

6.8 Example: two-way crossed PERMANOVA allowing heterogeneity

We shall look at the diets of $N = 346$ juvenile steelhead / rainbow trout (*Oncorhynchus mykiss*) obtained from 3 different rivers draining into Hood Canal, in the state of Washington, USA.[¶] Some of the fish caught had been reared in a hatchery (identifiable by a clipped fin), and some were wild, of natural origin. Scientists studying these fish wanted to understand more about how the location (i.e., the river system) and the rearing of the fish (hatchery or wild-type) might affect the diets of these juvenile salmonid fish, both in terms of *what* they were eating on average (i.e., centroids) and *how variable* their diets were (i.e., dispersion).

The data file (named 'Hood_Canal_juv_salmonid_diets.pri', found inside the 'Examples_P8 > Hood_Canal_fish' folder) contains abundances of $p = 44$ taxa found in the stomachs of individual fish (obtained by flushing). Note that each row of the file is an individual juvenile fish, and each column is a variable corresponding to items found in the stomachs of these fish (insects of various types, molluscs, etc.). The factor 'Hatchery' indicates whether the fish was hatchery-reared ('H') or wild type ('W'). The factor 'River' indicates which river system each fish was sampled from ('Dewatto', 'Duckabush', or 'Skokomish'). These data are unbalanced, as there is natural uncertainty regarding how many of the fish caught would end up being of hatchery origin. The number of replicates in each cell of the 2-factor crossed design is shown in Table 6.1 below.

Table. 6.1. Sample sizes (n_{ij}) in each of the $3 \times 2 = 6$ cells of the 2-factor crossed study design examining salmonid diets from Hood canal.

	Dewatto	Duckabush	Skokomish	Total
Wild	61	114	103	278
Hatchery-reared	35	22	11	68
Total	96	136	114	346

Pre-treat data, then calculate resemblances

1. **Open data in PRIMER** - Start by opening the file in PRIMER by clicking **File > Open...** and navigating to the file named 'Hood_Canal_juv_salmonid_diets.pri'.

PRIMER

File Edit Select View Wizards Pre-treatment Analyse Plots PERMANOVA+ Tools Window Help

Workspace

Hood_Canal_juv_salmonid_diets

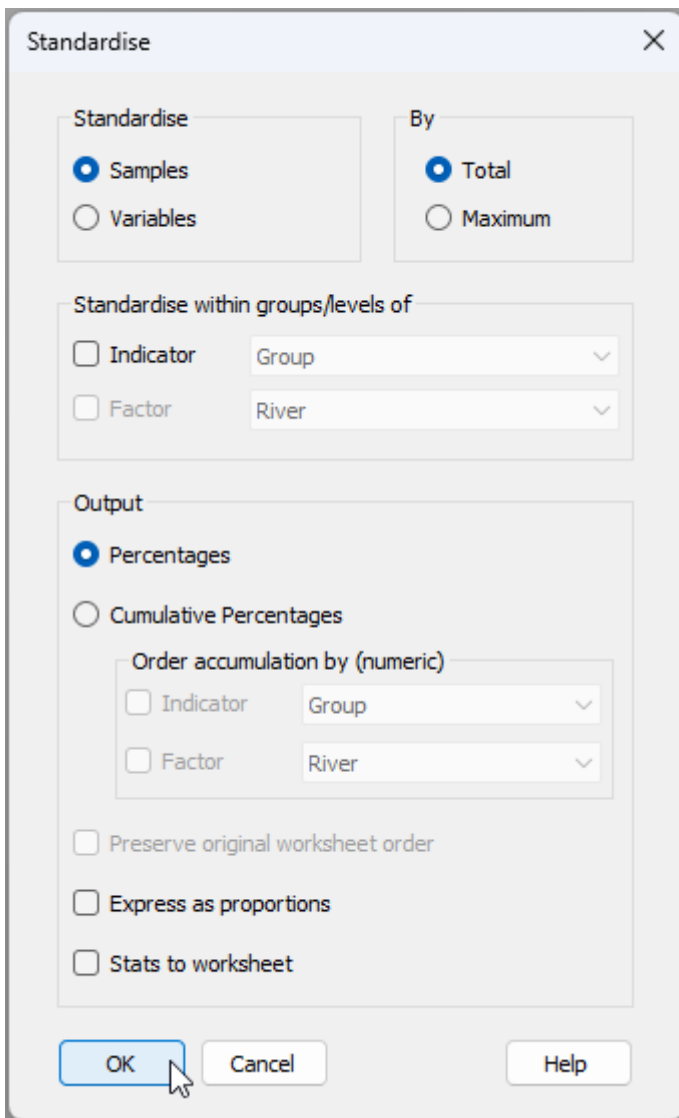
Hood Canal juvenile salmonid diets

Abundance

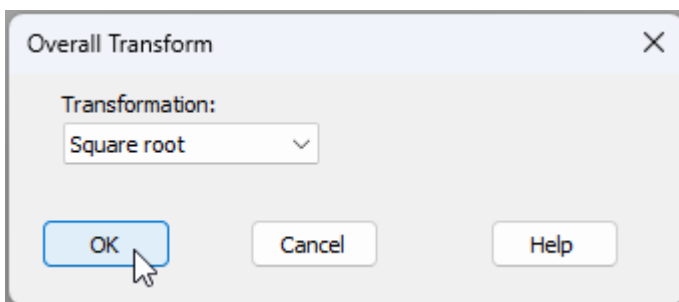
	Variables							
	Arachnida: A	Arthropoda: A	Arthropoda: T	Arthropoda: T	Coleoptera A	Coleoptera A	Coleoptera T	Coleoptera T
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	1	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0
13	0	1	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0
16	1	0	0	1	0	0	0	0
17	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0
19	2	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0

Samples - Individual fish

2. **Standardise data** - We first need to standardise the data by total sample abundances (rows) to account for the fact that individual fish are different sizes and naturally would have had a different volume of prey in their stomachs at the time each one was caught. Click **Pre-treatment > Standardise...** > (Standardise: $\bullet$ Samples) & (By: $\bullet$ Total) & (Output: $\bullet$ Percentages), then click 'OK'. The resulting sheet of standardised data will be called 'Data1'.



3. **Transform data** - From the standardised data ('Data1' in the Explorer tree), apply a square-root transformation by clicking **Pre-treatment > Transform(overall)...**. Choose Transformation: **Square root**, then click '**OK**'. This will create a data sheet of standardised and transformed data called 'Data2'.



4. **Calculate resemblances** - From the standardised and transformed data ('Data2' in the Explorer tree), calculate Bray-Curtis resemblances among all pairs of individual fish by clicking **Analyse > Resemblance...**, then choose (Measure $\bullet$ Bray-Curtis similarity) & (Analyse between $\bullet$ Samples), and click '**OK**'.

Resemblance

Measure

☒ Bray-Curtis similarity
 ☐ Euclidean distance
 ☐ Index of association
 ☐ Other

☒ Similarity
 ☒ Distance/dissimilarity
 ☐ Quantitative
 ☐ P/A
 ☐ Correlation
 ☐ Taxonomic P/A

 (exc0-0 = excluding joint absences)

S1 Simple matching

Analyse between

☒ Samples
 ☐ Variables

☐ Add dummy variable
 Value:

1

OK

Cancel

Help

The resulting Bray-Curtis resemblances will be shown in the item named 'Resem1' in the Explorer tree, as shown below:

PRIMER

File Edit Select View Analyse PERMANOVA+ Tools Window Help

Workspace

Hood_Canal_juv_salmonid_diets

Standardise1

Data1

Overall Transform1

Data2

Resemblance

Resem1

Resem1

Hood Canal juvenile salmonid diets

Similarity (0 to 100)

Samples - individual fish

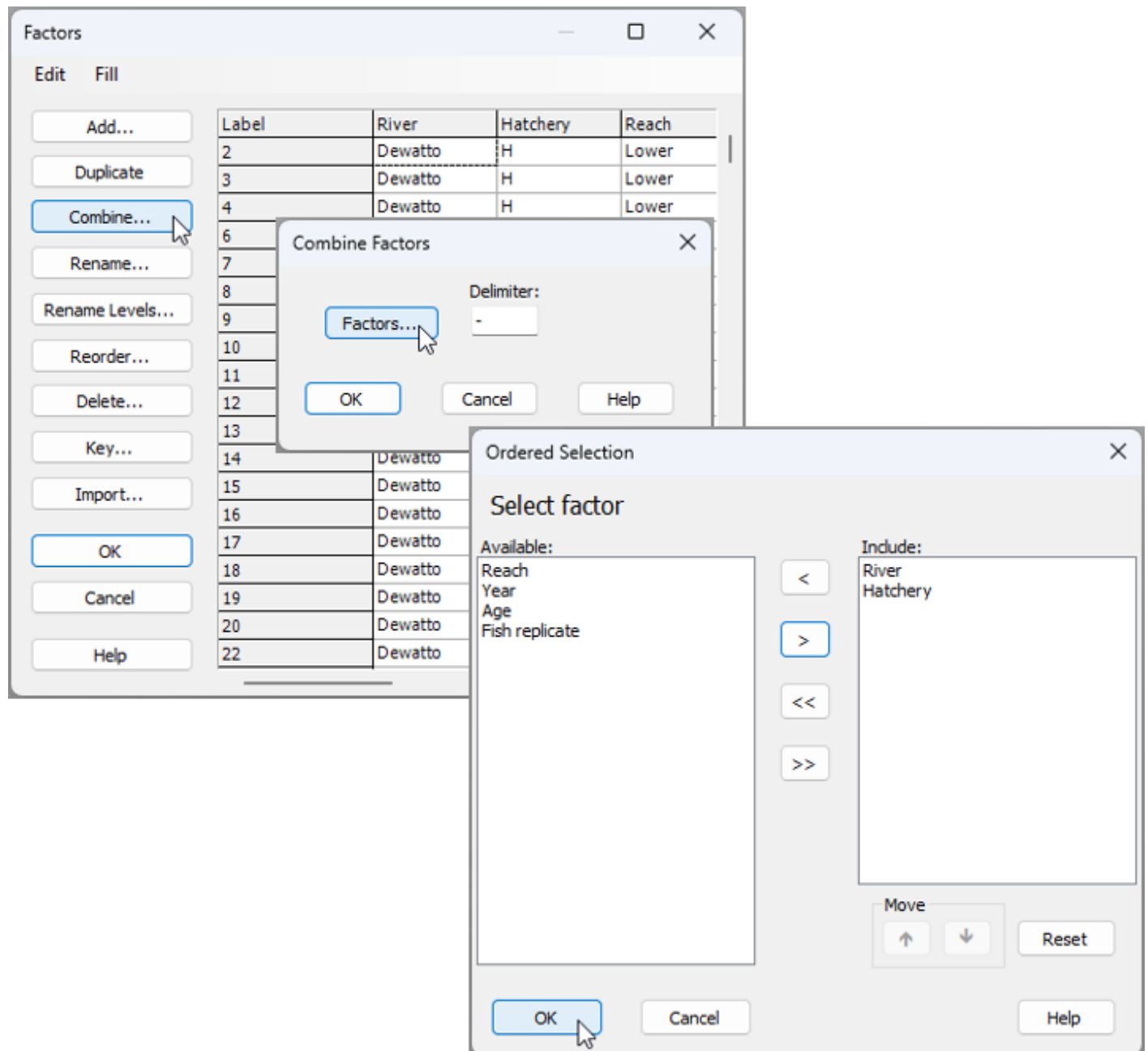
	2	3	4	6	7	8	9	10	11
2									
3	39.448								
4	91.618	40.46							
6	0	0	0						
7	40.926	52.031	41.941	0					
8	32.537	49.336	31.98	20.227	14.738				
9	86.66	41.121	77.902	0	42.804	34.718			
10	33.169	35.286	32.579	0	18.204	56.867	35.488		
11	34.685	21.132	39.239	0	0	32.947	21.244	27.379	
12	14.89	16.016	14.581	0	16.622	31.517	16.124	49.063	
13	21.397	22.879	20.987	0	23.67	30.631	23.021	52.811	
14	39.448	50	40.46	0	52.031	34.529	41.121	57.094	
15	55.11	34.68	53.985	0	35.968	51.269	59.596	72.332	
16	26.283	27.85	25.844	0	48.951	27.182	27.999	30.216	
17	0	58.579	0	0	0	39.401	0	19.948	
18	40.687	51.702	41.702	0	69.555	35.411	42.531	41.498	
19	29.849	32.038	29.246	0	58.319	32.02	32.248	52.782	
20	63.279	0	58.757	0	0	38.109	52.046	42.795	
22	53.939	26.555	53.022	0	27.357	39.956	57.528	50.628	
23	27.261	18.47	24.773	0	10.018	27.74	18.560	45.261	

Row 1 Col 1

Test for homogeneity of dispersions across the cells of the 2-way design

Let's test for heterogeneity of dispersions among the cells in this 2-way crossed design.

5. **Create a combined factor** - First, create a combined factor consisting of all combinations of 'River' and 'Hatchery'. Click **Edit > Factors**, click on the button labeled **Combine...**, then click on the **'Factors...'** button. Next, click on each of the factors of 'River' and 'Hatchery' in turn (shown in the 'Available:' column on the left), then click on '>' to move them over to the 'Include:' column on the right.



You will need to click '**OK**' in both the 'Ordered Selection' dialog box, and also the 'Combine Factors' dialog box (as shown above), which will return you to the 'Factors' dialog box, where, if you scroll to the right, you should now see your newly created combined factor, called '**River-Hatchery**' (see below):

Factors

Edit Fill

Add... Duplicate Combine... Rename... Rename Levels... Reorder... Delete... Key... Import... OK Cancel Help

Label	Age	Fish replicate	River-Hatchery
2	NA	14	Dewatto-H
3	1	15	Dewatto-H
4	NA	16	Dewatto-H
6	2	20	Dewatto-H
7	NA	21	Dewatto-H
8	NA	23	Dewatto-H
9	2	25	Dewatto-H
10	1	28	Dewatto-H
11	2	32	Dewatto-H
12	NA	33	Dewatto-H
13	2	37	Dewatto-H
14	2	4	Dewatto-H
15	1	45	Dewatto-H
16	NA	48	Dewatto-H
17	NA	5	Dewatto-H
18	NA	6	Dewatto-H
19	NA	8	Dewatto-H
20	2	9	Dewatto-H
22	1	11	Dewatto-W

Click 'OK' and you are ready for the next step.

- Do a PERMDISP test** - Let's do the test for homogeneity of multivariate dispersions across the cells corresponding to this combined factor. From the 'Resem1' matrix, click **PERMANOVA+ > PERMDISP** > (Group factor: River-Hatchery) & (P-values are from \$\bullet\$Permutation) & (\$\checkmark\$Do pairwise tests) & (\$\checkmark\$Output individual deviation values to worksheet).

PERMDISP

Group factor: River-Hatchery

Num. permutations: 9999

P-values are from

☒ Permutation

☐ Tables

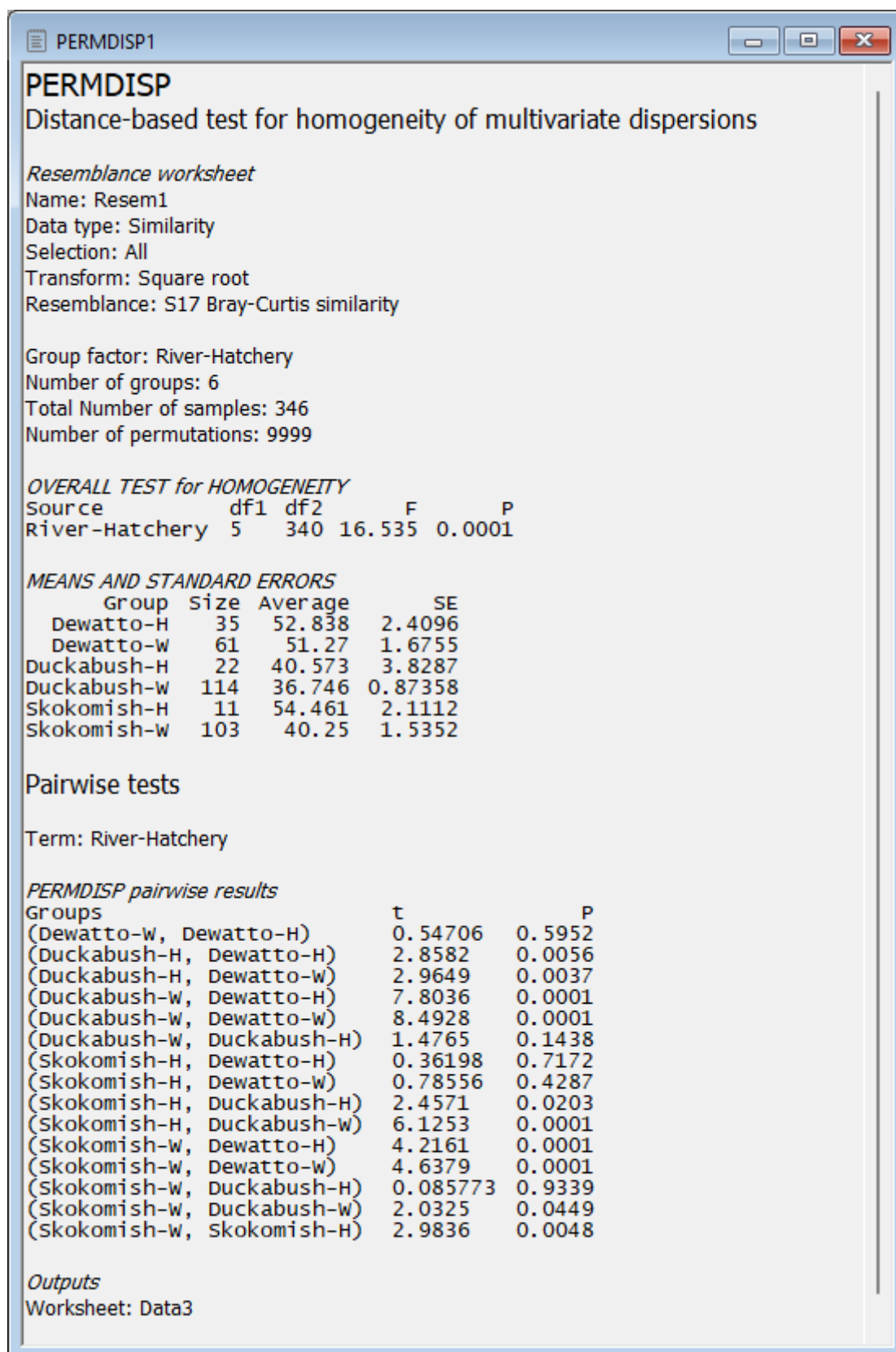
☒ Do pairwise tests

☒ Output individual deviation values to worksheet

☐ Output mean deviations to worksheet

OK Cancel Help

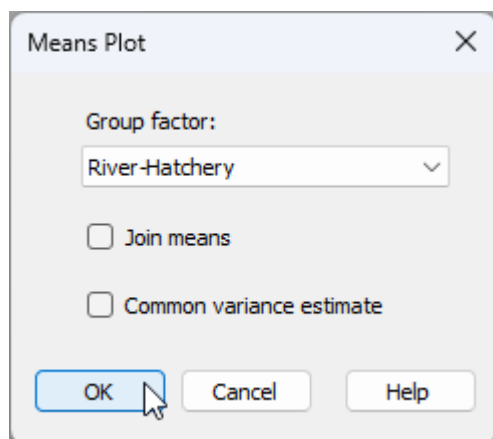
The output shows a highly significant result ($F_{5,340} = 16.535$, $P = 0.0001$), indicating there are highly significant differences in dispersion (variability in the diets) across these 6 cells.



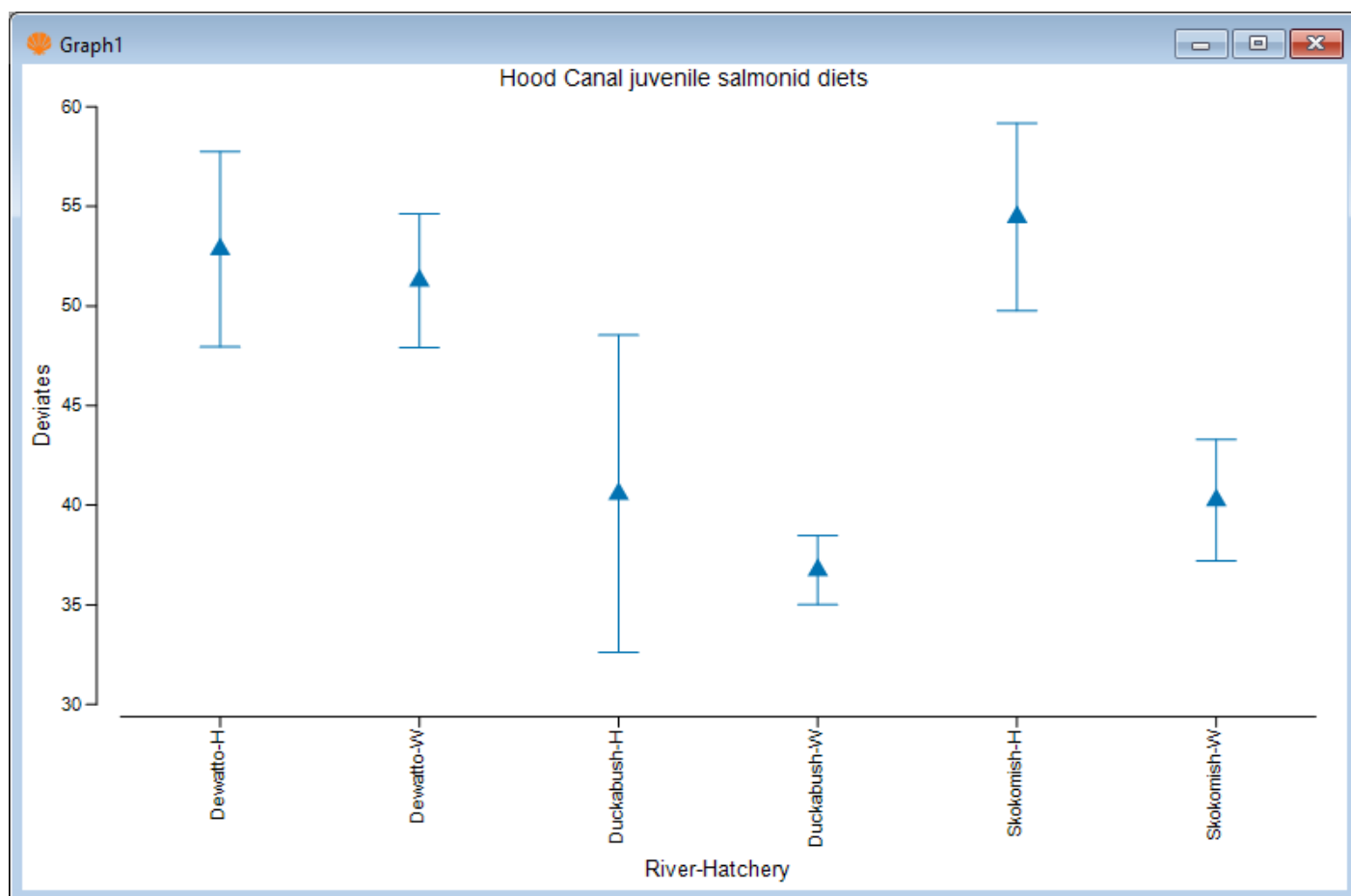
The mean distance-to-centroid values also show that for each river system, there is apparently greater dispersion (variability in diets) for the fish that were hatchery-reared, compared to the wild-type fish, on average; however, not all pairwise comparisons were statistically significant in this regard (e.g., for the Dewatto comparison of 'H' vs 'W', $t = 0.547$ and $P > 0.50$). Also provided

in the output are the individual deviation values; that is, the distance to the cell centroid for each salmonid in the Bray-Curtis space. These are provided in the item named 'Data3' of the Explorer tree.

7. **Examine dispersion differences among cells** - Let's create a *means plot* of the deviations from cell centroids as a useful way to visualise differences in dispersions (on average) across the cells. From the data sheet 'Data3', click **Plots > Means Plot... >** (Group factor: **River-Hatchery**) and untick the box in front of the words '\$\square\$Join means', then click '**OK**'.



You should see the following plot ('Graph1'):



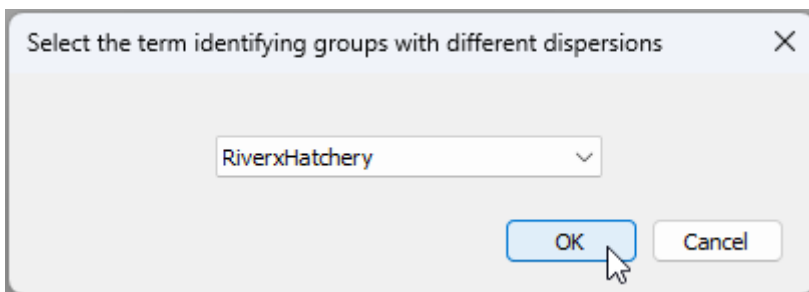
This plot suggests that the differences in the dispersions between the hatchery-reared and wild-type fish (with hatchery-reared fish being, on average, more variable in their diets) was greatest in the Skokomish river.

Test for equality of centroids, allowing for heterogeneous dispersions among cells.

Heterogeneity in dispersions among the cells of the study design is very clear and significant, and the number of replicates per cell is quite unbalanced; thus, we should run a two-way PERMANOVA examining the potential effects of river and hatchery-rearing on salmonid diets *allowing for these differences in dispersions*. We shall create an appropriate design file, then run the PERMANOVA analysis.

8. **Create a design file** - From the 'Resem1' matrix, click **PERMANOVA+ > Create PERMANOVA Design...** to create an appropriate design file, as follows:

- Add a row by clicking on the button that says 'Add row' (), so there will be two rows in the design file, one for each factor in the design.
- In the first row, click in the blank cell in the first column and choose the factor of 'River', then in the second row, choose the factor of 'Hatchery'. These are both fixed and are crossed with one another, so no further changes to the design file are needed.
- Under the word 'Dispersions', tick the box that says '\$\checkmark\$Allow for heterogeneity' and click on the 'Groups' button. In the resulting pop-up box that says 'Select the term identifying groups with different dispersions' choose 'RiverxHatchery', then click **OK**.



The resulting design file (called 'Design1') will look like this:

Design1

Hood Canal juvenile salmonid diets

Add row Remove row

Factor	Nested in	Type	Contrasts
River		Fixed	
Hatchery		Fixed	

Sources of variation

Terms... Pool...

☐ Use short names

Sums of Squares

☒ Type I (sequential)

☐ Type II (conditional)

☐ Type III (partial)

Covariables

Worksheet:

☐ Group covariables (indicator)

Make interaction terms available:

☒ Factor-by-covariable

☐ Covariable-by-covariable

Dispersions

☒ Allow for heterogeneity

Groups...

Term identifying groups with different dispersions:

RiverxHatchery

9. **Run a 2-way PERMANOVA (main test)** - Now that we have the design file, we can run the PERMANOVA analysis. From the 'Resem1' matrix, click **PERMANOVA+** > **PERMANOVA....** Choose (Design worksheet: **Design1**) & (Action: $\bullet$ Main test), with everything else in the dialog being left as the defaults, then click '**OK**'.

PERMANOVA

Design worksheet:

Design1

Action

☒ Main test

☐ Pair-wise test

For term:

River

For pairs of levels of factor:

River

☐ Output residual distance matrix

P-values

Num. permutations:

9999

Permute

☐ Raw data

☒ Reduced-model residuals

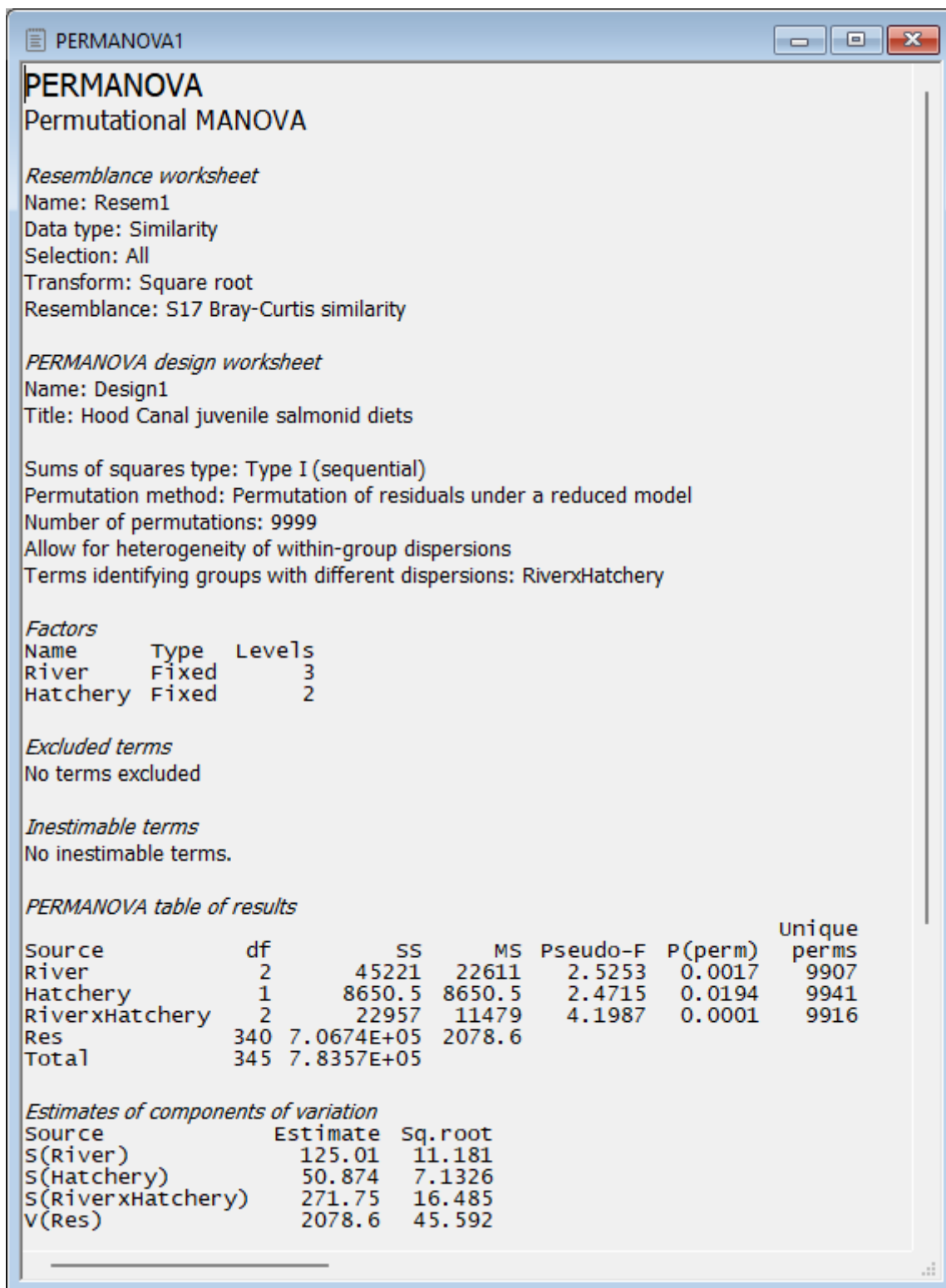
☐ Full-model residuals

☐ Do Monte Carlo tests

☐ Plot (pseudo-)F values under permutation

OK Cancel Help

The results of the analysis (in the item called 'PERMANOVA1' in the Explorer tree), show that there is a statistically significant interaction between the two factors of River and Hatchery in their effects on salmonid diets ($F_{2,44.21} = 4.20$, $P = 0.0001$).



A casual glance at the expectations of the means squares and the construction of the F statistics for each term in the model shows the complexity underlying these tests performed by the PERMANOVA routine.

PERMANOVA table of results						
Source	df	SS	MS	Pseudo-F	P(perm)	Unique perms
River	2	45221	22611	2.5253	0.0017	9907
Hatchery	1	8650.5	8650.5	2.4715	0.0194	9941
RiverxHatchery	2	22957	11479	4.1987	0.0001	9916
Res	340	7.0674E+05	2078.6			
Total	343	7.8357E+05				

Estimates of components of variation		
Source	Estimate	Sq.root
S(River)	125.01	11.181
S(Hatchery)	50.874	7.1326
S(RiverxHatchery)	271.75	16.485
V(Res)	2078.6	45.592

Details of the expected mean squares (EMS) for the model

Source	EMS
River	$0.30291 \cdot V(\text{Res}(6)) + 0.03235 \cdot V(\text{Res}(5)) + 0.25438 \cdot V(\text{Res}(4)) + 0.04909 \cdot V(\text{Res}(3)) + 0.22956 \cdot V(\text{Res}(2)) + 0.13171 \cdot V(\text{Res}(1)) + 21.822 \cdot S(\text{RiverxHatchery}) + 4.0165 \cdot S(\text{Hatchery}) + 114.17 \cdot S(\text{River})$
Hatchery	$0.018945 \cdot V(\text{Res}(6)) + 0.17739 \cdot V(\text{Res}(5)) + 0.058933 \cdot V(\text{Res}(4)) + 0.30538 \cdot V(\text{Res}(3)) + 0.16018 \cdot V(\text{Res}(2)) + 0.27917 \cdot V(\text{Res}(1)) + 3.135 \cdot S(\text{RiverxHatchery}) + 101.24 \cdot S(\text{Hatchery})$
RiverxHatchery	$0.038773 \cdot V(\text{Res}(6)) + 0.36306 \cdot V(\text{Res}(5)) + 0.051416 \cdot V(\text{Res}(4)) + 0.26643 \cdot V(\text{Res}(3)) + 0.1022 \cdot V(\text{Res}(2)) + 0.17812 \cdot V(\text{Res}(1)) + 32.179 \cdot S(\text{RiverxHatchery})$

Construction of Pseudo-F ratio(s) from mean squares

Source	Numerator	Denominator	Num. df
River	$0.21949 \cdot \text{Res}(5) + 0.14267 \cdot \text{Res}(3) + 1 \cdot \text{River}$	$0.27601 \cdot \text{Res}(6) + 0.21737 \cdot \text{Res}(4) + 0.15429 \cdot \text{Res}(2) + 0.00053132 \cdot \text{Res}(1) + 0.67429 \cdot \text{RiverxHatchery} + 0.039673 \cdot \text{Hatchery}$	2.18
Hatchery	$1 \cdot \text{Hatchery}$	$0.015168 \cdot \text{Res}(6) + 0.14202 \cdot \text{Res}(5) + 0.053923 \cdot \text{Res}(4) + 0.27942 \cdot \text{Res}(3) + 0.15022 \cdot \text{Res}(2) + 0.26182 \cdot \text{Res}(1) + 0.097424 \cdot \text{RiverxHatchery}$	1
RiverxHatchery	$1 \cdot \text{RiverxHatchery}$	$0.038773 \cdot \text{Res}(6) + 0.36306 \cdot \text{Res}(5) + 0.051416 \cdot \text{Res}(4) + 0.26643 \cdot \text{Res}(3) + 0.1022 \cdot \text{Res}(2) + 0.17812 \cdot \text{Res}(1)$	2

Source	Den. df
River	2.91
Hatchery	17.88
RiverxHatchery	44.21

It should also be noted, in passing, that we have used Type I SS here (sequential tests) and, as our sample sizes are unbalanced, the order in which we have chosen to fit these terms *will matter* to the results. We leave it up to the reader to consider re-running the analysis after changing the order of the factors (if desired) and/or to fit the PERMANOVA model using a different choice of sums of squares for the partitioning.

10. Pair-wise tests - A natural next step, having observed a significant interaction, is to do pair-wise comparisons. We can consider two different sets of pair-wise comparisons that would each be of interest:

- (i) compare the diets for hatchery-reared fish vs wild-type fish *separately* within each river; and/or
- (ii) compare the diets for fish caught in different rivers (there will be three tests here: one for every pair of rivers) *separately* for each of the hatchery-reared fish and the wild-type fish.

The pair-wise tests can be done in a way that also allows for heterogeneity in dispersions. To pursue (i), from the resemblance matrix ('Resem1'), click **PERMANOVA+ > PERMANOVA....** Choose (Design worksheet: **Design1**) & (Action: $\bullet$ Pair-wise test > For term: **RiverxHatchery** > For pairs of levels of factor: **Hatchery**), with everything else being left as the defaults, then click '**OK**'.

PERMANOVA

Design worksheet:
Design1

Action

☐ Main test

☒ Pair-wise test

For term:
RiverxHatchery

For pairs of levels of factor:
Hatchery

☐ Output residual distance matrix

P-values

Num. permutations:
9999

Permute

☐ Raw data

☒ Reduced-model residuals

☐ Full-model residuals

☐ Do Monte Carlo tests

☐ Plot (pseudo-)F values under permutation

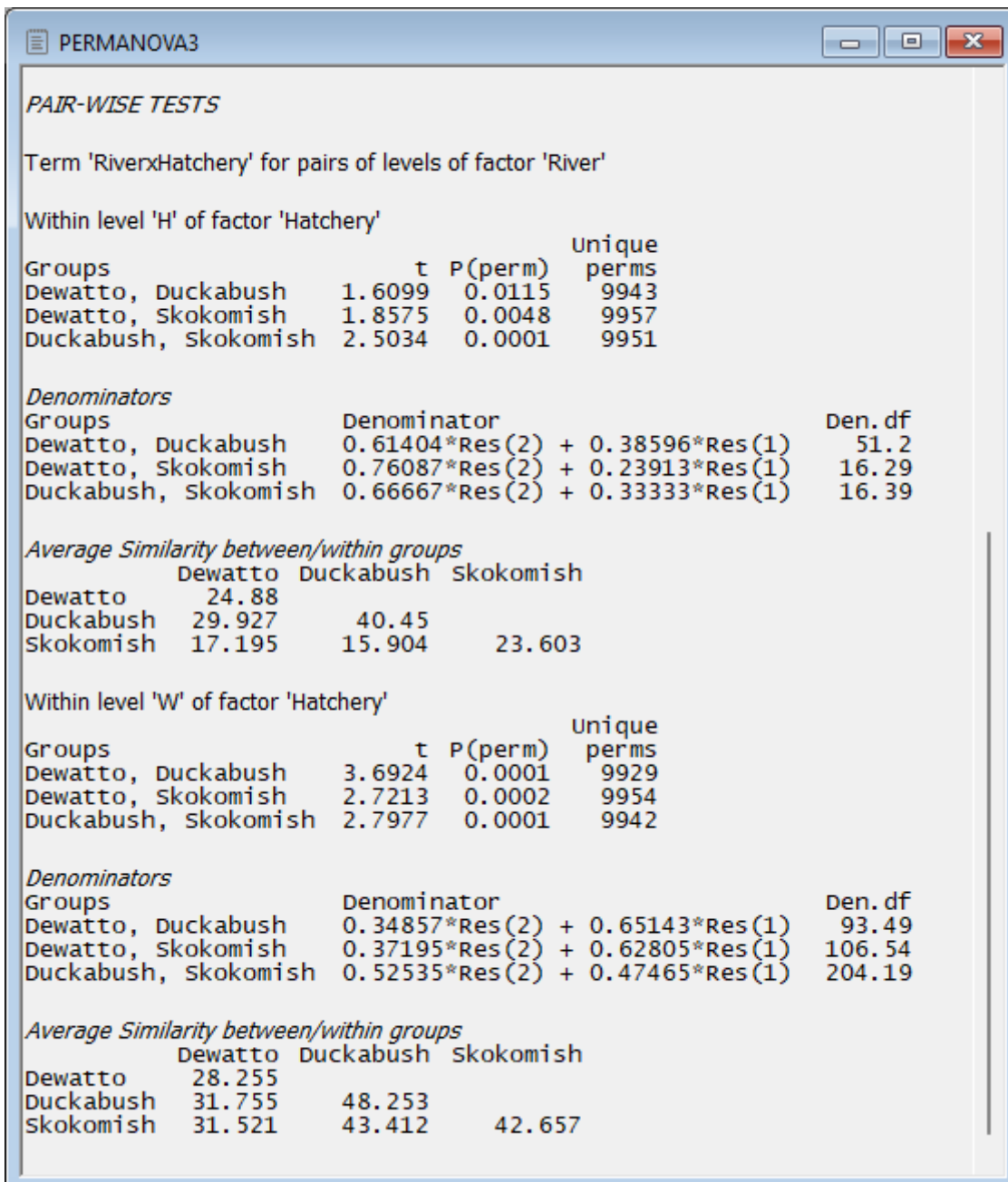
OK Cancel Help

The results of these tests (i) are shown below in the output file 'PERMANOVA2'.

PERMANOVA2				
PAIR-WISE TESTS				
Term 'RiverxHatchery' for pairs of levels of factor 'Hatchery'				
Within level 'Dewatto' of factor 'River'				
Groups	t	P(perm)	Unique perms	
H, w	1.1067	0.2828	9933	
Denominators				
Groups	Denominator		Den. df	
H, w	0.36458*Res(2) + 0.63542*Res(1)		68.67	
Average Similarity between/within groups				
	H	W		
H	24.88			
w	26.358	28.255		
Within level 'Duckabush' of factor 'River'				
Groups	t	P(perm)	Unique perms	
H, w	1.6553	0.0146	9950	
Denominators				
Groups	Denominator		Den. df	
H, w	0.16176*Res(2) + 0.83824*Res(1)		27.04	
Average Similarity between/within groups				
	H	W		
H	40.45			
w	42.355	48.253		
Within level 'Skokomish' of factor 'River'				
Groups	t	P(perm)	Unique perms	
H, w	2.6686	0.0008	9945	
Denominators				
Groups	Denominator		Den. df	
H, w	0.096491*Res(2) + 0.90351*Res(1)		11.24	
Average Similarity between/within groups				
	H	W		
H	23.603			
w	17.166	42.657		

They show that, although there were no significant differences in the diets of hatchery-reared vs wild-type fish for the Dewatto river ($t_{68.67} = 1.107$, $P > 0.25$), there were differences between them detected in both the Duckabush ($t_{27.04} = 1.655$, $P < 0.02$) and Skokomish ($t_{11.24} = 2.67$, $P < 0.001$) river systems.

We can also do a set of tests comparing diets of fish in different rivers (ii) by repeating this procedure in precisely the same way, but choosing '(Action: $\bullet$ Pair-wise test > For term: RiverxHatchery > For pairs of levels of factor: River)' in the PERMANOVA dialog instead (all else remaining the same). The results of those analyses are shown below ('PERMANOVA3'):



From this, we see that the diets of fish caught in each of the three rivers systems differed significantly from one another, whether they were hatchery-reared or wild-type fish (all pairwise tests had $P < 0.015$).

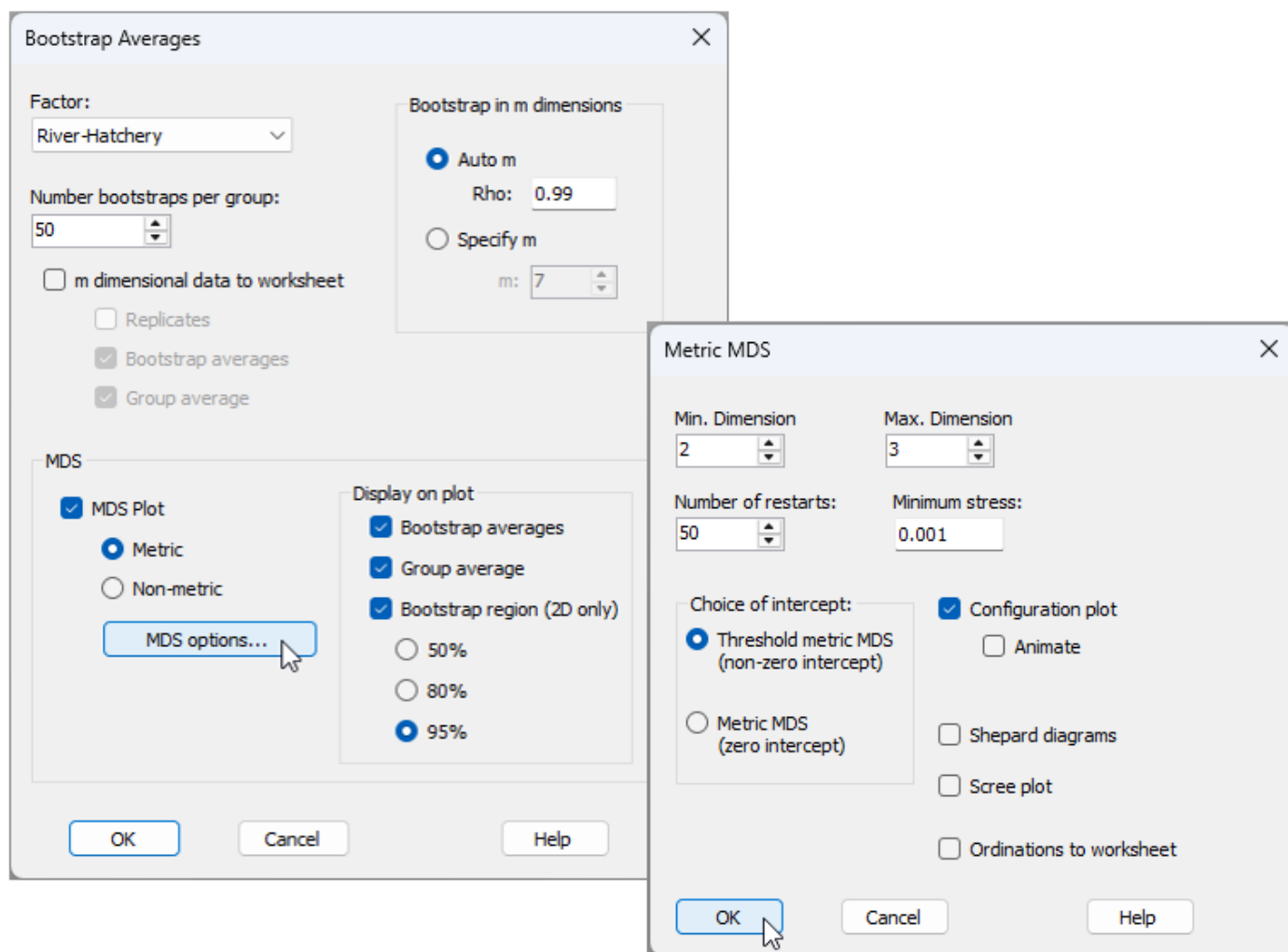
Visualise centroid and dispersion differences

We may consider using a *bootstrap average plot* for a holistic view, along with *ordinations of subsets* of the data (e.g., corresponding to pair-wise tests), to elucidate differences in centroids and/or dispersions that may have been detected among cells in our study design. We shall demonstrate each of these with the salmonid dataset next.

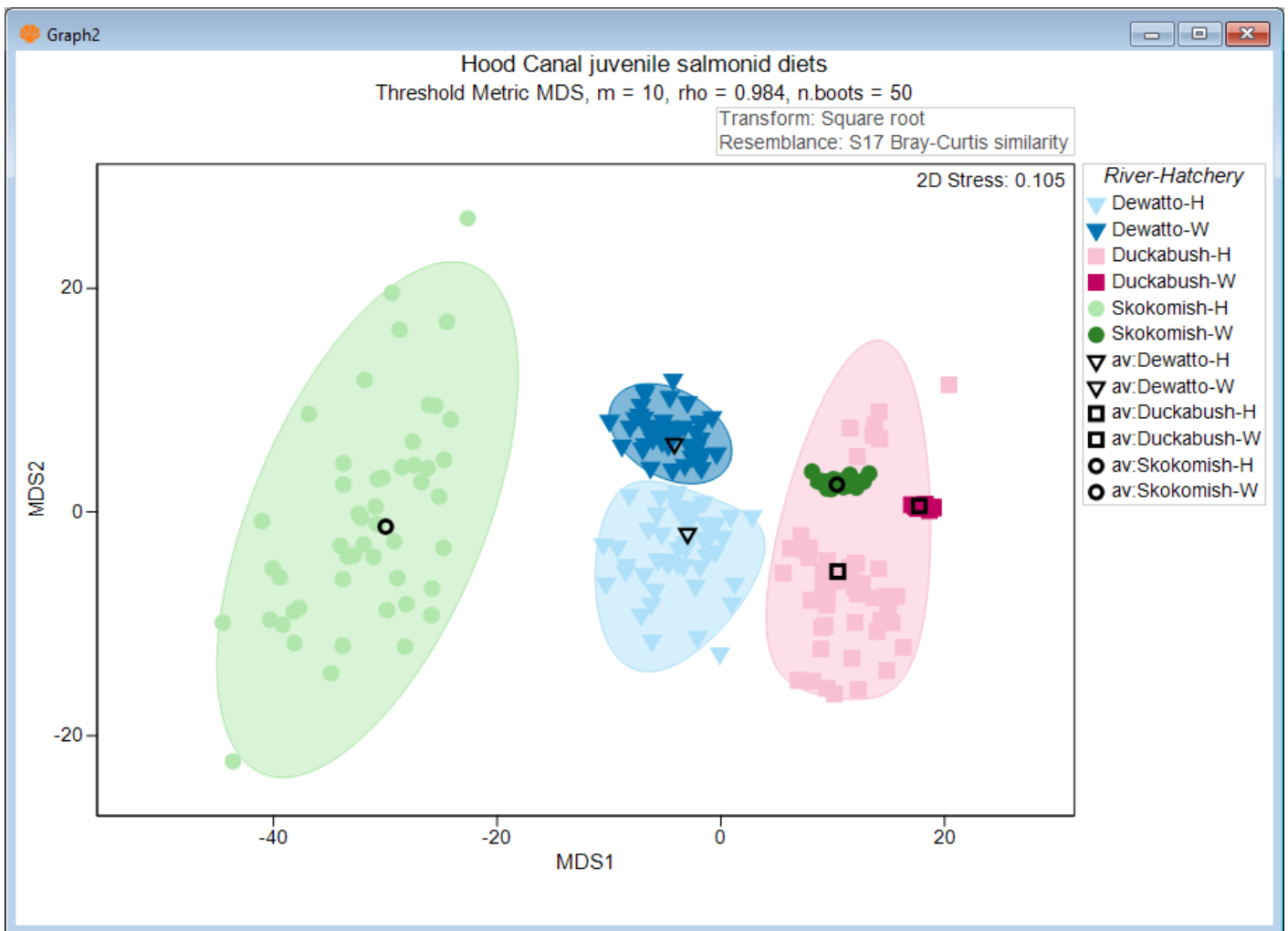
- Holistic view: bootstrap averages** - One useful way to visualise differences in centroids and in dispersions among cells simultaneously in a two-way (or multi-way) design such as this is to create a *bootstrap average* plot on the combined factor. Although the process of averaging bootstrapped data loses the details regarding inter-sample

relationships among the original individual replicate sampling units, this type of plot does have the advantage of permitting a holistic view of the overall study and relationships among the cells in a single plot.

For the salmonid data set, begin at the resemblance matrix ('Resem1') and click **Analyse > Bootstrap Averages...** and choose (Factor: River-Hatchery), then in the 'MDS' section of the dialog, click on the 'MDS options...' button **MDS options...**. This will bring up a separate dialog box, and under 'Choice of intercept:' choose (• Threshold metric MDS (non-zero intercept)), then click '**OK**' (you can leave the defaults for all the rest).[†]



The resulting bootstrap-average plot ('Graph2' in the Explorer tree), after a few alterations to symbols and colours[‡], looks like this:



From this plot, we can see the following patterns, further supporting results of the statistical tests we have done:

- There is generally greater dispersion (variability) in the diets of hatchery-reared fish (light colours) compared to wild-type fish (dark colours).
- A difference between the diets of hatchery-reared fish vs wild-type fish (a shift in centroid) was detected for the Duckabush river (light red vs dark red) and the Skokomish river (light green vs dark green), but not for the Dewatto river (light blue vs dark blue).
- There were differences in the diets of fish caught from different river systems when we considered (separately) either the hatchery-reared fish (light green, light red and light blue are all distinct from one another), or the wild-type fish (dark green, dark red and dark blue are also all quite distinct from one another).
- The interaction between the factors is explained largely by the difference in the diets between hatchery-reared vs wild-type fish being much larger for fish caught in the Skokomish river than for those caught in the other river systems. For the Dewatto river, neither centroid nor dispersion effects were detected.

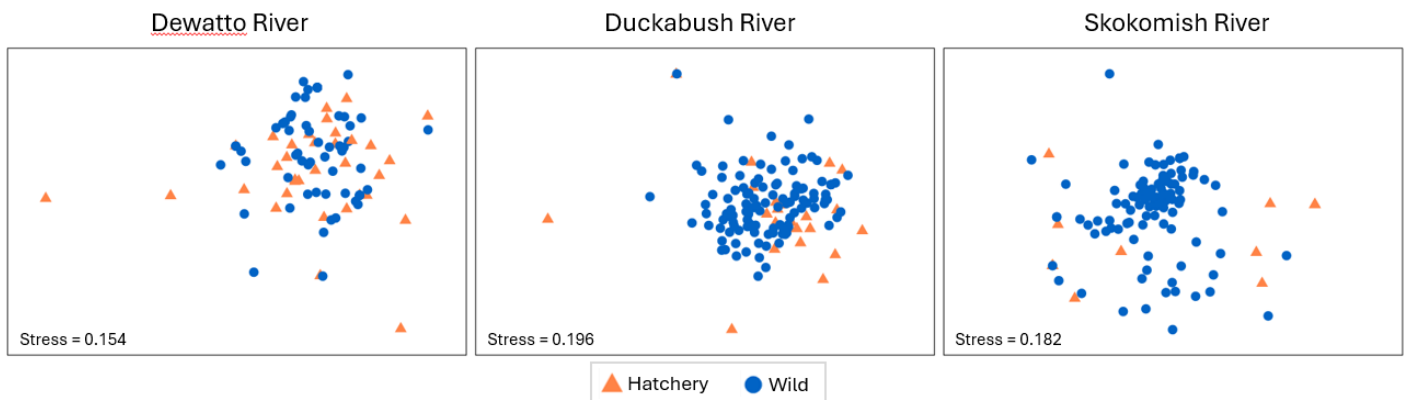
There are two important additional things to note about the above bootstrap average plot:

- *First*, the default colours and symbols in this bootstrap average plot produced by PRIMER 8 have been altered in the image above to make it easier to see changes in centroid and/or dispersion for the factors of interest.[‡]

- *Second*, we must not forget that every symbol on this plot is an *average*, and the dispersion of averages is therefore going to reflect not just the dispersion of the original data, but also the *sample size*. More specifically, we already should expect that the larger the sample size, the smaller the group's dispersion is expected to appear, due simply to the central limit theorem.⁵ For this reason, we must take observed differences in dispersion seen in a bootstrap average plot with a certain grain of salt when sample sizes differ among the groups (as here). Examining individual plots of dispersions of replicate sampling units (for subsets of the data, if necessary, as in step 12 below) will help to clarify the extent of genuine underlying dispersion differences among groups.

12. **Detailed view: ordinations of subsets** - Another way to visualise various aspects of these results is to 'zoom in' and examine *ordination plots of sub-sets* of the full dataset, such as those corresponding to pair-wise tests. This is a bit more direct than the bootstrap average approach, permitting investigation of the original inter-sample relationships among replicates, but of course each plot is restricted in its focus to a particular sub-set of the data, so the 'big-picture' information about relationships among *all* of the cells in the study design (as seen in the bootstrap average plot) is not able to be seen.

For example, let's consider aiming to visualise the pair-wise results from tests done according to point 10(i) above. We need to split the data into three groups according to the three river systems and run nMDS on each one, removing the labels and showing symbols for the factor 'Hatchery' on the resulting plots. You could start from the standardised and transformed data sheet (called 'Data2' in the Explorer tree) and click on **Tools > Split Data...** > (Samples > \$\checkmark\$Split by factor: **River**). Rename the resulting 3 datasheets according to the appropriate 3 river names, and proceed from there to calculate Bray-Curtis, then nMDS in each case. Doing this yields the following graphics:



Relevant things to note about the above plots are:

- For the Duckabush river, there is no apparent difference between the diets of hatchery-reared vs wild-type fish in terms of either their centroids or their relative dispersions; this is in line with the non-significant test results for PERMDISP and PERMANOVA that were found (above) for this river.
- For the other two rivers (Duckabush and Skokomish), a shift in centroid and a change in dispersion is apparent in both nMDS plots, although the stress in both of these final 2D configurations is quite large (approaching 0.2). We should therefore refrain from making any further interpretations of fine-scale patterns.

A similar set of ordinations on sub-sets of data may also be constructed to help visualise pair-wise results from tests done according to point 10(ii) above. We shall leave it to the reader to consider examining those on their own, if desired.

[¶]Data courtesy of Katie Doctor-Shelby, formerly based at the Northwest Fisheries Science Centre, National Oceanic and Atmospheric Administration (NOAA), Seattle, WA, USA.

[†]Note: the default here is to calculate 50 bootstrap averages per group. This can take a long time to execute! You may wish to speed up the process by choosing to do just 25 or 30 bootstraps instead.

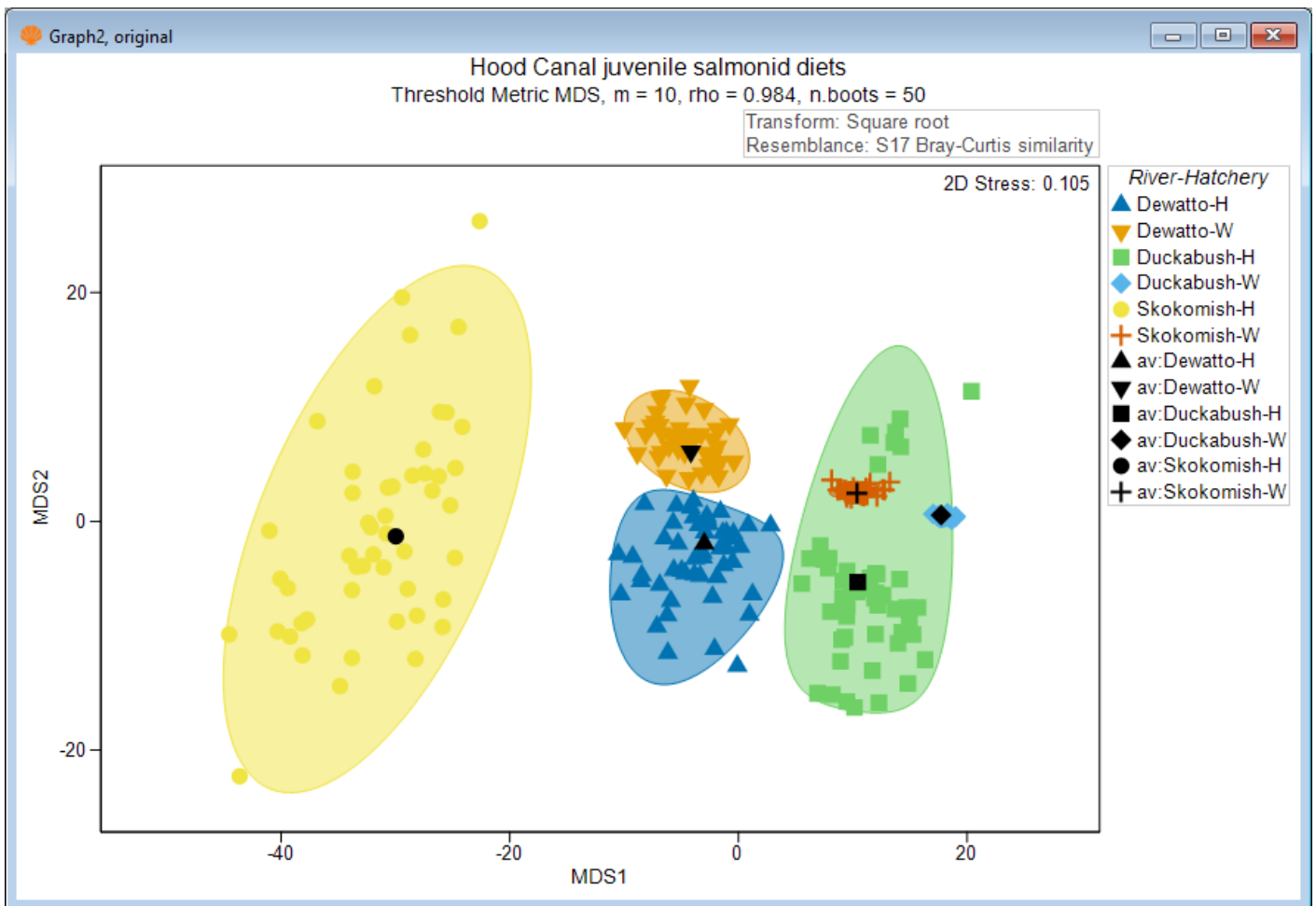
[‡] If you click on the legend itself inside the bootstrap average plot, you will see a complete legend key, as shown below. To the left is the default legend in PRIMER 8, and to the right are the customised symbols and colours I chose for the example. I find that when I do a bootstrap average plot (or a plot of distances among centroids) where the 'averages' are actually 'cells' in a factorial design (as here, where we have $3 \times 2 = 6$ cells in a two-way crossed design), it is useful to create a colour and symbol scheme that will make it easy to compare levels of factors of interest. For example, in the present case, I used blue, red and green for the three different rivers, then chose a lighter tint for the hatchery-reared fish and a darker shade for the wild-type fish (see the 'customised' key on the right in the image below). I also used a common symbol for each river system as well (although one could use, say open vs closed symbols as another option here). However, if you wish to cater carefully for any type of colour-blindness, then you will find that the default colours now used in PRIMER 8 are designed especially to do this. You might also find that maintaining the use of different readily distinguishable symbols for all of the cells in the design is a good idea. For this example, see the default legend shown at left below, and the associated default bootstrap average plot below that.

Default

Level	Symbol	Line	Hatch	Colour
Dewatto-H	▲	-----	▨	Blue
Dewatto-W	▼	-----	▨	Red
Duckabush-H	■	-----	▨	Green
Duckabush-W	◆	-----	▨	Blue
Skokomish-H	●	-----	▨	Yellow
Skokomish-W	+	-----	▨	Orange
av:Dewatto-H	▲	-----	▨	Black
av:Dewatto-W	▼	-----	▨	Black
av:Duckabush-H	■	-----	▨	Black
av:Duckabush-W	◆	-----	▨	Black
av:Skokomish-H	●	-----	▨	Black
av:Skokomish-W	+	-----	▨	Black

Customised

Level	Symbol	Line	Hatch	Colour
Dewatto-H	▼	-----	▨	Blue
Dewatto-W	▼	-----	▨	Red
Duckabush-H	■	-----	▨	Green
Duckabush-W	■	-----	▨	Blue
Skokomish-H	●	-----	▨	Yellow
Skokomish-W	●	-----	▨	Orange
av:Dewatto-H	▼	-----	▨	Black
av:Dewatto-W	▼	-----	▨	Black
av:Duckabush-H	■	-----	▨	Black
av:Duckabush-W	■	-----	▨	Black
av:Skokomish-H	○	-----	▨	Black
av:Skokomish-W	○	-----	▨	Black



[§]Recall the central limit theorem from classical univariate statistics. Let Y be a random variable with an unknown distribution that has a mean of μ_Y and a variance of σ_Y^2 . Now suppose you take a random sample of size n with realised values $\{y_1, y_2, \dots, y_n\}$ and calculate the average of that sample as: $\bar{y} = \sum_{i=1}^n y_i$. This average is then a random representative of variable $\bar{Y}$, which has a distribution that converges to a normal distribution, with a mean of $\mu_Y = \mu_{\bar{Y}}$ and a variance of $\sigma_{\bar{Y}}^2 = \sigma_Y^2/n$. Thus, it is clear that the distribution of the averages has a variance that gets smaller and smaller, the greater the sample size. Similarly, we expect the dispersion of averages (whether bootstrapped or otherwise) for a group of multivariate sampling units will get smaller and smaller (all else being equal) the larger the sample size of that group.

Revision #125

Created 31 March 2025 22:46:39 by Marti

Updated 4 December 2025 03:03:16 by Marti