4.6 (Holdfast invertebrates)

To demonstrate conditional tests in DISTLM, we will consider the number of species inhabiting holdfasts of the kelp *Ecklonia radiata* in the dataset from New Zealand, located in the file <mark>hold.pri</mark> in the 'HoldNZ' folder of the 'Examples add-on' directory. These multivariate data were analysed in the section entitled <mark>Designs with covariates</mark> in chapter 1. Due to the well-known species-area relationship in ecology, we expect the number of species to increase with the volume (size) of the holdfast (Smith, Simpson & Cairns (1996) , Anderson, Diebel, Blom *et al.* (2005) , Anderson, Millar, Blom *et al.* (2005)). In addition, the density of the kelp forest surrounding a particular kelp plant might also affect the number of species able to immigrate or emigrate to/from the holdfast (Goodsell & Connell (2002)) and depth can also play a role in structuring holdfast communities (Coleman, Vytopil, Goodsell *et al.* (2007)). Volume (in ml), density of surrounding *Ecklonia* plants (per m^2) and depth (in m) were all measured as part of this study and are contained in the file <mark>holdenv.pri</mark>. The associated draftsman plot (Fig. 1.45) shows that these variables are reasonably evenly distributed across their range (see <mark>Assumptions & diagnostics</mark>, below).

Here, we will focus on the number of species of molluscs only, as members of this phylum were all identified down to species level. Open the file <mark>hold.pri</mark> in PRIMER, select the molluscan species only and duplicate these into their own sheet, named <mark>molluscs,</mark> as shown in Fig. 1.27 in the section entitled <mark>Nested designs</mark> of chapter 1. Choose **Analyse > DIVERSE** and retain only the tick mark next to $\checkmark$Total species: S on the 'Other' tab, but remove all others. Also choose to get $\checkmark$Results to worksheet and name the resulting sheet <mark>No.species</mark>. This will be the (in this case univariate) response variable for the analysis. Calculate a Euclidean distance matrix among samples on the basis of this variable. Next, open up the <mark>holdenv.pri</mark> file in PRIMER. Here we shall focus on running the analysis on only two predictor variables: volume and density, asking the specific question: does density explain a significant portion of the variation in the number of species of molluscs after volume is taken into account? Do this by choosing **PERMANOVA+ > DISTLM** > (Predictor variables worksheet: <mark>holdenv</mark>) and click on the button marked 'Select variables/groups'. This opens up a 'Selection' dialog where you can choose to force the inclusion or exclusion of particular variables from the model and you can also change the order in which the variables (listed in the 'Available' column) are fitted. Choose to exclude depth for the moment and to fit volume first, followed by density (see Fig. 4.8). Back in the DISTLM dialog, choose (Selection procedure •All specified) & (Selection criterion •R^2) & (Num. permutations: <mark>9999</mark>).
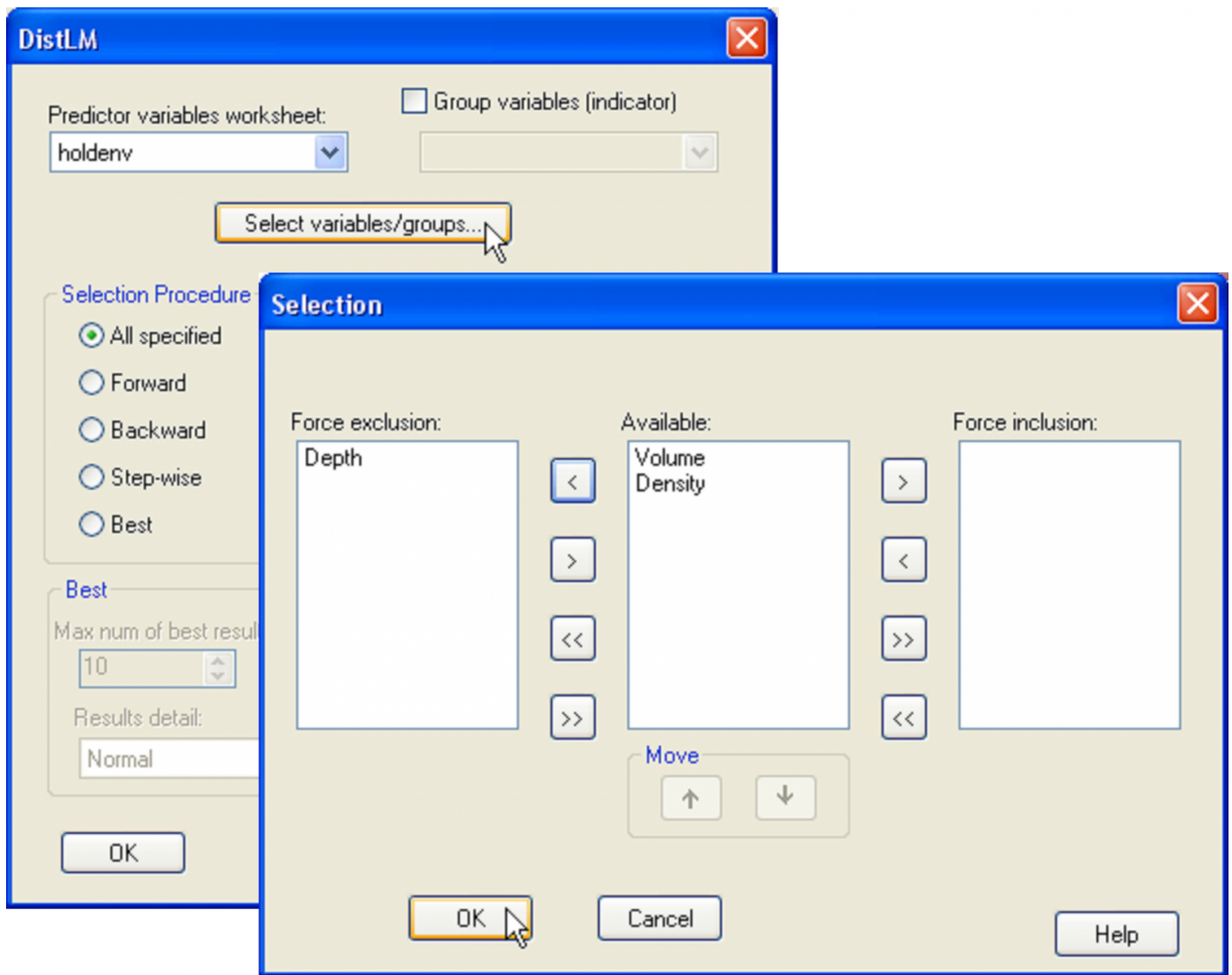
Fig. 4.8. Dialog in DISTLM for specifying a model with two variables, volume and density (in that order), but excluding the variable of depth.

The results (Fig. 4.9) show first the *marginal* tests of how much each variable explains when taken alone, *ignoring* all other variables. Here, we see that volume explains a significant amount of about 32.8% of the variation in the number of mollusc species ($P = 0.0001$). Density, when considered alone, only explains about 5.7% of the variation, but this small amount is nevertheless found to be statistically significant when tested by permutation ($P = 0.037$). Note that these marginal tests do not include any corrections for multiple testing. Although this is not such a big issue in the present example, where there are only two marginal tests, clearly the probability of rejecting a null hypothesis by chance will increase the more tests we do, and if there are a lot of predictor variables, the user may wish to adjust the level of significance at which individual tests reject their null hypothesis. The philosophy taken here, as in the case of multiple pair-wise tests in the PERMANOVA routine, is that the permutation approach provides an exact test for each of these individual hypotheses. It is then up to the user to decide whether or not to apply subsequent corrections, if any, to guard against the potential inflation of overall rates of error.
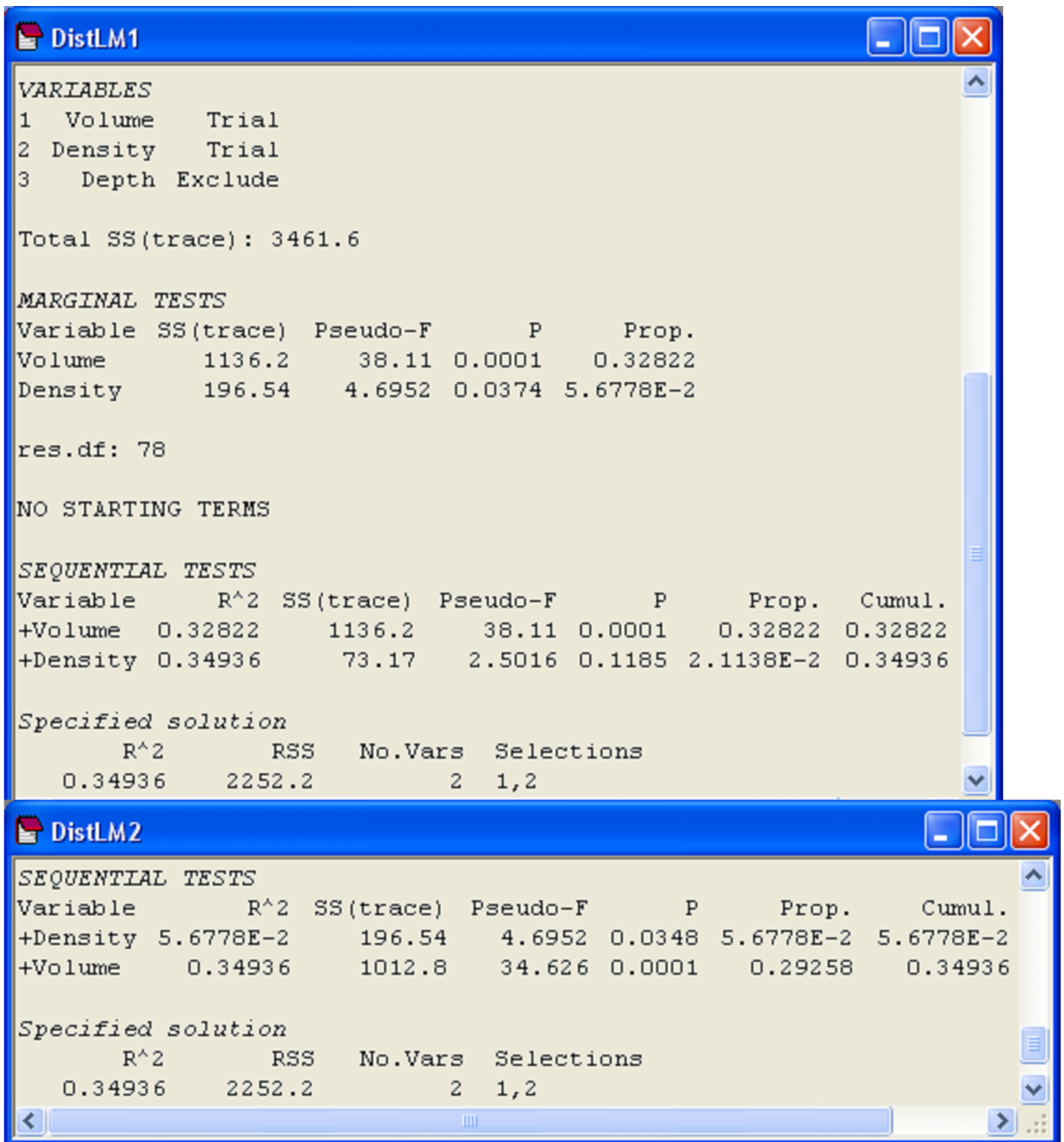
```
DistLM1                                                    [_][口][X]

VARIABLES
1   Volume    Trial
2  Density    Trial
3    Depth  Exclude


Total SS(trace): 3461.6


MARGINAL TESTS
Variable  SS(trace)  Pseudo-F       P       Prop.
Volume        1136.2     38.11 0.0001    0.32822
Density       196.54    4.6952 0.0374 5.6778E-2


res.df: 78


NO STARTING TERMS


SEQUENTIAL TESTS
Variable        R^2 SS(trace)  Pseudo-F       P       Prop.    Cumul.
+Volume   0.32822     1136.2      38.11 0.0001    0.32822  0.32822
+Density 0.34936       73.17     2.5016 0.1185 2.1138E-2  0.34936


Specified solution
        R^2        RSS    No.Vars  Selections
   0.34936     2252.2          2   1,2
```

```
DistLM2                                                    [_][口][X]

SEQUENTIAL TESTS
Variable         R^2 SS(trace)  Pseudo-F        P       Prop.      Cumul.
+Density 5.6778E-2     196.54    4.6952 0.0348 5.6778E-2  5.6778E-2
+Volume    0.34936     1012.8    34.626 0.0001    0.29258    0.34936


Specified solution
        R^2        RSS    No.Vars  Selections
   0.34936     2252.2          2   1,2
```

Fig. 4.9. Results from DISTLM obtained by first fitting volume followed by density ('DistLM1'), including marginal tests, and then by fitting density followed by volume ('DistLM2').

Following the marginal tests in the output file, there is a section that contains sequential tests. These are the conditional tests of individual variables done in the order specified. All variables appearing above a particular variable are fitted as covariates, and each test examines whether adding that particular variable contributes significantly to the explained variation. In our case, we have the test of whether density adds a significant amount to the explained variation, given that volume has already been included in the model. These sequential conditional tests correspond directly to the notion of Type I SS met in chapter 1 (e.g., see Fig. 1.41). That is, we fit variable 1, then variable 2 *given* variable 1, then variable 3 *given* variables 1 and 2, and so on. In the present

example (Fig. 4.9, 'DistLM1'), density adds only about an additional 2.1% to the explained variation once volume has been fitted, and this is not statistically significant ($P = 0.12$). The column labeled 'Prop.' in the section of the output entitled 'Sequential tests' indicates the increase in the proportion of explained variation attributable to each variable that is added, and the column labeled 'Cumul.' provides a running cumulative total. For this example, these two variables together explained 34.9% of the variation in number of mollusc species in the holdfasts.

We can re-run this analysis and change the order in which the variables are fitted. This is done by clicking on particular variables and using the 'Move' arrows under the names of variables listed in the 'Available' column of the 'Selection' dialog box within DISTLM. There is no need to include marginal tests again, so we can also remove the tick in front of 'Do marginal tests' in the ordinary DISTLM dialog. When we fit density first, followed by volume, it is not at all surprising to find that: (i) volume explains a significant proportion of variation even after fitting density and (ii) the variation explained by both variables together remains the same, at 34.9% (Fig. 4.9, 'DistLM2').

Revision #6
Created 11 August 2022 14:54:07 by Arden
Updated 5 September 2022 11:12:25 by Arden