11.2 Example: Plankton (revisited)

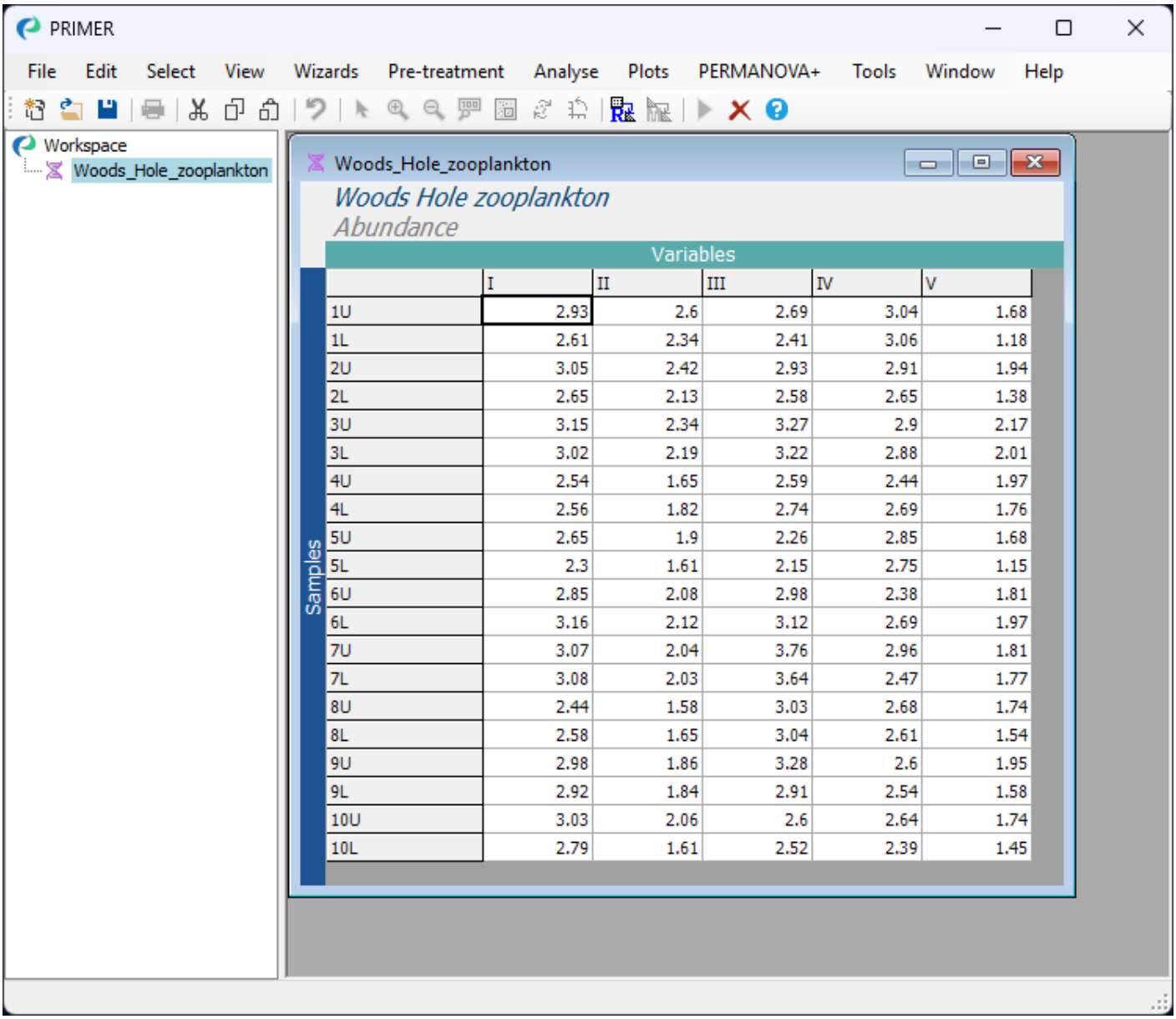
We shall show the utility of being able to construct a residual distance matrix and, from this, a residual ordination plot, by reference to a study of differential catches in plankton nets by Winsor & Clarke (1940) , provided by Snedecor (1946) . We met this dataset earlier, as an example of the use of the Wilcoxon signed-rank test. Recall that these data consist of the total log abundance values for each of five different types of plankton (hence, five variables, named using Roman numerals I, II, III, IV and V) that were caught in each of 2 nets towed simultaneously at 2 different depths: one at 29 m and the other at 31 m. There were ten hauls done in this way. The factors associated with these data are:

- Position (either the upper (U) or the lower (L) net: a fixed factor); and
- Haul (10 hauls labeled 1-10: a random factor).

Earlier (in the context of the Wilcoxon signed-rank test) we considered only a single variable - the total sum of the log abundances across all plankton types. Here, however, we are interested in examining the full multivariate dataset with all original p = 5 variables. These data are located in the file ‘Woods_Hole_zooplankton.pri’, found inside the 'Examples_P8 > Woods_Hole_zooplankton' folder.

Visualise the main patterns using ordination

1. Open up the file (‘Woods_Hole_zooplankton.pri’) in PRIMER. The data values provided here 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
9L	2.92	1.84	2.91	2.54	1.58
10U	3.03	2.06	2.6	2.64	1.74
10L	2.79	1.61	2.52	2.39	1.45

2. From the 'Woods_Hole_zooplankton' data sheet, calculate a Bray-Curtis resemblance matrix by clicking **Analyse** > **Resemblance** and taking all the defaults (click '**OK**'). The result will be 'Resem1', as shown below:

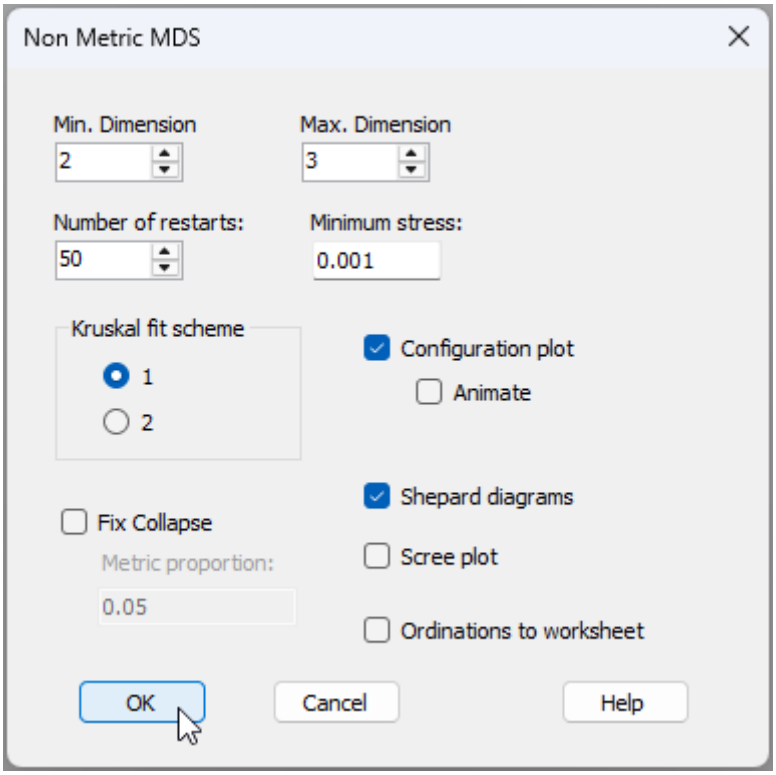
Resem1

Woods Hole zooplankton

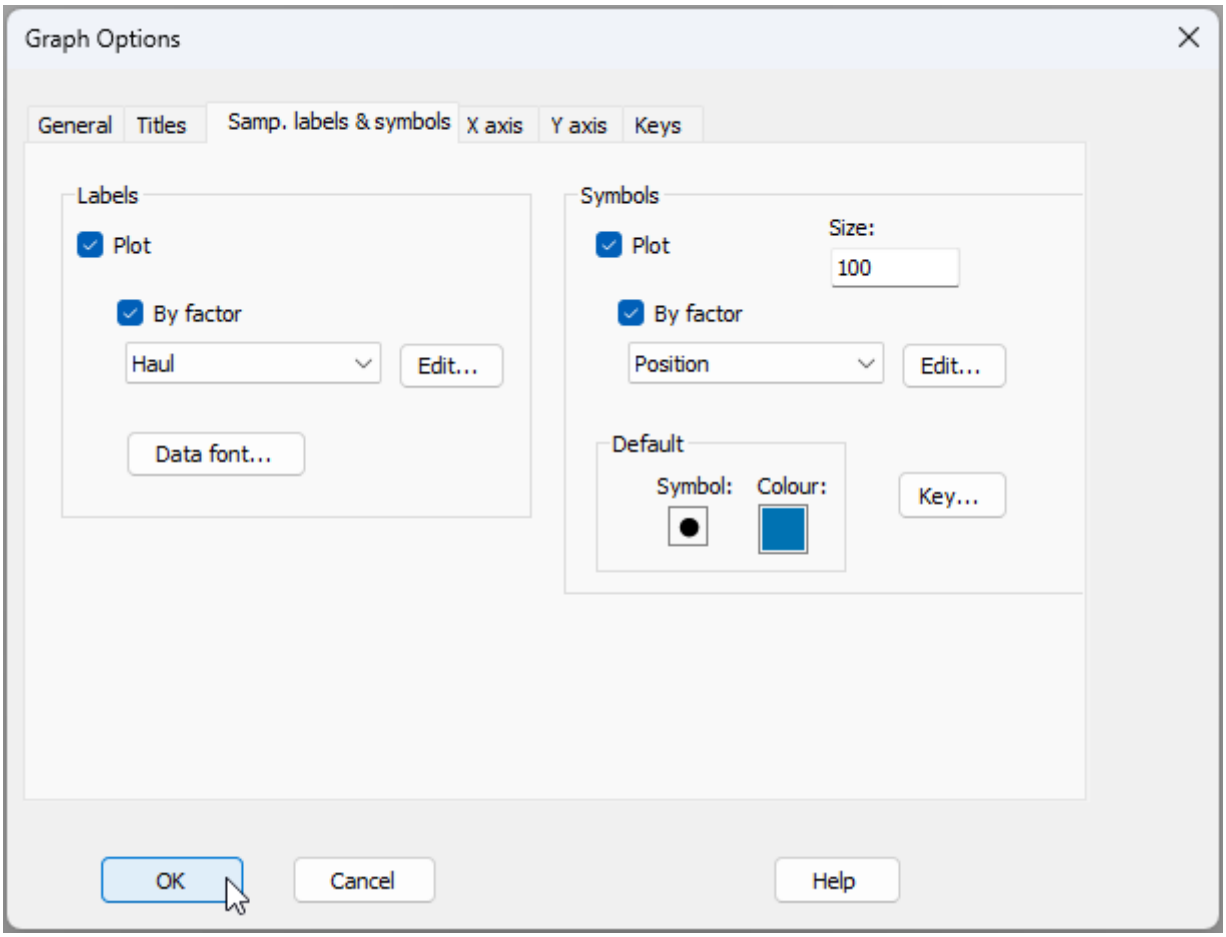
Similarity (0 to 100)

Samples	1U	1L	2U	2L	3U	3L	4U	4
1U								
1L	94.377							
2U	96.449	92.153						
2L	93.629	95.52	92.451					
3U	93.687	89.972	97.194	90.325				
3L	94.212	90.449	97.554	92.189	98.122			
4U	90.344	89.688	91.326	93.8	89.448	91.31		
4L	93.35	92.016	93.231	95.732	91.102	92.969	96.485	
5U	93.41	94.159	92.233	95.381	90.107	91.971	93.83	
5L	86.987	92.393	85.825	92.365	83.733	85.567	91.253	
6U	93.291	89.958	95.069	94.253	93.328	95.201	94.204	
6L	93.154	89.294	96.769	93.088	97.062	97.953	92.289	
7U	92.551	88.748	94.756	90.292	95.304	96.217	88.844	
7L	91.014	87.027	93.369	91.14	94.109	95.097	90.902	
8U	90.701	89.207	91.99	93.001	90.672	92.537	95.234	
8L	90.887	90.617	91.69	94.695	90.455	92.32	95.179	
9U	91.839	87.845	94.985	92.02	95.547	97.037	93.63	
9L	93.571	90.466	94.169	94.823	92.037	93.907	93.995	

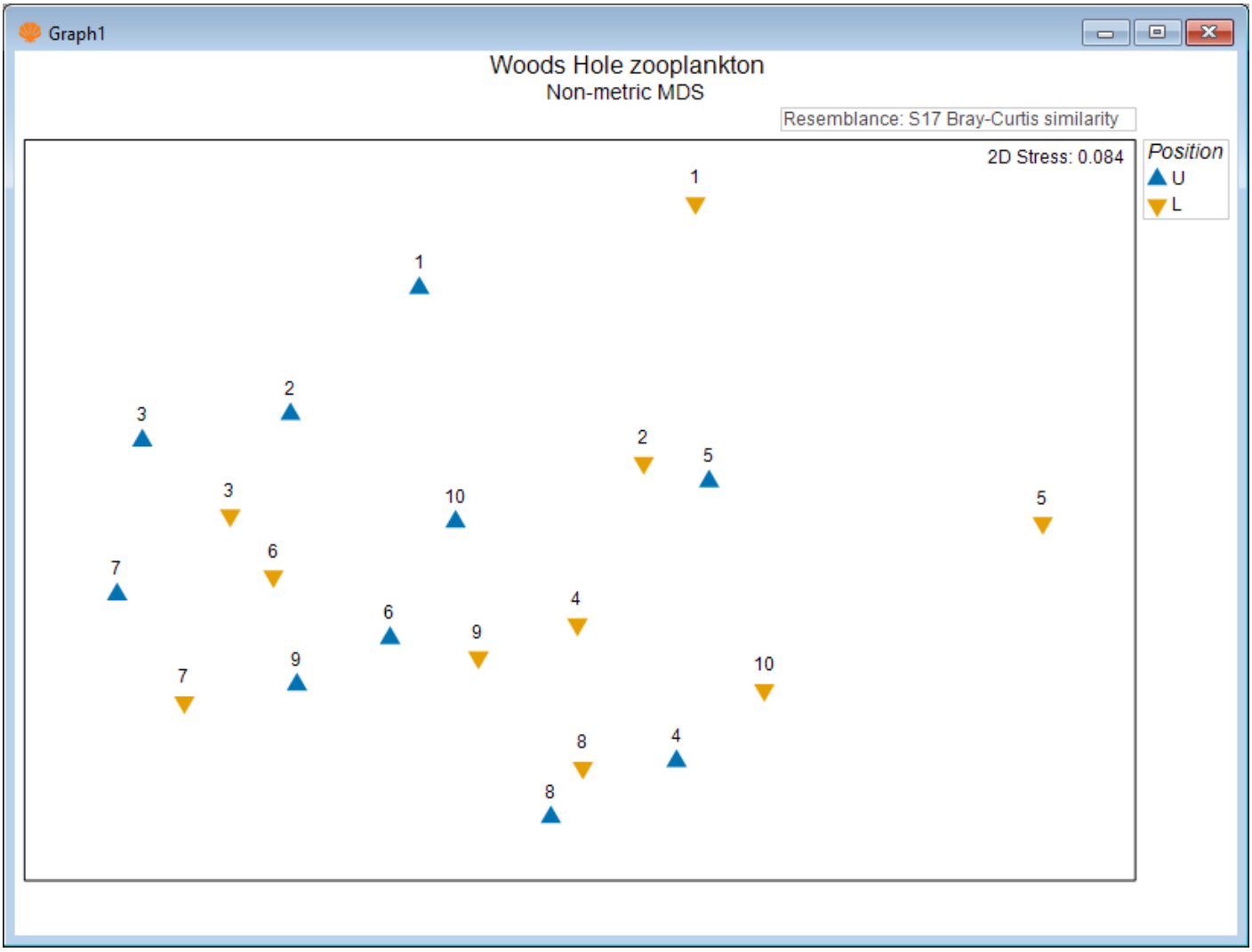
3. Next, create a non-metric MDS plot to visualise the patterns of inter-sample relationships among the plankton communities caught in these nets, based on the Bray-Curtis resemblances. From 'Resem1', click **Analyse** > **MDS** > **Nonmetric MDS (nMDS)...**. Accept all of the defaults in the 'Non Metric MDS' dialog and just click '**OK**'.



4. When you get the multi-plot output, you will see that 'Graph1' in the Explorer tree has the lowest-stress 2D nMDS plot achieved by the routine. From 'Graph1', click **Graph** > **Sample Labels & Symbols...** and choose the Labels to be plotted according to the factor of 'Haul', and the Symbols to be plotted according to the factor of 'Position', like so:



The resulting nMDS plot looks like this:



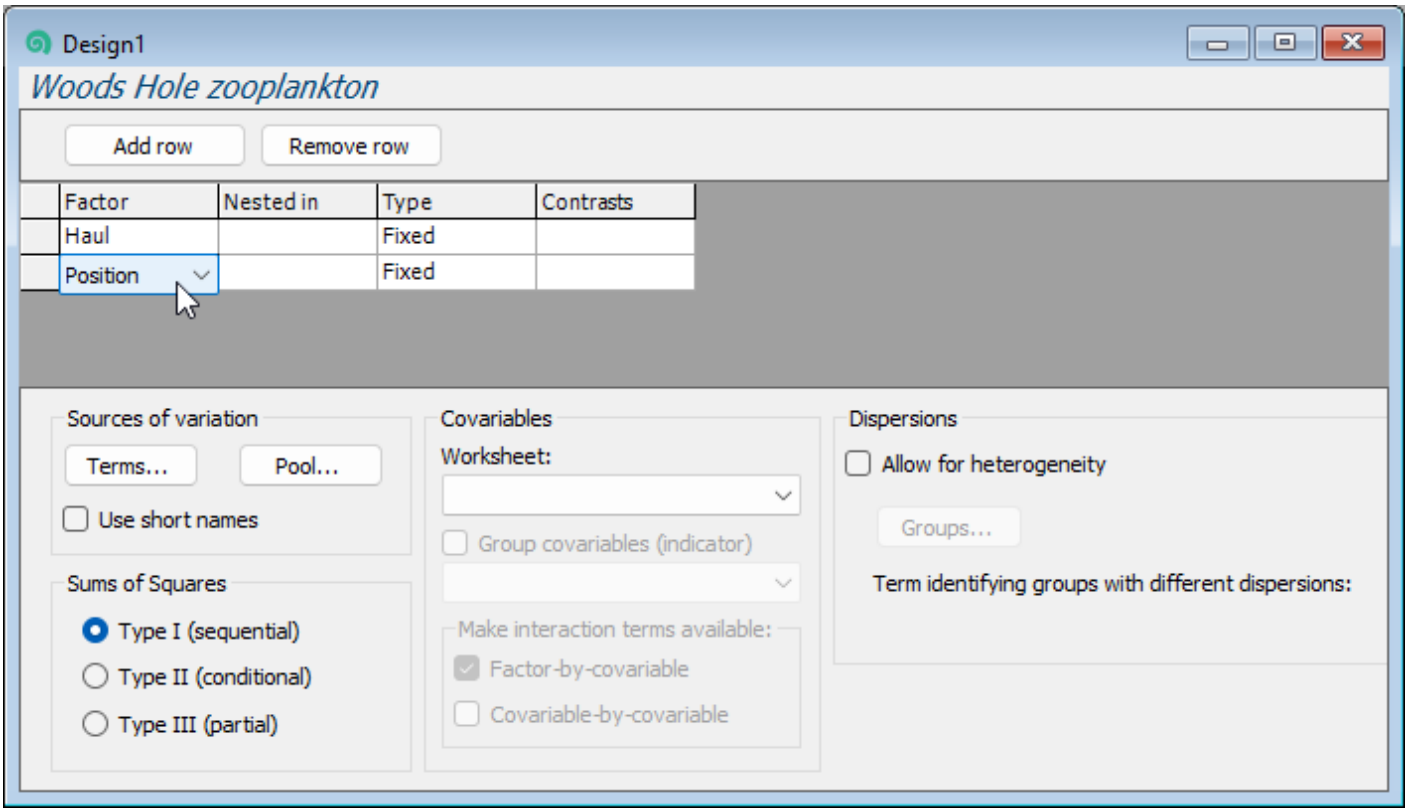
In the above ordination, there do not appear to be any obvious effects on these plankton communities attributable to the factor of 'Position'. Specifically, the symbols corresponding to assemblages caught in Upper ('U') *vs* Lower ('L') nets are scattered throughout the plot and look to be well mixed throughout this 2D MDS space. If anything, individual 'Hauls' seem to be more important in explaining variation here; we can see assemblages associatd with specific hauls (i.e., the two matching numbers, such as the two 7's, the two 3's, the two 8's, etc.) are often fairly close to one another in this nMDS plot.

However, it is possible that differences between the assemblages from one haul to the next are masking genuine differences due to Position, which may be smaller in size and/or may occur in a different direction than is captured by the nMDS ordination plot of reduced dimension. Of course, we should rely upon a two-way PERMANOVA (or ANOSIM)[‡] to formally test for significant effects in either of these factors.[¶]

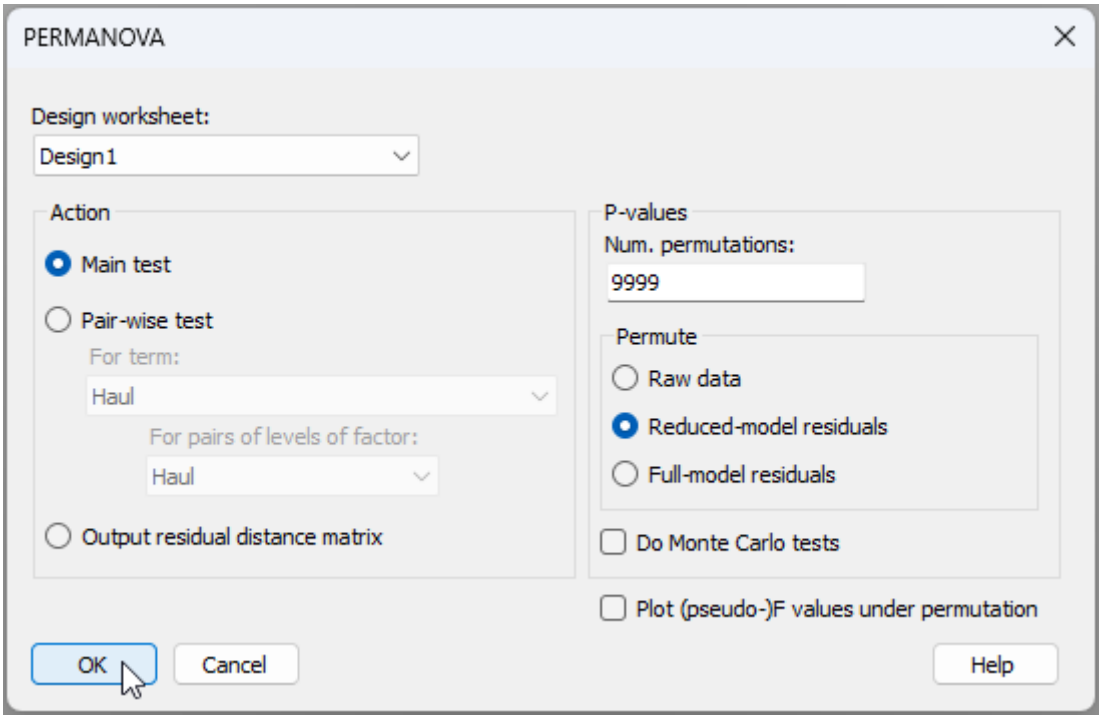
Run a two-way PERMANOVA analysis

5. Create a design file in accordance with the desired two-factor PERMANOVA model. From 'Resem1', click **PERMANOVA+** > **Create PERMANOVA Design...**. In the resulting design file, called 'Design1', add a row (so there are 2 rows in total), then click inside the appropriate cells in the first column to choose the two factors: 'Haul' and 'Position' in rows

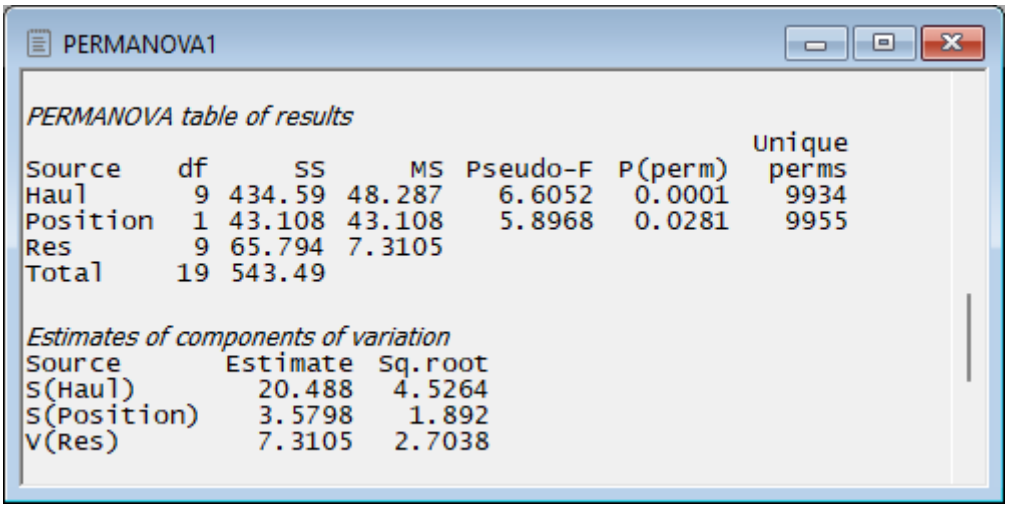
1 and 2, respectively. Also specify the factor of 'Haul' as 'Random' (under 'Type' in column 3), while 'Position' will remain as 'Fixed'. The resulting design file should look like this:



6. Now run the PERMANOVA analysis on the basis of this design. From 'Resem1', click **PERMANOVA+** > **PERMANOVA**, then click '**OK**' to take all the defaults here (the 'Design worksheet' should default to 'Design1' and all else is fine).



The key results in the output file (called 'PERMANOVA1' in the Explorer tree) look like this:



There is significant variation in these plankton assemblages from haul to haul ($F_{9,9}$ = 6.6, P < 0.001). There is also a significant effect caused by the position (relative depth) of the nets ($F_{1,9}$ = 5.9, P < 0.05). The effect size for the factor of 'Position' was much smaller than that for 'Haul' (see the relative sizes of these in the table labeled '*Estimates of components of variation*' in the above output). We can therefore conclude that, in the initial nMDS plot, the significant effects due to the position of the nets are effectively being masked by substantial variation among hauls.

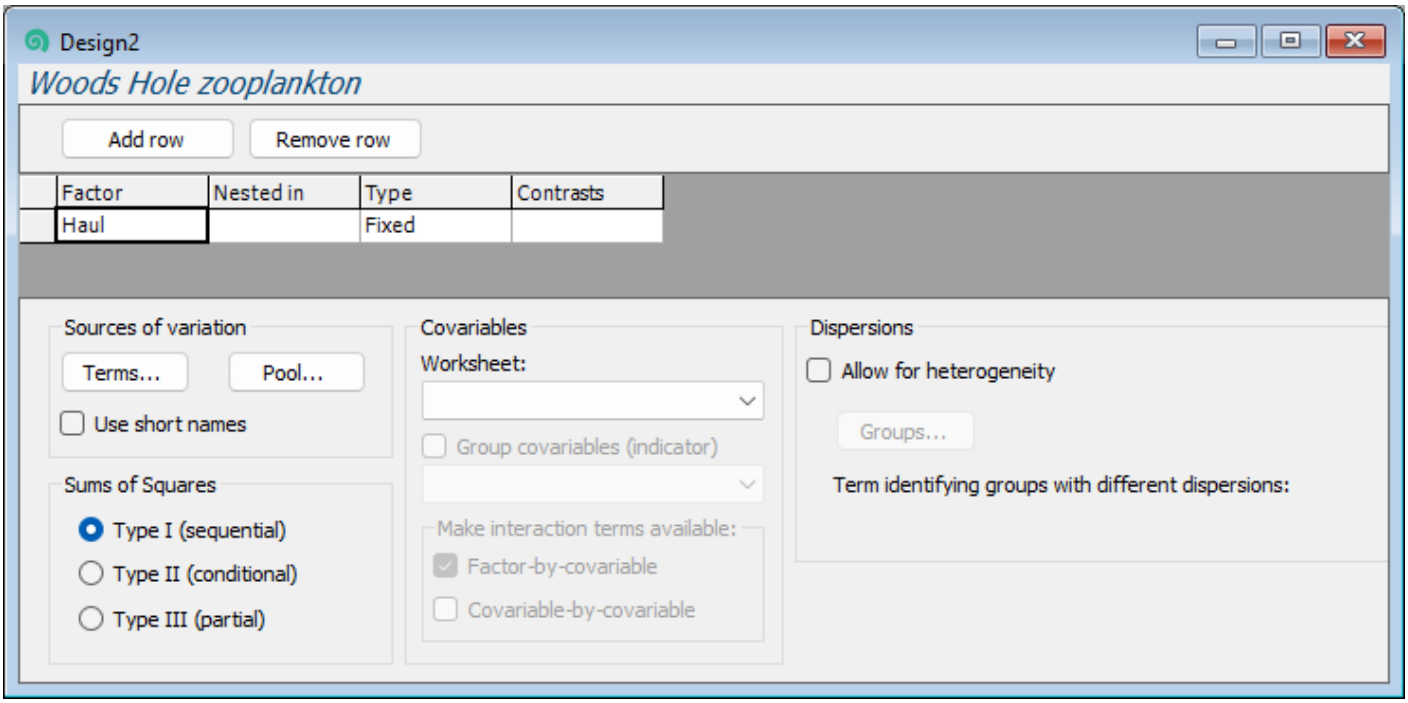
It would be good to *remove the effects of different hauls* so we can visualise effects due to Position. To do this, we need to:

- Fit a PERMANOVA model with one factor only: 'Haul', and choose the option to output a *residual distance matrix*.
- Do a non-metric MDS of the residual distance matrix (with the same symbols as above) to examine patterns with respect to variation in the remaining factor: 'Position'.

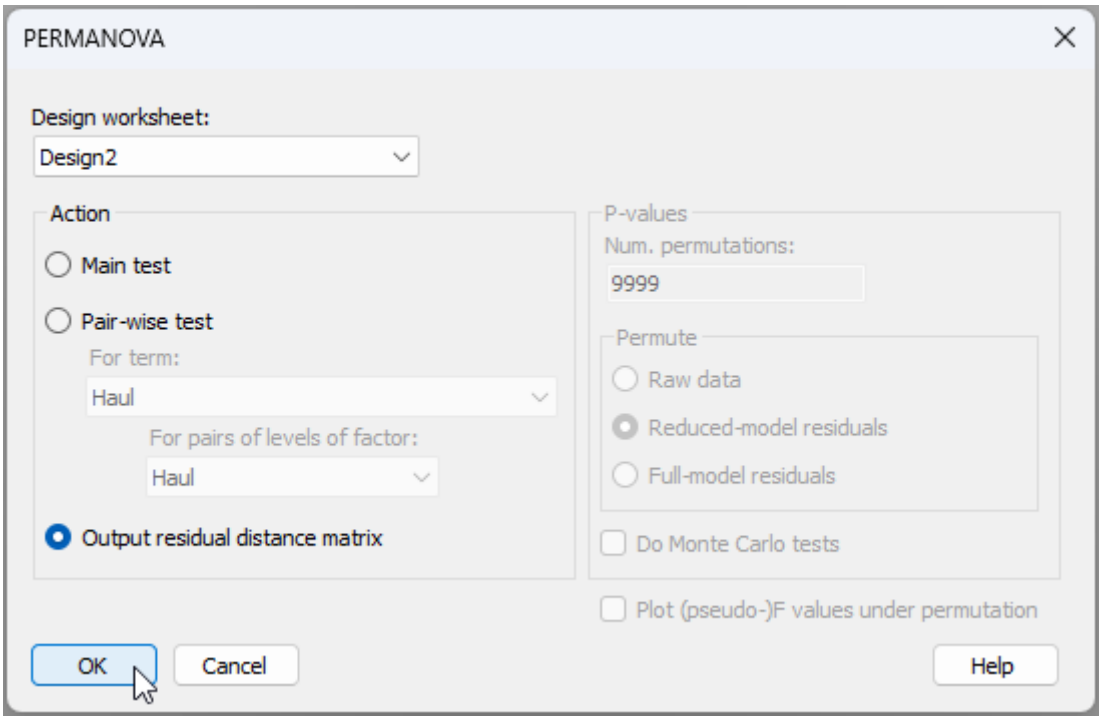
Note that, if the PERMANOVA analysis had been done and no significant effect of 'Position' had been detected, then there would be no particular reason to pursue the idea of 'digging deeper' in an attempt to visualise its effects.

Obtain residual distances and a residual ordination plot

7. To 'remove' the haul effects, we need to fit a one-way PERMANOVA model with the factor of 'Haul' alone. You could just modify the existing design file, but let's make a new design file, for clarity. From 'Resem1', click **PERMANOVA+** > **Create PERMANOVA Design...**, and make the resulting design file (called 'Design2') have just one factor (i.e., 'Haul', the thing we are trying to remove), so that it looks like this:†

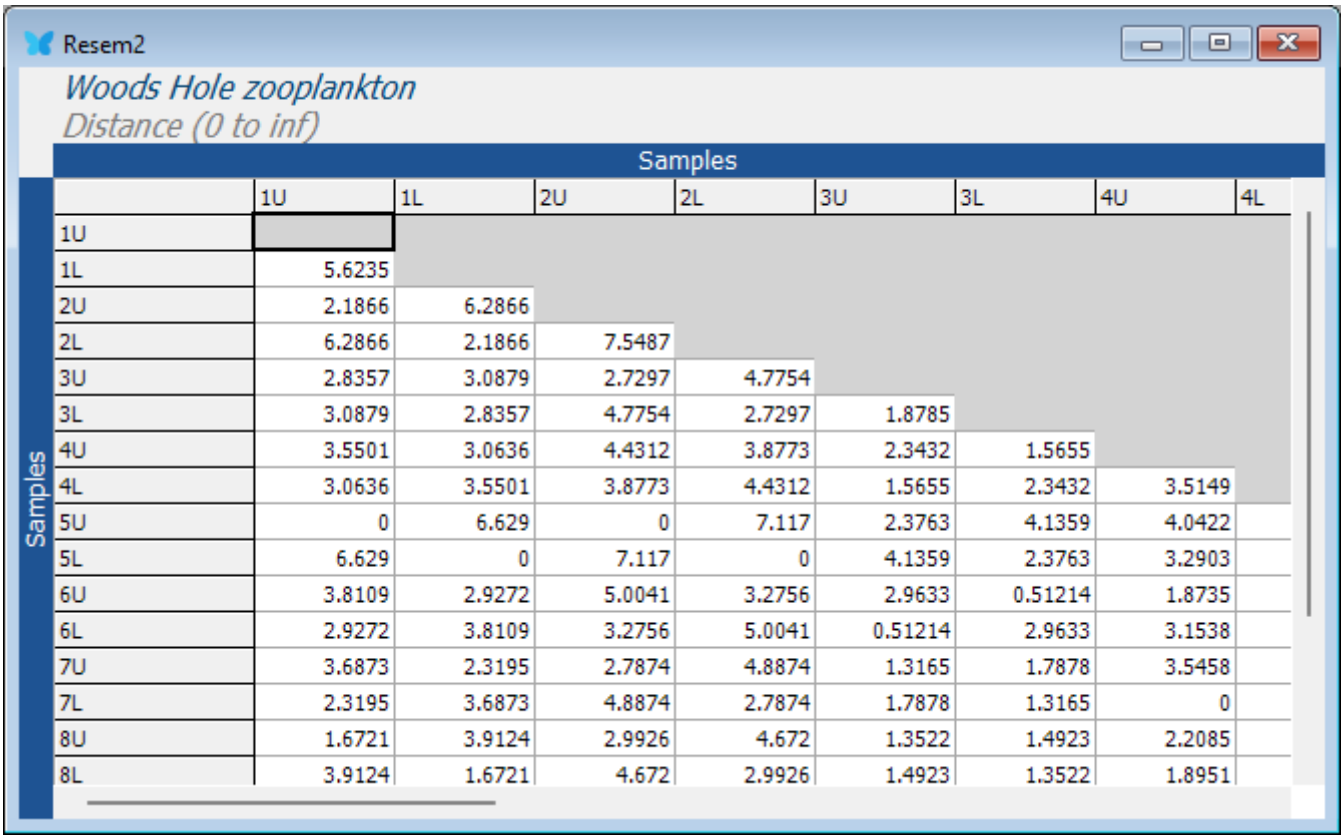


8. Let's get the *residual distance matrix*. From 'Resem1', click **PERMANOVA+** > **PERMANOVA**, choose the 'Design worksheet' as 'Design2', then choose to '$\bullet$Output residual distance matrix', like so:



Notice here, in passing, that the right-hand side of the dialog is 'greyed out' when we choose the option to output a residual distance matrix, as no permutations or tests need to be done in that case. Click '**OK**'.

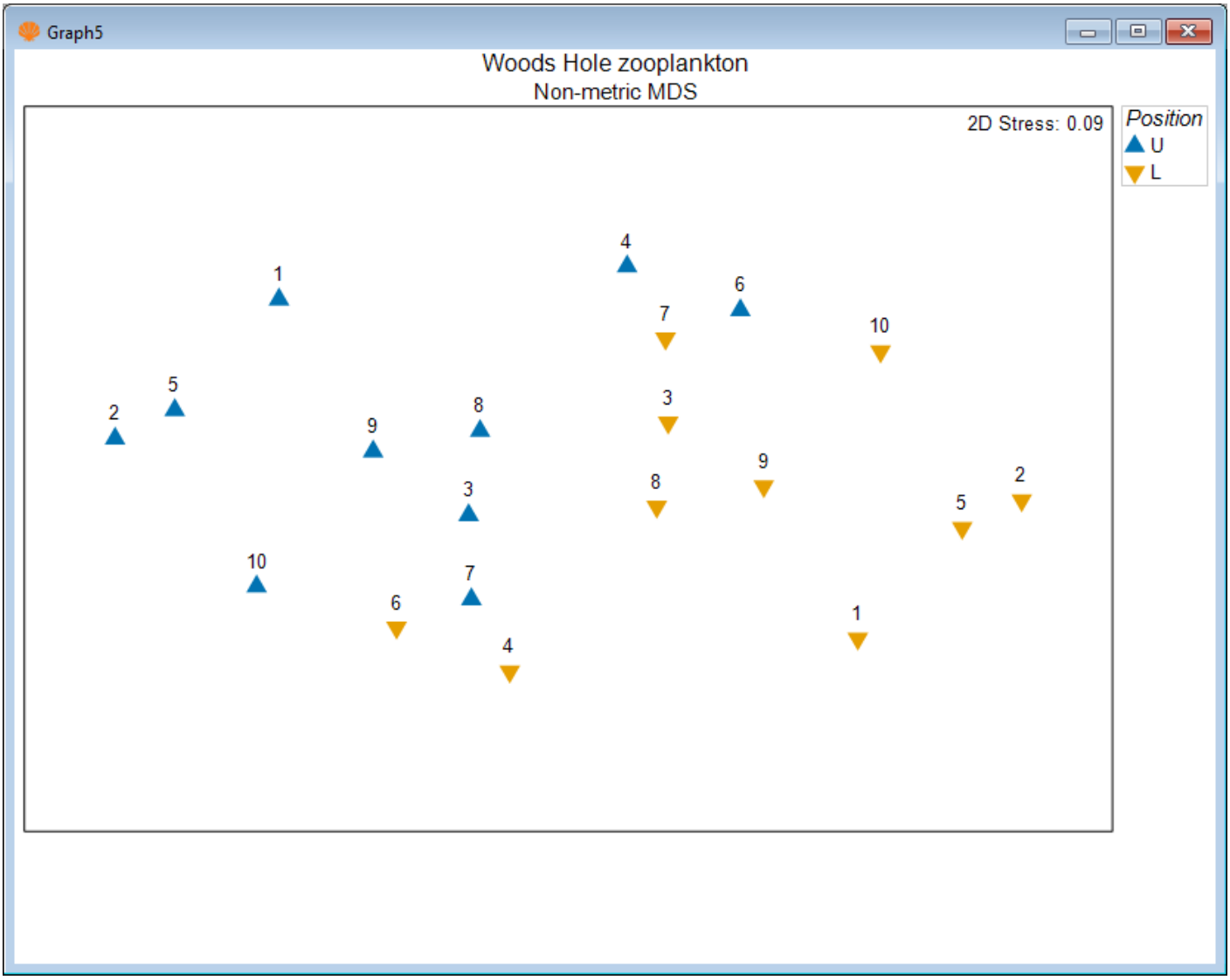
The residual distance matrix is output as 'Resem2' in the Explorer tree, as shown below:



	1U	1L	2U	2L	3U	3L	4U	4L
1U								
1L	5.6235							
2U	2.1866	6.2866						
2L	6.2866	2.1866	7.5487					
3U	2.8357	3.0879	2.7297	4.7754				
3L	3.0879	2.8357	4.7754	2.7297	1.8785			
4U	3.5501	3.0636	4.4312	3.8773	2.3432	1.5655		
4L	3.0636	3.5501	3.8773	4.4312	1.5655	2.3432	3.5149	
5U	0	6.629	0	7.117	2.3763	4.1359	4.0422	
5L	6.629	0	7.117	0	4.1359	2.3763	3.2903	
6U	3.8109	2.9272	5.0041	3.2756	2.9633	0.51214	1.8735	
6L	2.9272	3.8109	3.2756	5.0041	0.51214	2.9633	3.1538	
7U	3.6873	2.3195	2.7874	4.8874	1.3165	1.7878	3.5458	
7L	2.3195	3.6873	4.8874	2.7874	1.7878	1.3165	0	
8U	1.6721	3.9124	2.9926	4.672	1.3522	1.4923	2.2085	
8L	3.9124	1.6721	4.672	2.9926	1.4923	1.3522	1.8951	

9. Finally, we are ready to take a look at our multivariate residuals in an ordination diagram. From 'Resem2', click **Analyse** > **MDS** > **Nonmetric MDS (nMDS)...**. Accept all of the defaults in the 'Non Metric MDS' dialog and just click '**OK**'. Once the default plot has been produced, click **Graph** > **Sample Labels & Symbols...** and choose the Labels to be plotted according to the factor of 'Haul', and the Symbols to be plotted according to the factor of 'Position', just as we had done before.

The resulting lowest-stress 2D nMDS solution for our *residual ordination plot* (having removed effects due to 'Haul') looks like this:



This is very different from the plot we saw before! Now the effects due to the 'Position' are abundantly clear! This might seem like a spot of magic, but if you sneak a peek back at the original nMDS plot (above), you might notice that, for most of the pairs of observations within a given haul, the 'U' (blue) symbol occurs to the left of the 'L' (amber) symbol. Thus, when we center all of the haul pairs (1 through 10) onto a common centroid, the blue symbols appear to the left and the amber symbols appear to the right. Of course, don't forget that the centering (i.e., the calculation of residual distances) is done in the full-dimensional space, and not in the 2d nMDS space. Being able to carefully work out the patterns we see in a residual plot for a given factor by reference to the original plot in this way is not always possible, however. In fact, it will rarely be the case, as typically the effects of factors are happening across a lot more than 2 dimensions. Nevertheless, this example definitely shows us this fantastic tool for visualising minor effects after removing the effects of some other factor(s) or nuisance variable(s).

---

‡*In this particular example, however, ANOSIM cannot be used for the test of Position for the two-way crossed design, because there is no replication within the cells of the design (i.e., within the Position-by-Haul combinations) and the Position factor has fewer than 3 groups.*

¶*This is a case of a two-way unreplicated ANOVA design. The lack of replicate nets at the same depth for each haul means that we will be unable to partition out and estimate any potential Position$\times$Haul interactive effects from the estimated Residual variation, with which it is confounded. Despite this inability to separately test for an interaction, it is still important, nevertheless, to fit a two-way design here that includes the 'Haul' factor, and not to fit a one-way model with only the 'Position' factor alone. Specifically, if we ignore the paired nature of the nets in the sampling protocol, we risk being unable to detect 'Position' effects at all. See the section on unreplicated designs in the original PERMANOVA+ manual for details.*

---

†*In this particular case (i.e., a one-way PERMANOVA), it does not matter whether we specify 'Haul' as being fixed or random; the residuals will be the same either way. We specify it as random here for clarity and consistency in the specification of our reduced (one-factor) model.*

Revision #62
Created 25 June 2025 22:14:06 by Marti
Updated 5 December 2025 04:00:50 by Marti