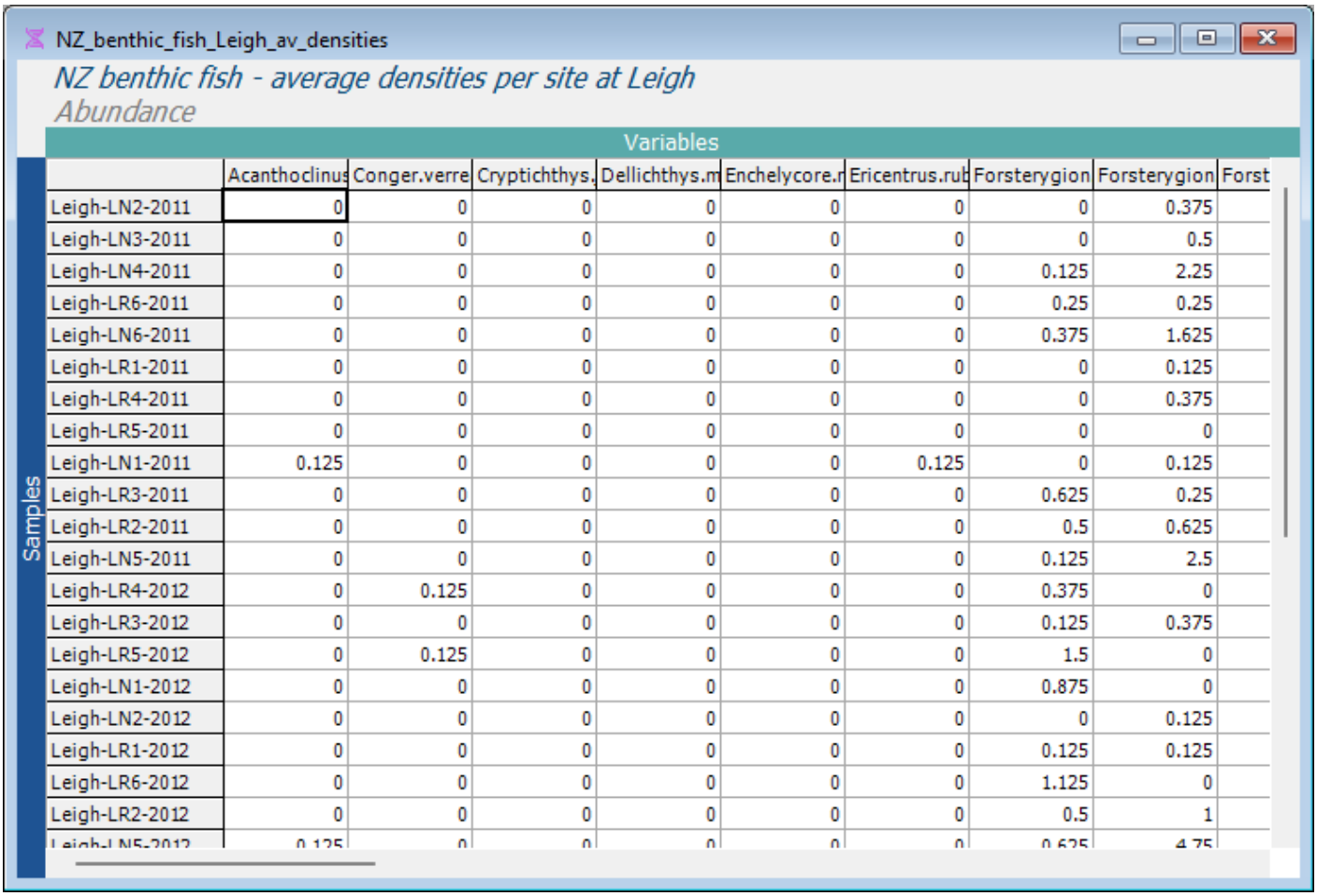
15.8 Output diagnostic plots from CAP

In PRIMER 8, it is now possible to output diagnostic plots from the CAP routine (i.e., canonical analysis of principal coordinates, an analysis obtained from a resemblance matrix by clicking **PERMANOVA+** > **CAP**). You simply tick the new option to '$\checkmark$Do diagnostic plots' in the CAP dialog window.

For example, consider a study of the assemblages of small benthic fishes living in rocky subtidal reef habitats in northeastern New Zealand, examined by Smith & Anderson (2016) . Data from this study are located in the file called 'NZ_benthic_fish.pri', in the 'Examples_P8' > 'NZ_benthic_fish' folder. Surveys were done of the benthic fish fauna, along with fine-scale habitat features in kelp forests and rocky reefs along the north-eastern coast of New Zealand. There were sites at a range of locations in and around several marine reserves (including Leigh, Tawharanui, Hahei and the Poor Knights Islands), and data were obtained over a period of 3 years (2011-2013). At each site, divers surveyed n = 8 transects, measuring 1 m $\times$ 5 m. Each transect was made up of five contiguous 1 m $\times$ 1 m quadrats. For each quadrat, divers first visually searched and recorded counts for all benthic fishes, then recorded the presence/absence of a set of pre-defined habitat features (see the file named 'NZ_benthic_fish_habitat.pri', located in the same folder).

Here, we shall consider only the potential effects of the marine reserve at Leigh on these fish communities in a fully balanced design. Note that the factor of 'Reserve' has two levels: 0 = outside the reserve, and 1 = inside the reserve. Due to the sparsity of the count data at small spatial scales, we shall analyse data summed to the transect level, then consider densities of each species (averages per transect) at the spatial scale of sites for the ensuing CAP analysis. Mean densities (per $5m^2$) of each fish species per site are located in the file named ' NZ_benthic_fish_Leigh_av_densities.pri'.

1. Open the file ('NZ_benthic_fish_Leigh_av_densities.pri') in PRIMER 8. It will look like this:


NZ_benthic_fish_Leigh_av_densities

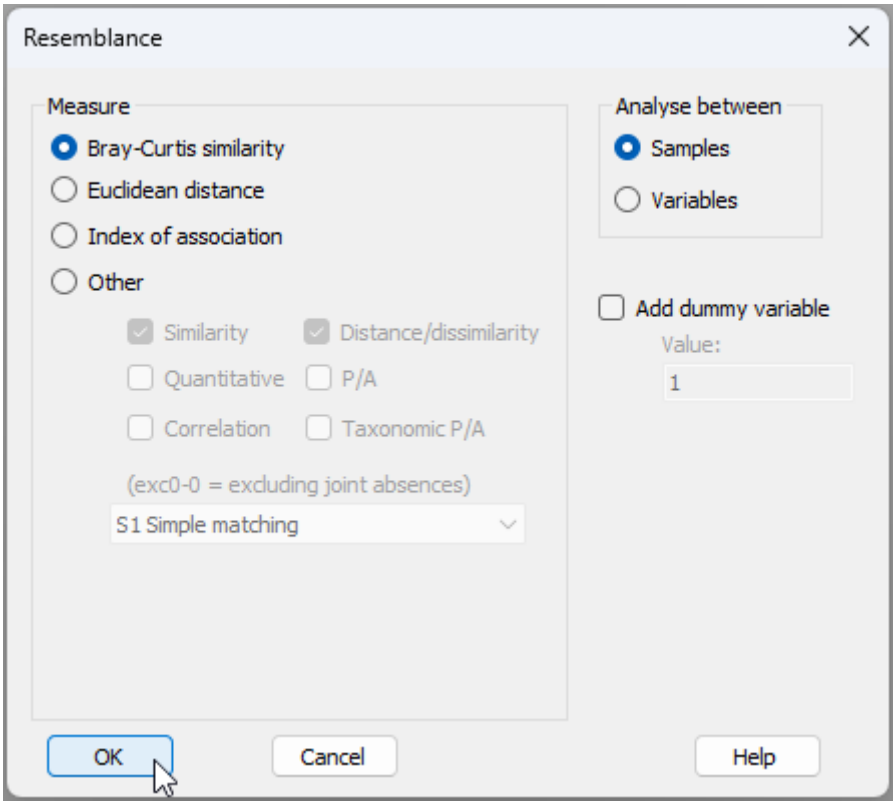
NZ benthic fish - average densities per site at Leigh

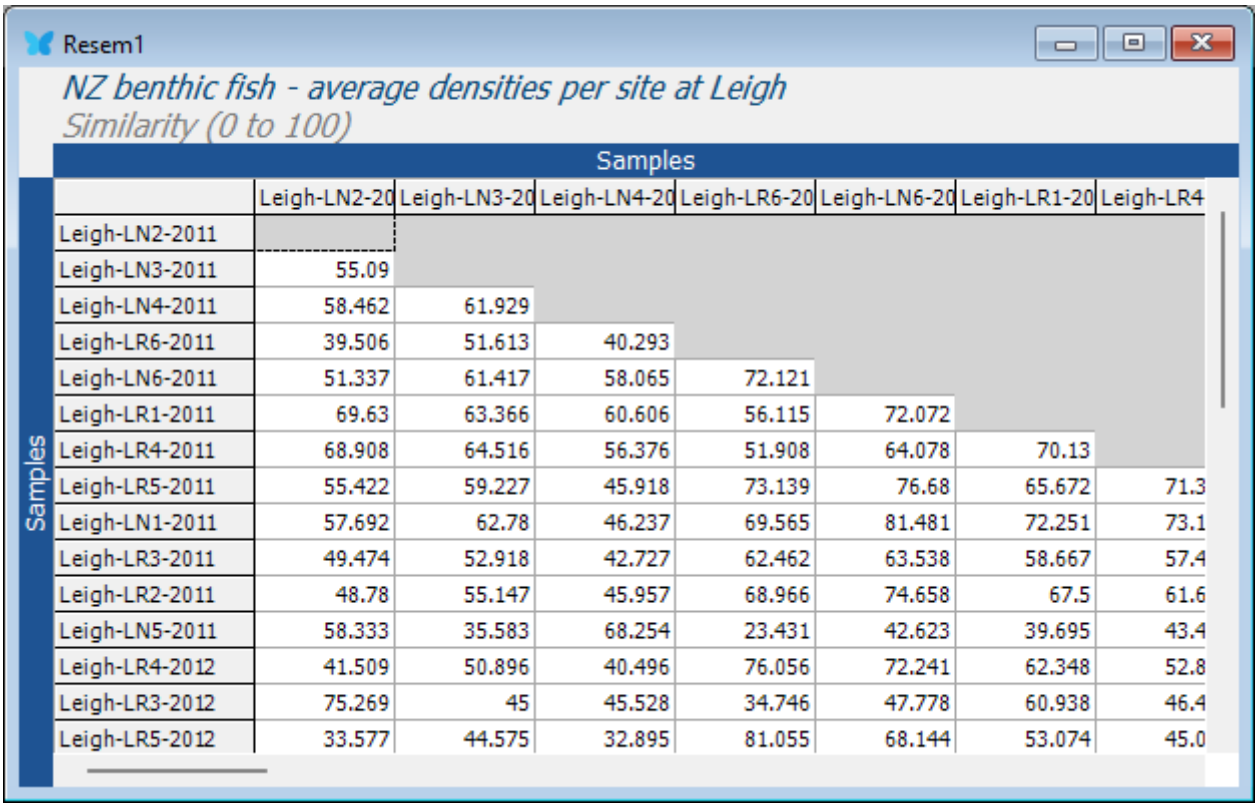
Abundance

Variables

Samples	Acanthoclinus	Conger.verre	Cryptichthys.	Dellichthys.m	Enchelycore.r	Ericentrus.rub	Forsterygion	Forsterygion	Forst
Leigh-LN2-2011	0	0	0	0	0	0	0	0.375	
Leigh-LN3-2011	0	0	0	0	0	0	0	0.5	
Leigh-LN4-2011	0	0	0	0	0	0	0.125	2.25	
Leigh-LR6-2011	0	0	0	0	0	0	0.25	0.25	
Leigh-LN6-2011	0	0	0	0	0	0	0.375	1.625	
Leigh-LR1-2011	0	0	0	0	0	0	0	0.125	
Leigh-LR4-2011	0	0	0	0	0	0	0	0.375	
Leigh-LR5-2011	0	0	0	0	0	0	0	0	
Leigh-LN1-2011	0.125	0	0	0	0	0.125	0	0.125	
Leigh-LR3-2011	0	0	0	0	0	0	0.625	0.25	
Leigh-LR2-2011	0	0	0	0	0	0	0.5	0.625	
Leigh-LN5-2011	0	0	0	0	0	0	0.125	2.5	
Leigh-LR4-2012	0	0.125	0	0	0	0	0.375	0	
Leigh-LR3-2012	0	0	0	0	0	0	0.125	0.375	
Leigh-LR5-2012	0	0.125	0	0	0	0	1.5	0	
Leigh-LN1-2012	0	0	0	0	0	0	0.875	0	
Leigh-LN2-2012	0	0	0	0	0	0	0	0.125	
Leigh-LR1-2012	0	0	0	0	0	0	0.125	0.125	
Leigh-LR6-2012	0	0	0	0	0	0	1.125	0	
Leigh-LR2-2012	0	0	0	0	0	0	0.5	1	
Leigh-LN5-2012	0.125	0	0	0	0	0	0.625	4.75	

2. From 'NZ_benthic_fish_Leigh_av_densities', click **Analyse** > **Resemblance...** and choose to calculate the Bray-Curtis similarity among samples. This will produce a resemblance matrix called 'Resem1'.




Resem1

NZ benthic fish - average densities per site at Leigh

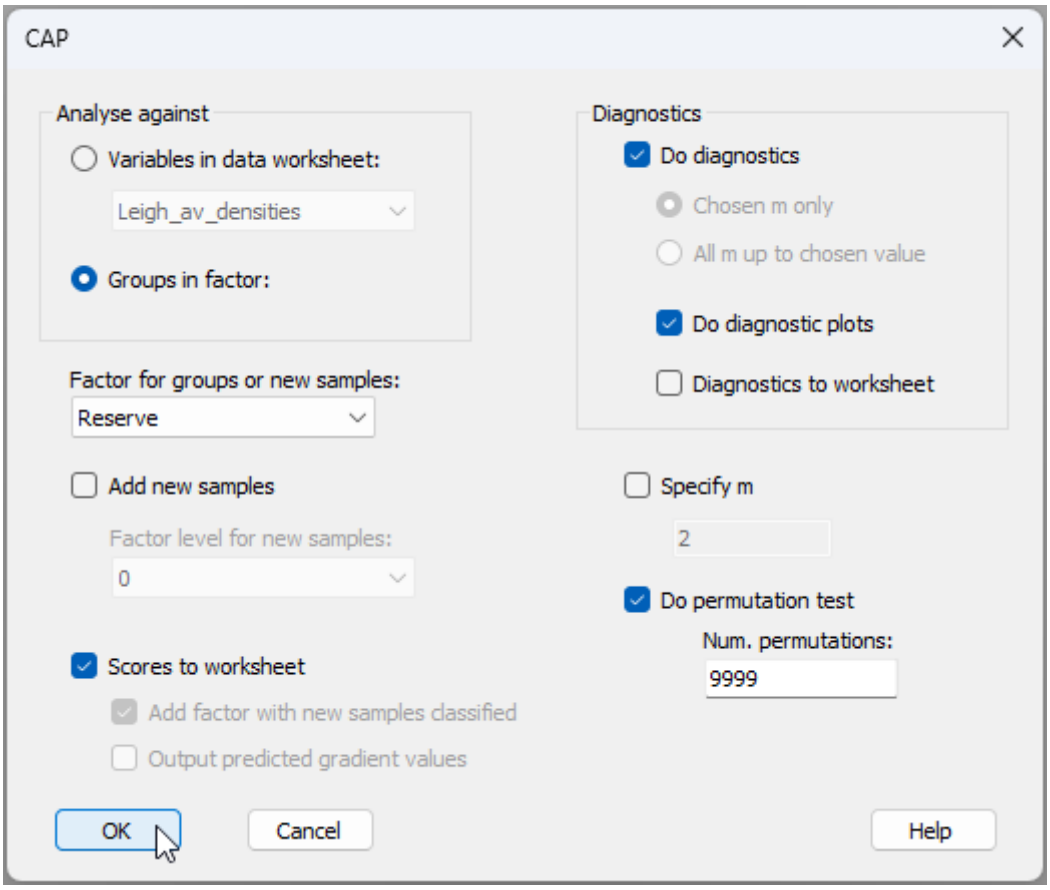
Similarity (0 to 100)

Samples

Samples	Leigh-LN2-20	Leigh-LN3-20	Leigh-LN4-20	Leigh-LR6-20	Leigh-LN6-20	Leigh-LR1-20	Leigh-LR4
Leigh-LN2-2011							
Leigh-LN3-2011	55.09						
Leigh-LN4-2011	58.462	61.929					
Leigh-LR6-2011	39.506	51.613	40.293				
Leigh-LN6-2011	51.337	61.417	58.065	72.121			
Leigh-LR1-2011	69.63	63.366	60.606	56.115	72.072		
Leigh-LR4-2011	68.908	64.516	56.376	51.908	64.078	70.13	
Leigh-LR5-2011	55.422	59.227	45.918	73.139	76.68	65.672	71.3
Leigh-LN1-2011	57.692	62.78	46.237	69.565	81.481	72.251	73.1
Leigh-LR3-2011	49.474	52.918	42.727	62.462	63.538	58.667	57.4
Leigh-LR2-2011	48.78	55.147	45.957	68.966	74.658	67.5	61.6
Leigh-LN5-2011	58.333	35.583	68.254	23.431	42.623	39.695	43.4
Leigh-LR4-2012	41.509	50.896	40.496	76.056	72.241	62.348	52.8
Leigh-LR3-2012	75.269	45	45.528	34.746	47.778	60.938	46.4
Leigh-LR5-2012	33.577	44.575	32.895	81.055	68.144	53.074	45.0

3. From the 'Resem1' matrix, run the CAP routine aiming to distinguish small benthic fish communities inside *vs* outside the marine reserve, by clicking **PERMANOVA+** > **CAP** and choose:
   (Analyse against $\bullet$Groups in factor > Factor for groups or new samples: 'Reserve') &
   ($\checkmark$Scores to worksheet) &
   (Diagnostics > $\checkmark$Do diagnostics > $\checkmark$Do diagnostic plots) &
   ($\checkmark$Do permutation test > Num. permutations: 9999)

then click **OK**, as shown below:



The option to *do diagnostic plots* is new to PRIMER 8.

The analysis will run and produce the following:

- 'CAP1' (CAP1), containing all of the essential results in a rich text format (*.rtf), and
- 'MultiPlot1' (MultiPlot1), containing the CAP plot itself (as 'Graph1') along with plots of the *diagnostics for choosing m*, where m = the number of principal coordinate axes (PCOs) used to produce the canonical (CAP) axis (as 'Graph2' through 'Graph5').

From 'CAP1' and the accompanying CAP plot ('Graph1'), shown below, we can see that the canonical correlation associated with the 'Reserve' effect is reasonably strong (δ_1 = 0.644), and that this is statistically significant (δ_1^2 = 0.415, P = 0.001). This CAP model was achieved with m = 4 PCO axes, which achieved a leave-one-out allocation success of 77.78% under cross-validation.

CAP1

CANONICAL ANALYSIS

Correlations

Eigenvalue	Correlation	Corr.Sq.
1	0.644	0.4147

DIAGNOSTICS

m	prop.G	ssres	d_1^2	%correct
1	0.4148	0.99277	0.0626	61.111
2	0.5625	0.74153	0.3958	72.222
3	0.6935	0.66856	0.411	75
4	0.7783	0.65859	0.4147	77.778
5	0.8554	0.68304	0.4397	75
6	0.9145	0.74106	0.4428	72.222
7	0.9528	0.74104	0.4496	75
8	0.9902	0.84985	0.4507	69.444

Cross Validation

Leave-one-out Allocation of Observations to Groups
(for the choice of m: 4)

	classified			
Orig. group	0	1	Total	%correct
0	14	4	18	77.778
1	4	14	18	77.778

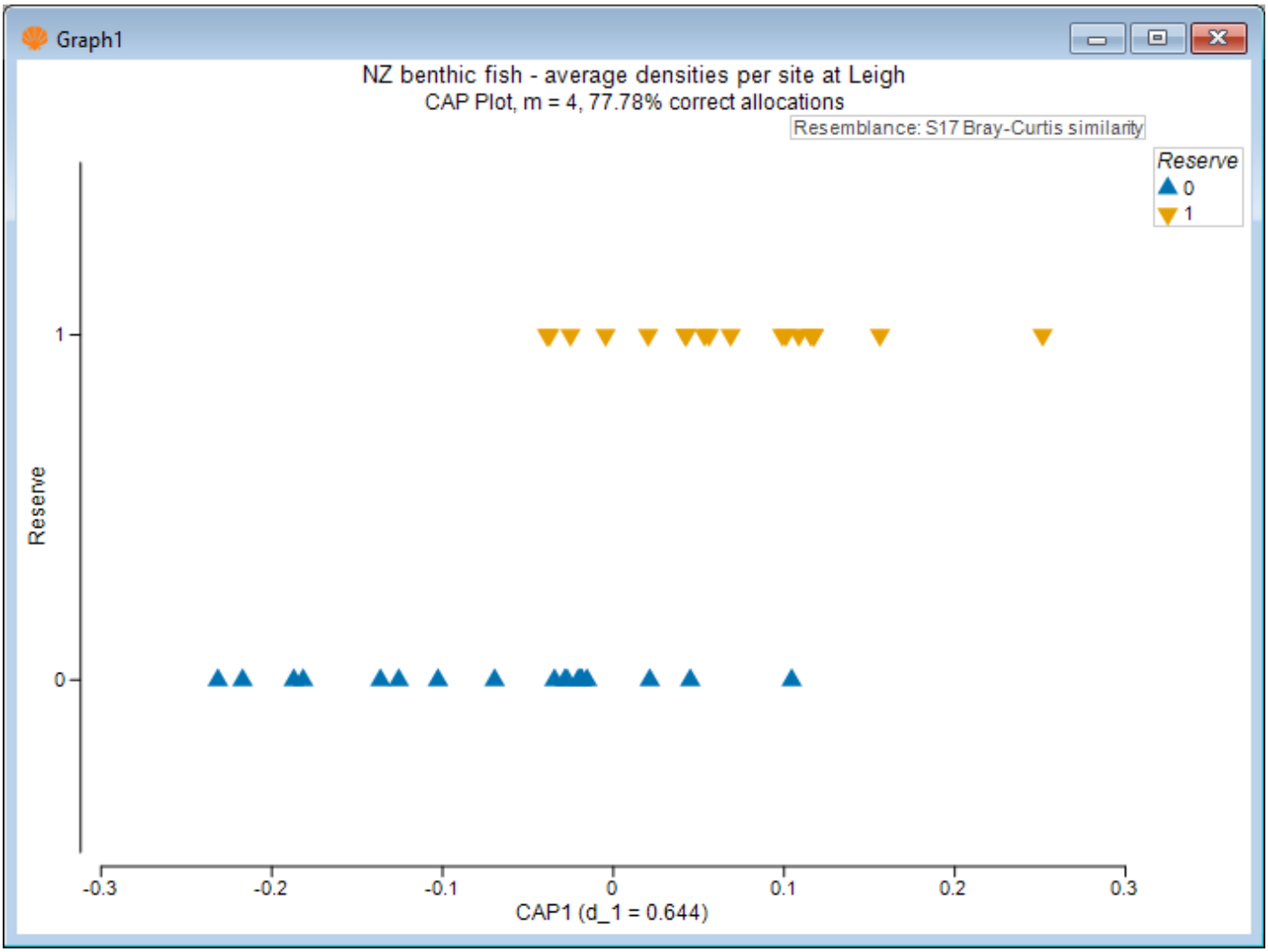
Total correct: 28/36 (77.778%)
Mis-classification error: 22.222%

Individual samples that were mis-classified

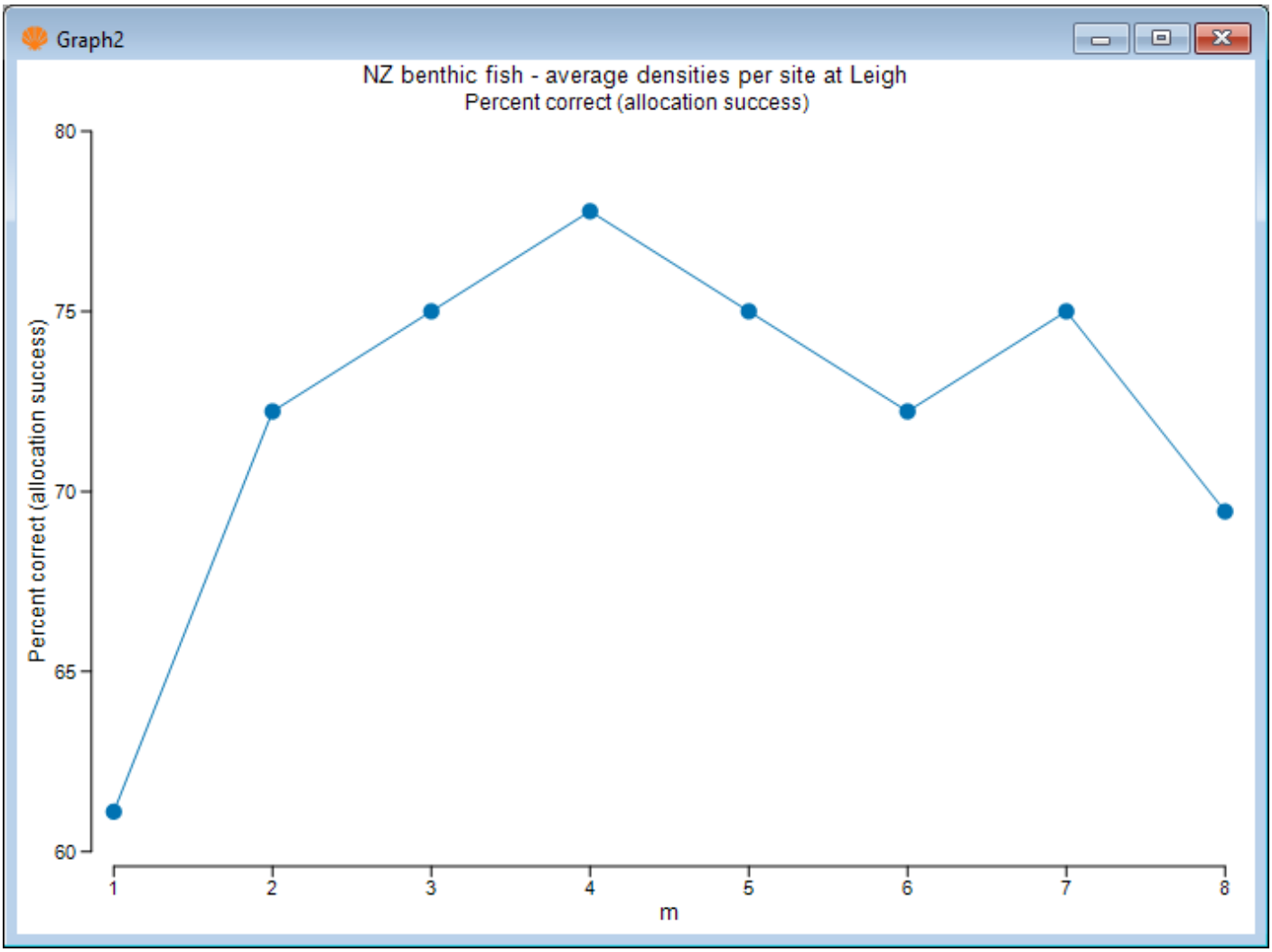
Sample	Orig.group	Class.group
Leigh-LN1-2011	0	1
Leigh-LN1-2012	0	1
Leigh-LN1-2013	0	1
Leigh-LN2-2013	0	1
Leigh-LR1-2011	1	0
Leigh-LR3-2012	1	0
Leigh-LR2-2012	1	0
Leigh-LR2-2013	1	0

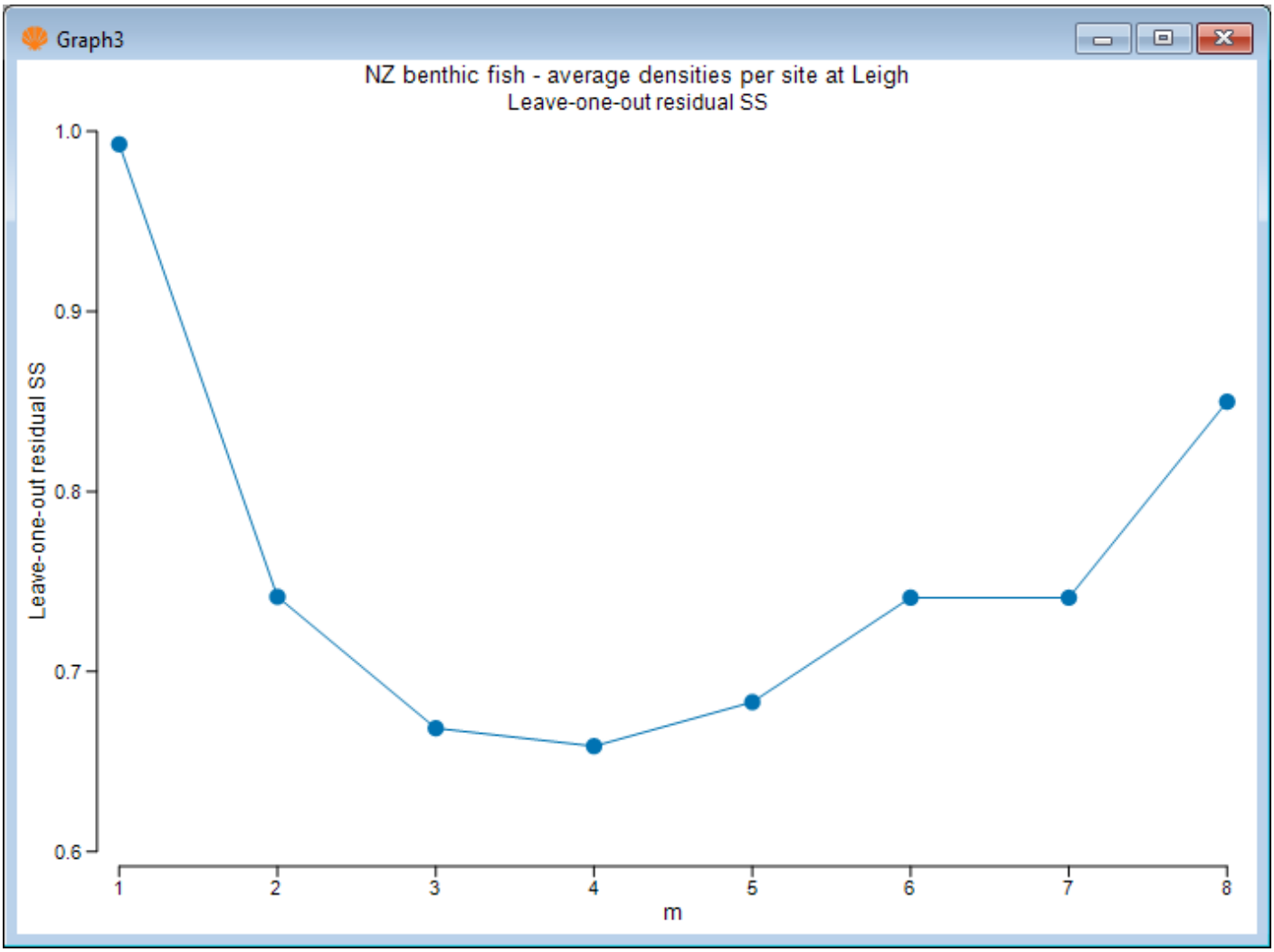
PERMUTATION TEST
trace statistic = (tr(Q_m'HQ_m))
first squared canonical correlation = (delta_1^2)

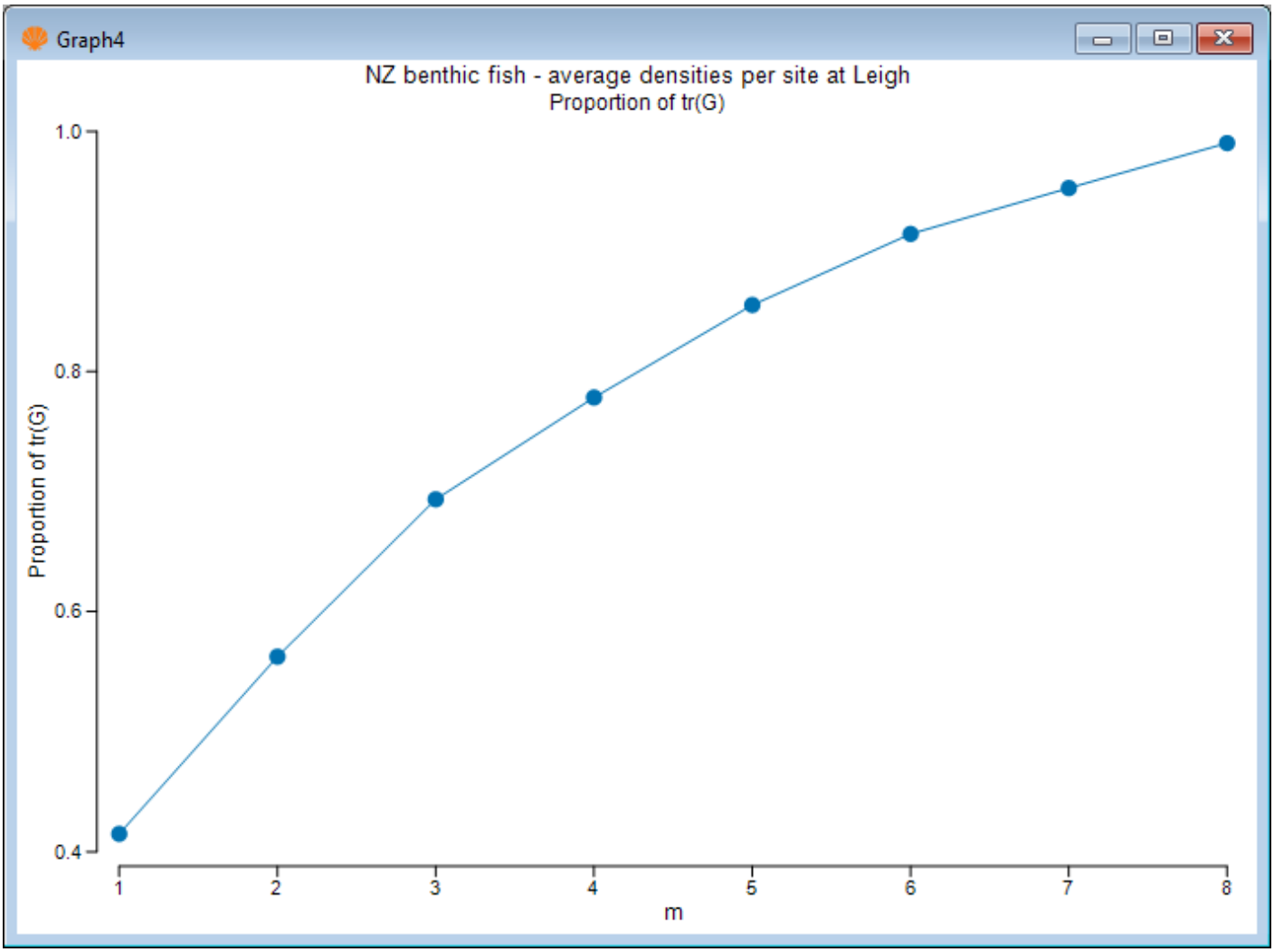
tr(Q_m'HQ_m): 0.41472 P: 0.001
delta_1^2: 0.41472 P: 0.001
No. of permutations used: 9999

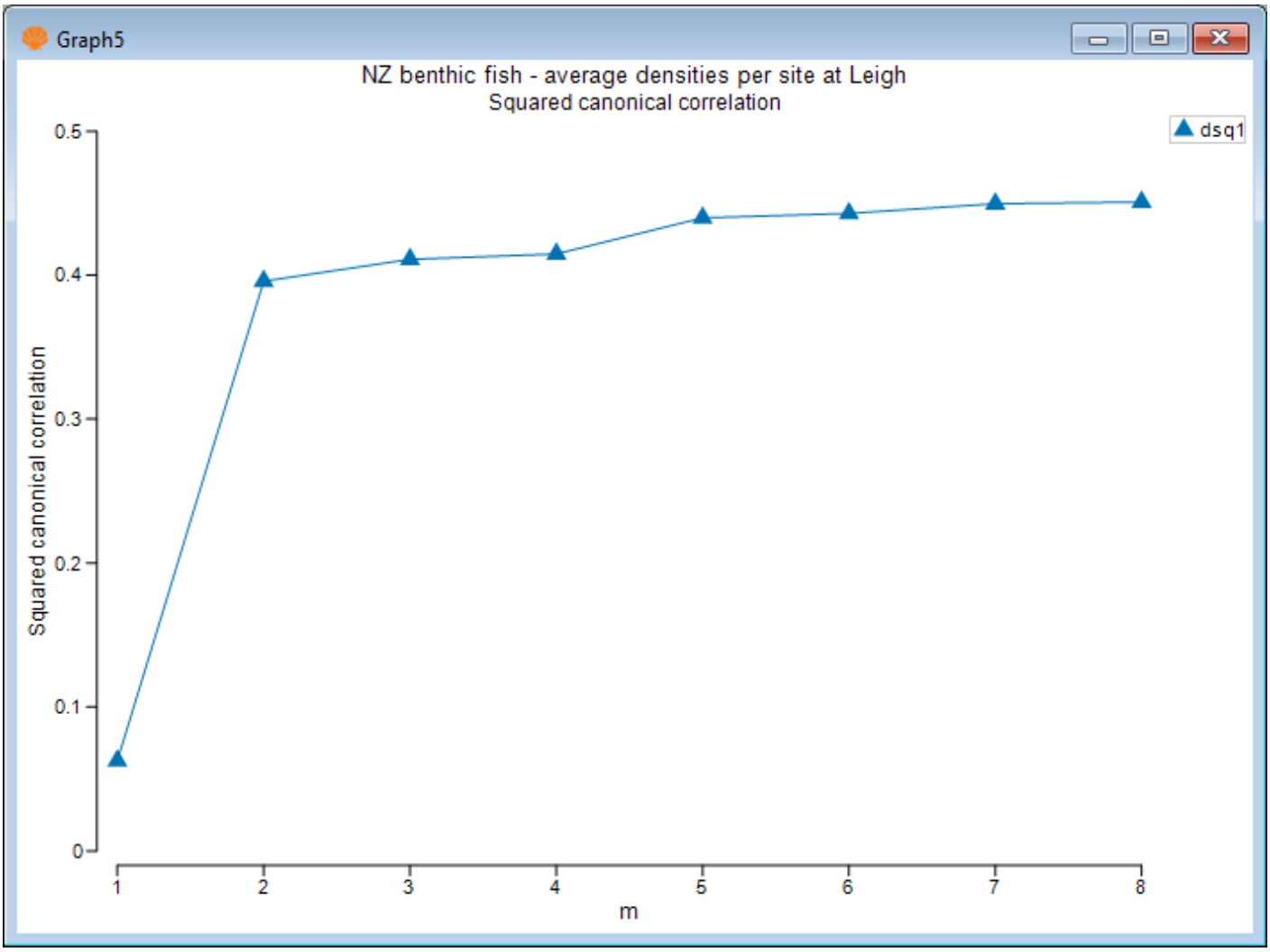


The four *diagnostic plots* are shown below.


Graph2
NZ benthic fish - average densities per site at Leigh
Percent correct (allocation success)
Percent correct (allocation success)
80
75
70
65
60
1
2
3
4
5
6
7
8
m


Graph3
NZ benthic fish - average densities per site at Leigh
Leave-one-out residual SS
Leave-one-out residual SS
1.0
0.9
0.8
0.7
0.6
1
2
3
4
5
6
7
8
m


Graph4
NZ benthic fish - average densities per site at Leigh
Proportion of tr(G)
1.0
0.8
0.6
0.4
Proportion of tr(G)
1
2
3
4
5
6
7
8
m



It is clear that the choice of m = 4 is a good one here, as this choice achieved the greatest leave-one-out allocation success ('Graph2') and the lowest leave-one-out residual sum-of-squares ('Graph3'). Although this can also (technically) be seen in the '*DIAGNOSTICS*' section of the 'CAP1' output above, it is certainly much clearer to see this information in diagnostic plots, whereby the wisdom of the choice made, overall, by reference to other values of m, can be readily assessed.

Revision #43
Created 19 July 2025 04:46:20 by Marti
Updated 1 February 2026 20:25:51 by Marti