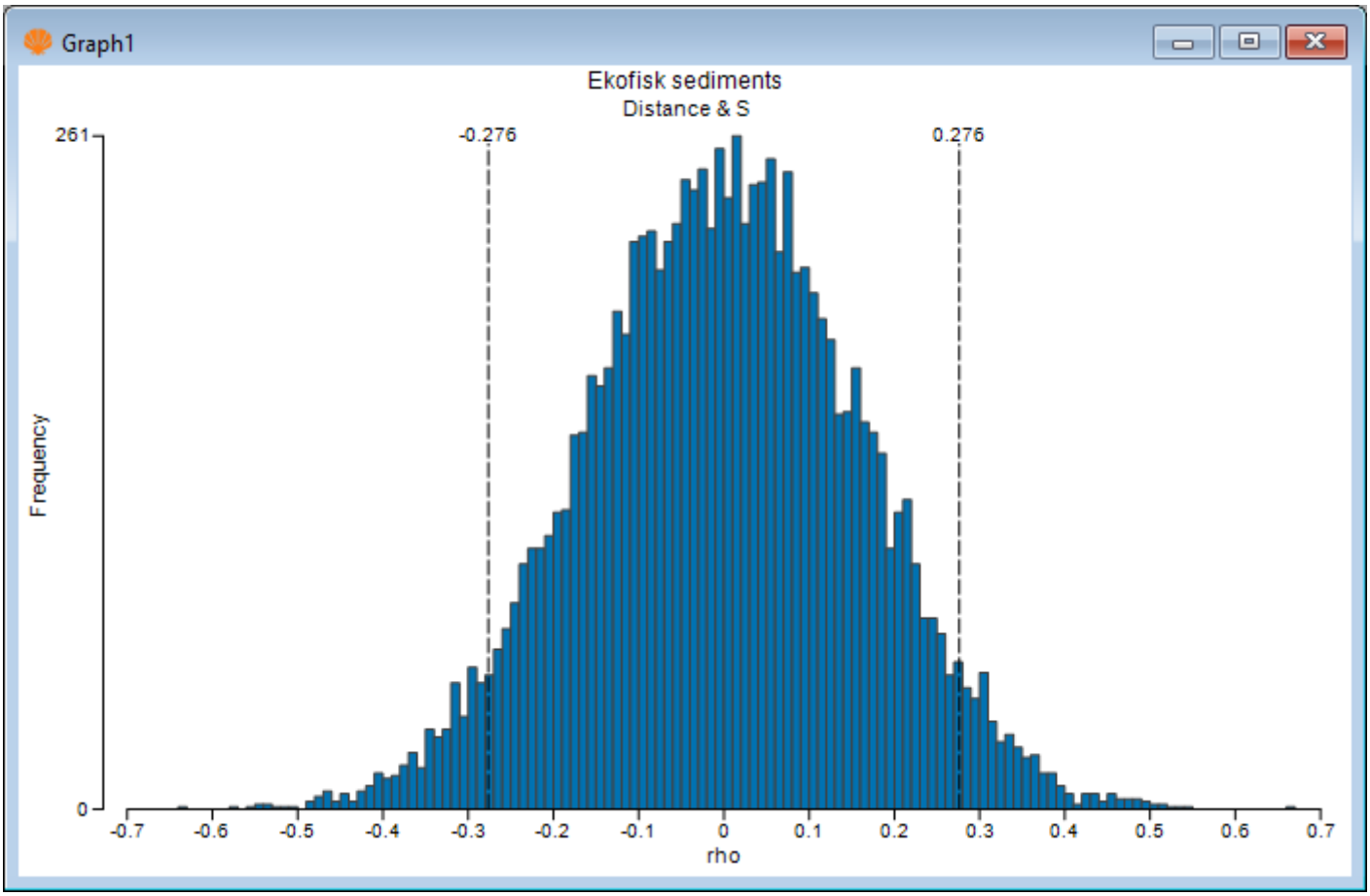


Results of the test of association

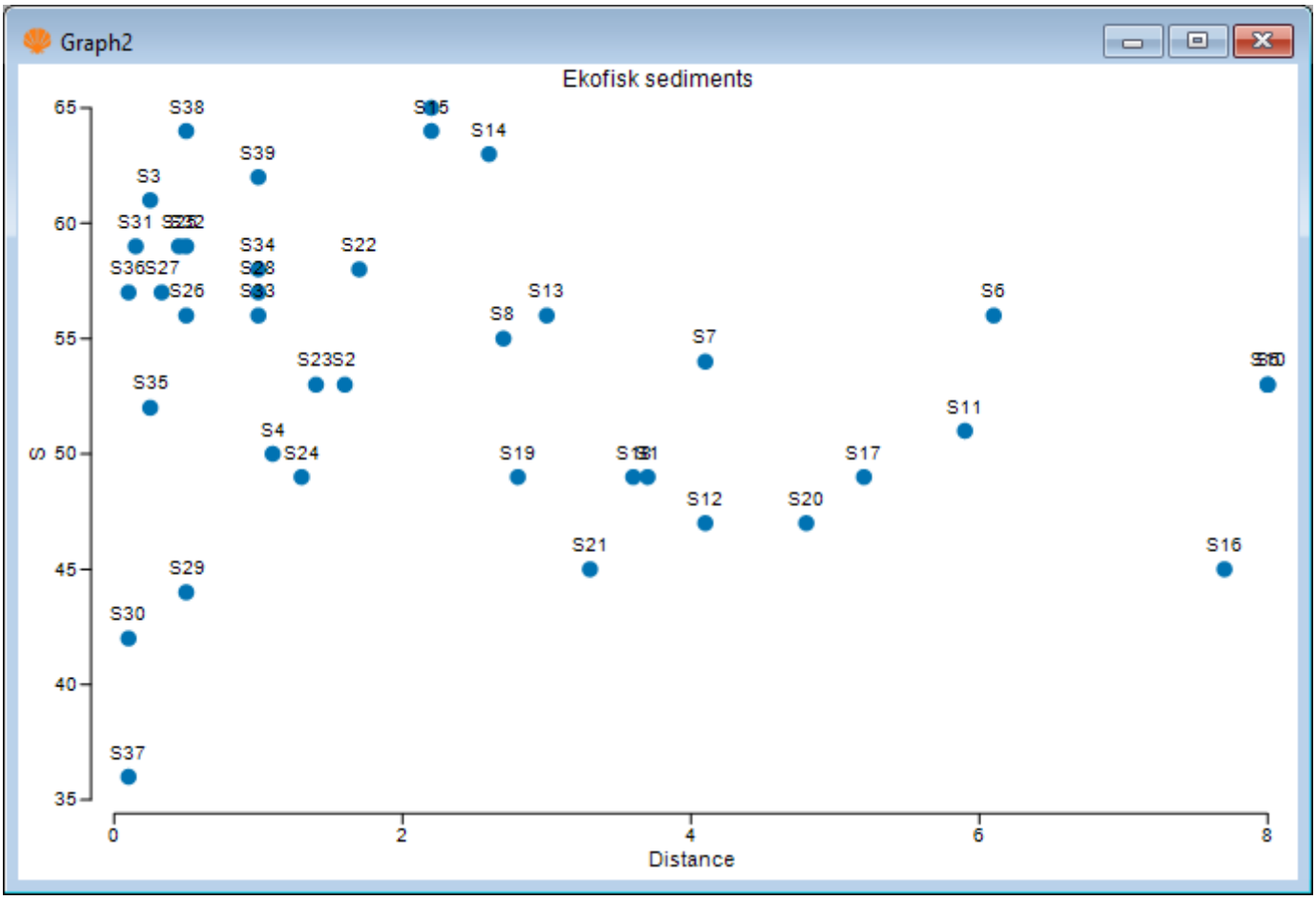
You will see the following output (called 'Test of association1' in the Explorer tree window).



Additional output for this example includes two graphics. First, there is a histogram of the values of the test statistic (here, Spearman's rank correlation, *rho* $= \rho_{\tiny{S}}$) obtained under permutation ('Graph1'), *viz*:



Second, there is a scatterplot of the two variables ('Graph2'), where whatever variable was provided as 'Variable 1' in the dialog will be the x-axis, and the variable named as 'Variable 2' will be the y-axis, as shown here:



Interestingly, although some samples close to the platform (i.e., S37, S30 and S29) had quite low values for species richness compared to most other sites, there is nevertheless a negative rank correlation measured between species richness and distance from the oil platform ($ \rho_{\tiny{S}}$ = -0.2762); however, this was not statistically significant overall (P > 0.08) according to the test. Indeed, in many practical ecological applications, univariate diversity measures may or may not show a very strong directional pattern of response to environmental impact. In contrast, multivariate analyses of community structure, holistically, will tend to pick up significant effects that can then be characterised in terms of simultaneous changes in relative occurrences and/or abundances across a suite of component species.

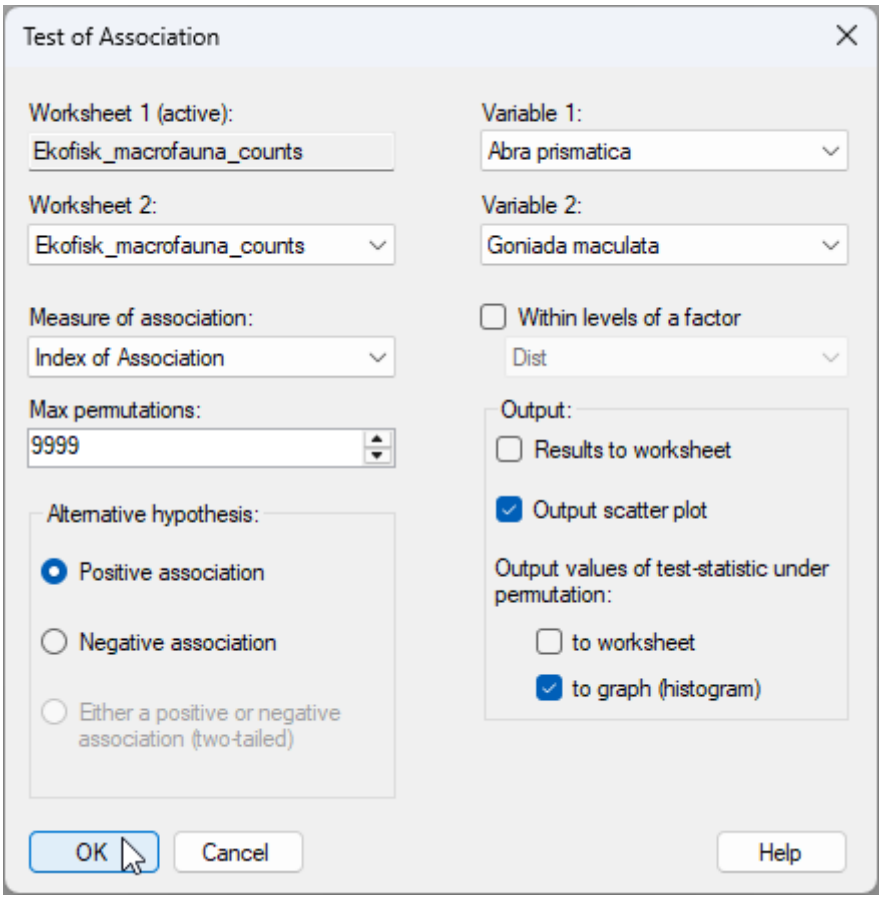
4.11 Example: Associations between species

It is instructive to consider some additional examples of the test of association where the variables are not evenly distributed. Specifically, we wish to cater for situations where the variables of interest are occurrences, densities or counts of species' abundances, as commonly encountered in community ecology.

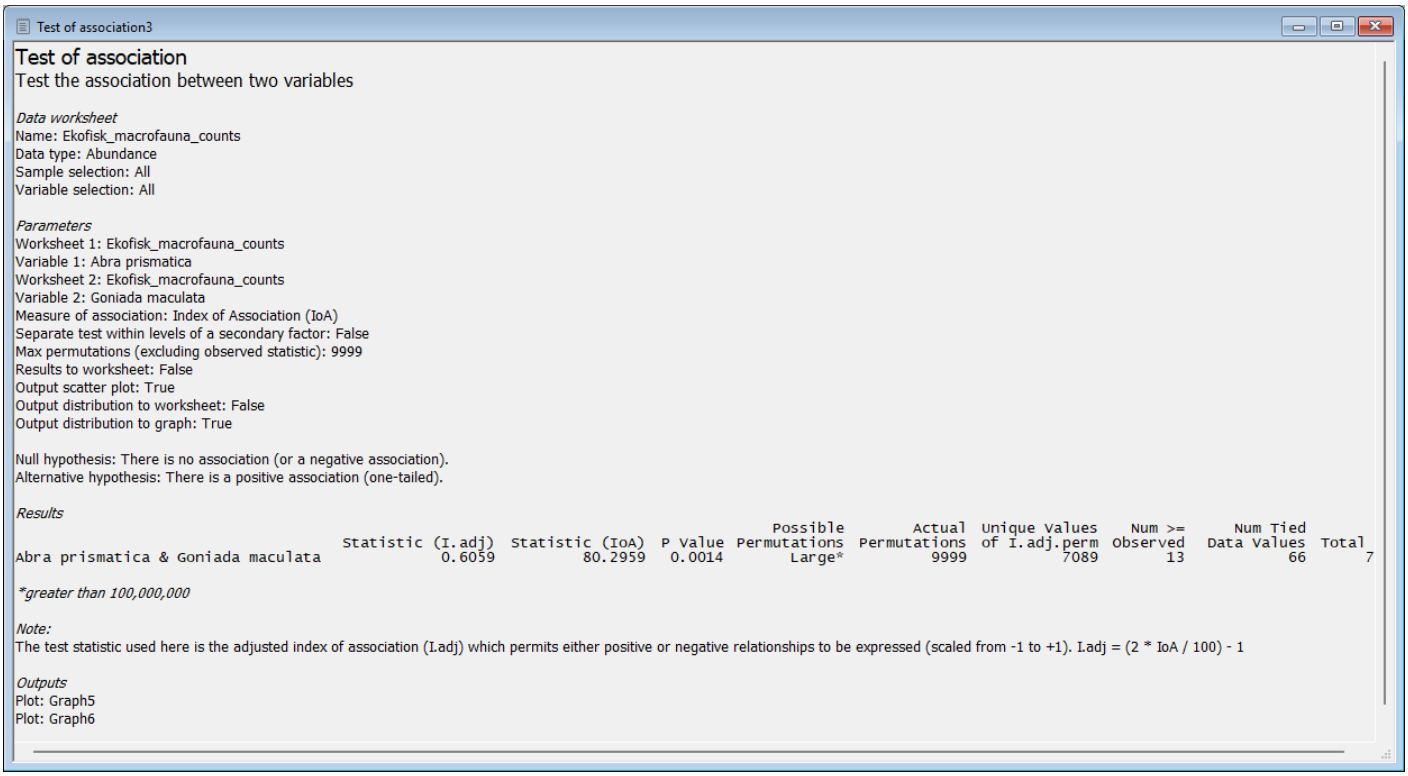
In such cases, we should consider using the *index of association* ($I_{\tiny{A}}$ or $I_{\tiny{A}}^\star$) as our measure, because, in the majority of cases, there can be no sense in including joint absences (i.e., where both species take a value of zero, corresponding to sites where neither species occurs) in the consideration of whether the individuals of the two species are likely to be found together or not. Sites that contain neither species at all are simply non-informative in this regard.

Abra prismatica & Goniada maculata

From the 'Ekofisk macrofauna counts' dataset (see the previous page for how to access this data set), choose **Analyse** > **Univariate** > **Association...** and choose the bivalve, *Abra prismatica*, as 'Variable 1' and the polychaete *Goniada maculata* as 'Variable 2' (leaving defaults for the rest), then click '**OK**'.



The output file shows a statistically significant positive association between these two variables; the index of association is $I_{\tiny{A}}$ = 80.3 and P < 0.01.



Test of association3

Test of association

Test the association between two variables

Data worksheet
Name: Ekofisk_macrofauna_counts
Data type: Abundance
Sample selection: All
Variable selection: All

Parameters
Worksheet 1: Ekofisk_macrofauna_counts
Variable 1: Abra prismatica
Worksheet 2: Ekofisk_macrofauna_counts
Variable 2: Goniada maculata
Measure of association: Index of Association (IoA)
Separate test within levels of a secondary factor: False
Max permutations (excluding observed statistic): 9999
Results to worksheet: False
Output scatter plot: True
Output distribution to worksheet: False
Output distribution to graph: True

Null hypothesis: There is no association (or a negative association).
Alternative hypothesis: There is a positive association (one-tailed).

Results

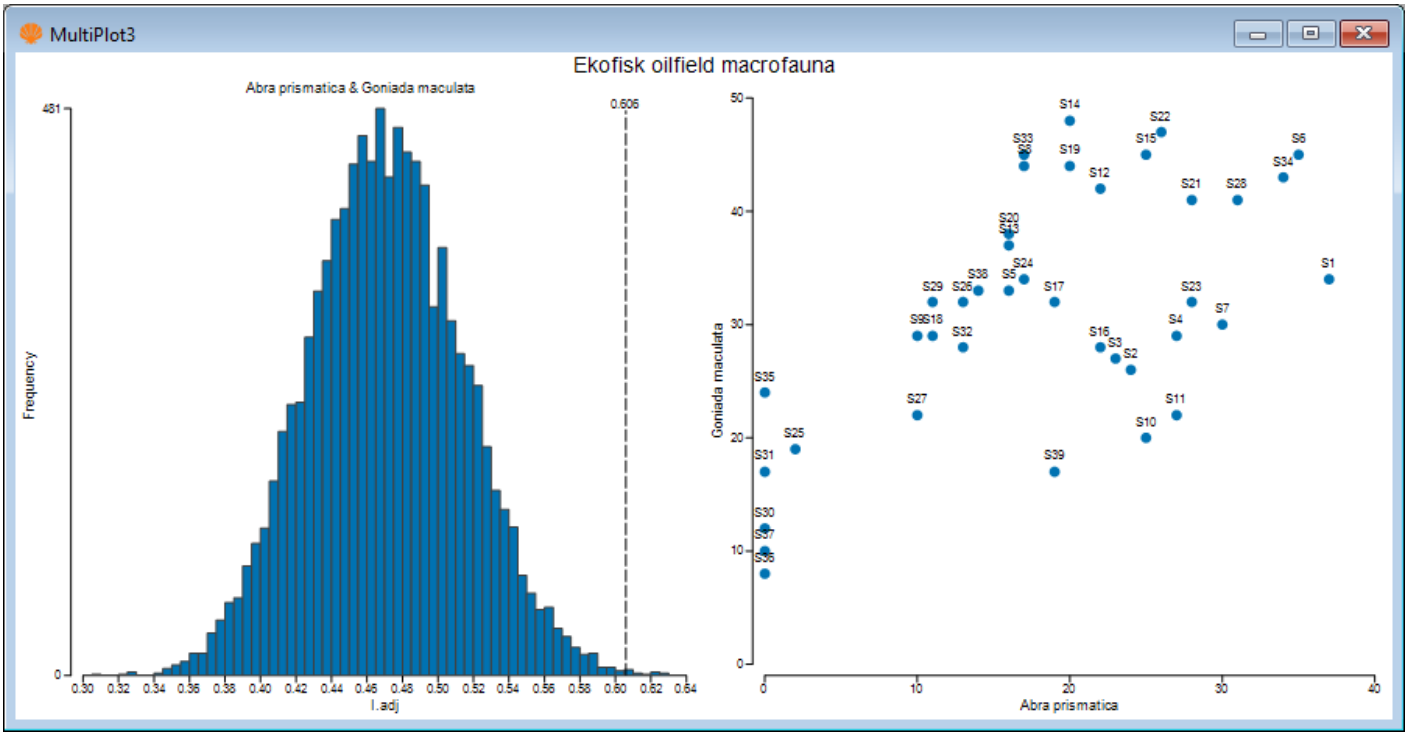
	Statistic (I.adj)	Statistic (IoA)	P Value	Possible Permutations	Actual Permutations	Unique Values of I.adj.perm	Num >= Observed	Num Tied Data Values	Total
Abra prismatica & Goniada maculata	0.6059	80.2959	0.0014	Large*	9999	7089	13	66	7

greater than 100,000,000

Note:
The test statistic used here is the adjusted index of association (Ladj) which permits either positive or negative relationships to be expressed (scaled from -1 to +1). Ladj = (2 * IoA / 100) - 1

Outputs
Plot: Graph5
Plot: Graph6

This association is also evident visually in the scatter plot:



Note that *standardising* the variables first (e.g., by their totals), as would be desirable here, is not necessary to do as a separate step, and would actually have no effect, as that operation is done automatically as part of the calculation of the index itself (see Somerfield & Clarke (2013)).

Amphictene auricoma & Trichobranchus roseus

Another example demonstrates how joint absences (double zeros) across the samples can occasionally yield counter-intuitive results when examining scatter plots. Let's run the test of association on the following two species of polychaete worms: *Amphictene auricoma* & *Trichobranchus roseus*.

The output file and graphics (based on the index of association) are shown below:

Test of association4

Test of association

Test the association between two variables

Data worksheet
Name: Ekofisk_macrofauna_counts
Data type: Abundance
Sample selection: All
Variable selection: All

Parameters
Worksheet 1: Ekofisk_macrofauna_counts
Variable 1: Amphictene auricoma
Worksheet 2: Ekofisk_macrofauna_counts
Variable 2: Trichobranchus roseus
Measure of association: Index of Association (IoA)
Separate test within levels of a secondary factor: False
Max permutations (excluding observed statistic): 9999
Results to worksheet: False
Output scatter plot: True
Output distribution to worksheet: False
Output distribution to graph: True

Null hypothesis: There is no association (or a negative association).
Alternative hypothesis: There is a positive association (one-tailed).

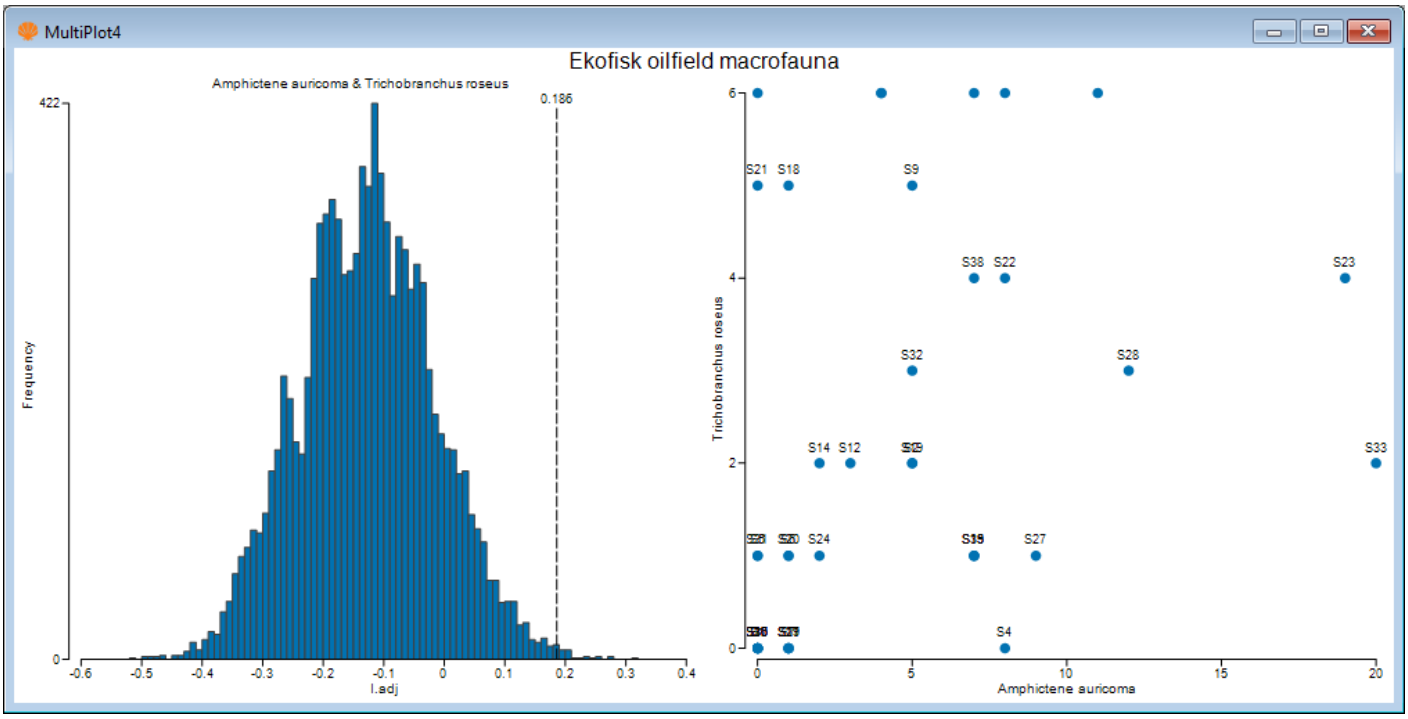
Results

	Statistic (I.adj)	Statistic (IoA)	P Value	Possible Permutations	Actual Permutations	Unique Values of I.adj.perm	Num >= Observed	Num Tied Data Values	Total N
Amphictene auricoma & Trichobranchus roseus	0.186	59.3023	0.0028	Large*	9999	1664	27	73	78

*greater than 100,000,000

Note:
The test statistic used here is the adjusted index of association (I.adj) which permits either positive or negative relationships to be expressed (scaled from -1 to +1). I.adj = (2 * IoA / 100) - 1

Outputs
Plot: Graph7
Plot: Graph8

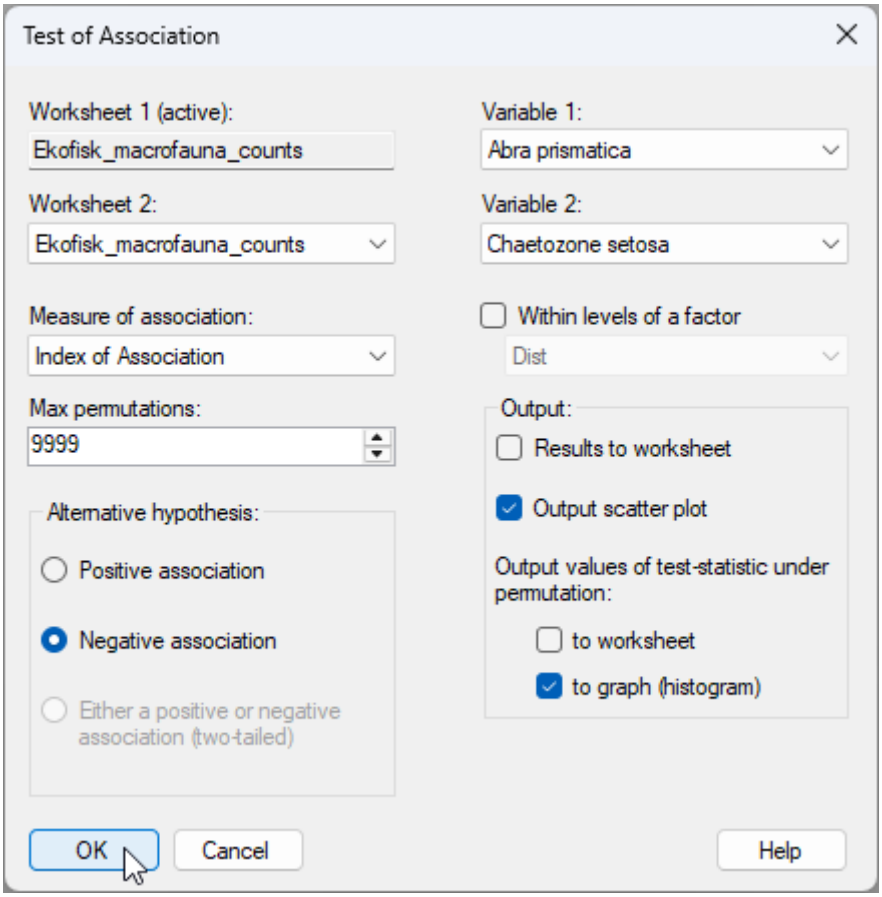


Despite having detected a significant positive association between these two species ($I_{\tiny{A}}$ = 59.3 and P < 0.01), a classic pattern of positive correlation is not particularly obvious in the scatter plot. For these two species, more than a quarter of the data values are equal to zero.¶ We can, however, trust the outcome of the test, but this example serves to show how a raw scatter plot of all joint values (including the joint absences) may not assist us in ascertaining the nature of species' co-occurrence relationships. Other types of plots may be helpful in clarifying similarity in the patterns of species across multiple samples (e.g., boxplots, means plots, line plots, coherence plots, etc.).

Abra prismatica & Chaetozone setosa

Another example demonstrates how the test of association works in the case of a negative association. Consider the following two species: *Abra prismatica* & *Chaetozone setosa*. *A. prismatica* is a bivalve mollusc that can be negatively affected by contaminants in the field, while *C. setosa* is an opportunistic species that can flourish in polluted areas.

Running the test of association on this pair of species is done under the alternative hypothesis of there being a *negative association* between them, like so:



The results are shown below:

Test of association5

Test of association

Test the association between two variables

Data worksheet
Name: Ekofisk_macrofauna_counts
Data type: Abundance
Sample selection: All
Variable selection: All

Parameters
Worksheet 1: Ekofisk_macrofauna_counts
Variable 1: Abra prismatica
Worksheet 2: Ekofisk_macrofauna_counts
Variable 2: Chaetozone setosa
Measure of association: Index of Association (IoA)
Separate test within levels of a secondary factor: False
Max permutations (excluding observed statistic): 9999
Results to worksheet: False
Output scatter plot: True
Output distribution to worksheet: False
Output distribution to graph: True

Null hypothesis: There is no association (or a positive association).
Alternative hypothesis: There is a negative association (one-tailed).

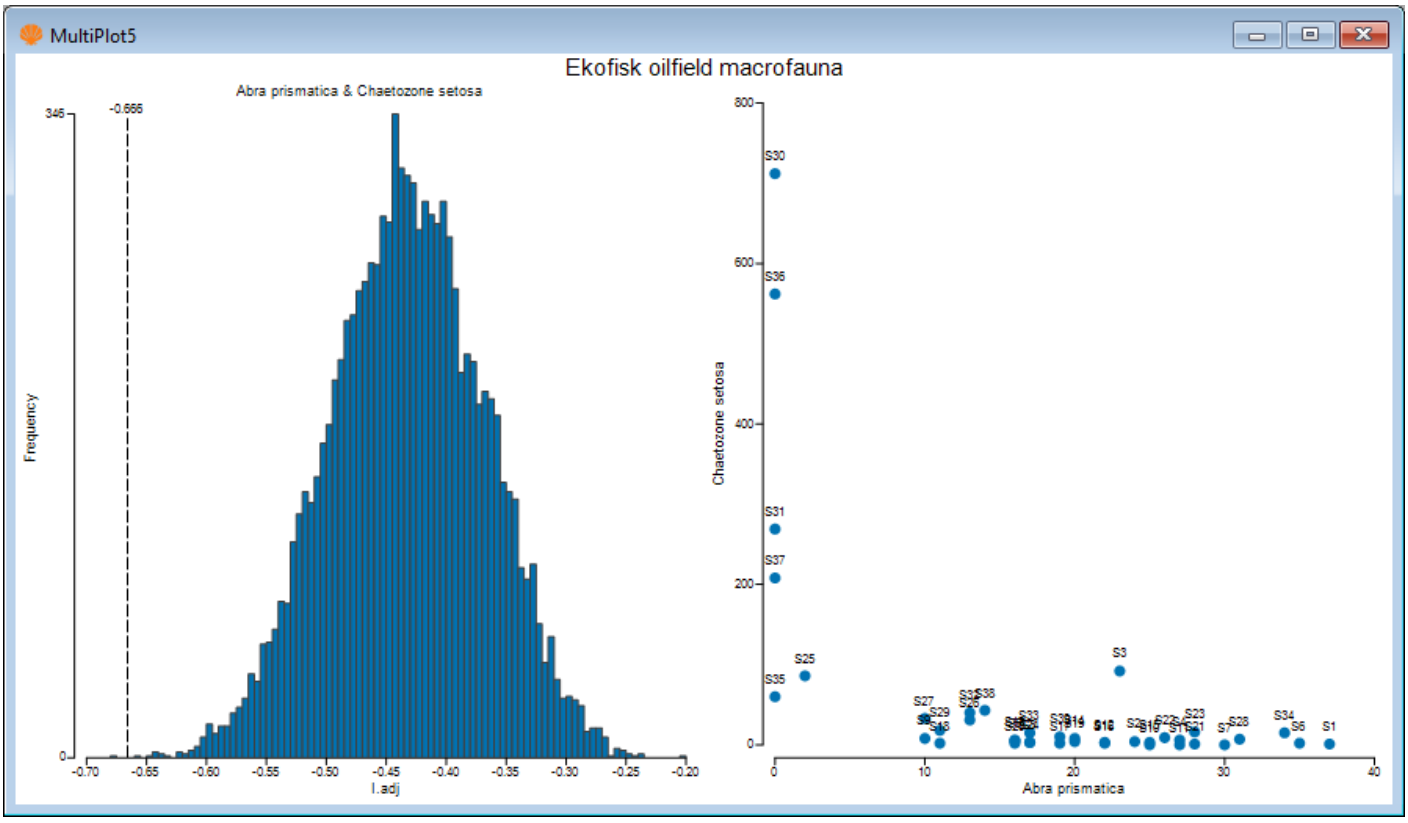
Results

	Statistic (I.adj)	Statistic (IoA)	P Value	Possible Permutations	Actual Permutations	Unique Values of I.adj.perm	Num <= Observed	Num Tied Data Values	Total N
Abra prismatica & Chaetozone setosa	-0.6659	16.7061	0.0002	Large*	9999	8549	1	56	78

*greater than 100,000,000

Note:
The test statistic used here is the adjusted index of association (I.adj) which permits either positive or negative relationships to be expressed (scaled from -1 to +1). I.adj = (2 * IoA / 100) - 1

Outputs
Plot: Graph9
Plot: Graph10



In the past, we would have looked at the (quite low) value for the index of association ($I_{\tiny{A}}$ = 16.7), but there would be no obvious way of asserting any particular statistical significance to this, one way or the other. Now, with the new tool for testing associations in PRIMER 8, we can use our adjusted index of association value as the test-statistic ($I_{\tiny{A}}^\star$ = -0.67) and, under permutation, it is clear that this negative association is highly statistically signifciant (P = 0.0001). In other words, for this dataset, where you find one of these species, you do not tend to find the other one, and *vice versa*.

Note that, in all three of the above tests, the distribution of the index of association under permutation is not centred on zero, nor is it necessarily symmetric; yet, in all three cases, it is easy to examine the output provided (consisting of the empirical permutation distribution and the observed value of the test statistic relative to it) in order to ascertain the appropriate alternative hypothesis for any given test.

¶ *We can very quickly calculate the number of zeros (and the number of non-zeros) for any variable(s) in any dataset by running the new 'Tools > Summary Stats...' in PRIMER 8.*