PERMANOVA+
for PRIMER:
Guide to Software and Statistical Methods

by
Marti J. Anderson†‡, Ray N. Gorleyπ & K. Robert Clarkeπ

†New Zealand Institute for Advanced Study
Massey University, Albany Campus
Private Bag 102 904, North Shore Mail Centre
Auckland 0745
New Zealand
tel: 64-9-213-6609
email: m.j.anderson@massey.ac.nz
web: http://www.nzias.ac.nz/

‡Founders of PRIMER-E Ltd, Plymouth, UK
# PERMANOVA+ for PRIMER: Guide to Software and Statistical Methods

## OVERVIEW

A. Contact details and installation of the PERMANOVA+ software
   - Getting in touch with us  
   - System requirements  
   - Installing PERMANOVA+  
   - Getting help  

B. Introduction to the methods of PERMANOVA+
   - Rationale  
   - Application  
   - Outline of routines  

C. Changes from DOS to PERMANOVA+ for PRIMER
   - The new Windows interface  
   - Development of the methods  

D. Using this manual
   - Typographic conventions  
   - Example data sets  
   - Finding your way around  

## GUIDE TO SOFTWARE AND STATISTICAL METHODS

1. Permutational ANOVA and MANOVA (PERMANOVA)
   - Key references  
   - General description  
   - Partitioning  
   - Huygens’ theorem  
   - Sums of squares from a distance matrix  
   - The pseudo-\(F\) statistic  
   - Test by permutation  
   - Assumptions  
   - One-way example (Ekofisk oil-field macrofauna)  
   - Creating a design file  
   - Running PERMANOVA  
   - Pair-wise comparisons  
   - Monte Carlo \(P\)-values (Victorian avifauna)  
   - PERMANOVA versus ANOSIM  
   - Two-way crossed design (Subtidal epibiota)  
   - Interpreting interactions  
   - Additivity  
   - Methods of permutation  
   - Additional assumptions  
   - Contrasts  
   - Fixed versus random factors (Tasmanian meiofauna)  
   - Components of variation  
   - Expected mean squares (EMS)  
   - Constructing \(F\) from EMS  
   - Exchangeable units  
   - Inference space and power  
   - Testing the design  
   - Nested design (Holdfast invertebrates)  
   - Estimating components of variation  
   - Pooling or excluding terms  
   - Designs that lack replication (Plankton net study)
2. Tests of homogeneity of dispersions (PERMDISP)
   Key references 87
   General description 87
   Rationale 87
   Multivariate Levene’s test (Bumpus’ sparrows) 88
   Generalisation to dissimilarities 90
   P-values by permutation 91
   Test based on medians 91
   Ecological example (Tikus Island corals) 92
   Choice of measure 92
   Dispersion as beta diversity (Norwegian macrofauna) 94
   Small sample sizes 96
   Dispersion in nested designs (Okura macrofauna) 96
   Dispersion in crossed designs (Cryptic fish) 100
   Concluding remarks 104

3. Principal coordinates analysis (PCO)
   Key references 105
   General description 105
   Rationale 105
   Mechanics of PCO 106
   Example: Victorian avifauna 107
   Negative eigenvalues 109
   Vector overlays 112
   PCO versus PCA (Clyde environmental data) 115
   Distances among centroids (Okura macrofauna) 118
   PCO versus MDS 121

4. Distance-based linear models (DISTLM) and distance-based redundancy analysis (dbRDA)
   Key references 123
   General description 123
   Rationale 123
   Partitioning 123
   Simple linear regression (Clyde macrofauna) 126
   Conditional tests 128
   (Holdfast invertebrates) 130
   Assumptions & diagnostics 132
   Building models 134
   Cautionary notes 136
   (Ekofisk macrofauna) 136
   Visualising models: dbRDA 142
   Vector overlays in dbRDA 143
   dbRDA plot for Ekofisk 145
   Analysing variables in sets (Thau lagoon bacteria) 147
   Categorical predictor variables (Oribatid mites) 150
   DISTLM versus BEST/BIOENV 154

5. Canonical analysis of principal coordinates (CAP)
   Key references 157
General description 157
Rationale (Flea-beetles) 157
Mechanics of CAP 159
Discriminant analysis (Poor Knights Islands fish) 160
Diagnostics 163
Cross-validation 166
Test by permutation (Anderson’s irises) 167
CAP versus PERMANOVA 169
Caveats on using CAP (Tikus Island corals) 171
Adding new samples 172
Canonical correlation: single gradient (Fal estuary biota) 175
Canonical correlation: multiple X’s 179
Sphericising variables 180
CAP versus RDA 181
Comparison of methods using SVD (Hunting spiders) 182

Acknowledgements 187
References 187
Index to mathematical notation and symbols 195
Index to data sets used in examples 197
General index 199
OVERVIEW

A. Contact details and installation of the PERMANOVA+ software

PERMANOVA+ for PRIMER was produced as a collaborative effort between Professor Marti Anderson (New Zealand Institute for Advanced Studies, Massey University, Albany, Auckland, New Zealand and current Director of PRIMER-e) and Ray Gorley & Professor Bob Clarke (formerly of PRIMER-E Ltd, Plymouth, UK). For the latest news about PERMANOVA+ or PRIMER, including details of upcoming training workshops, see: http://www.primer-e.com

Please report any bugs, technical problems, dislikes or suggestions for improvement to: tech@primer-e.com. For licensing and other general enquiries, contact Lyn Shave at: primer@primer-e.com. For queries related to the scientific methods, contact Marti Anderson at: marti@primer-e.com.

Our business postal address is:
PRIMER-e (Quest Research Limited)
ecentre, Gate 5, Oaklands Rd
Massey University Albany Campus
Auckland 0632
New Zealand
Tel: +64 (0)9 869 2230
(or you may, if you prefer, use the registered address of the company: PRIMER-e (Quest Research Limited), 67 Mahoenui Valley Rd, RD3 Albany, Auckland 0793, New Zealand)

PERMANOVA+ is an add-on package to the PRIMER v6 software program. PERMANOVA+ is not a stand-alone piece of software and can only be installed on machines that already have PRIMER v6 installed. Other requirements, in accordance with PRIMER v6, are:
- PC with Intel compatible processor.
- Windows 2000/XP/Vista or later, 32 and 64 bit.
- Memory - For Windows 2000 and XP: 128 Mb (256 Mb recommended). For Windows Vista or later: 512 Mb (1Gb recommended).
- Internet Explorer 5.01 or later is needed to install the .Net framework.
- To be able to read and write Excel worksheets you need Excel 2000, or later, installed.

Ensure that PRIMER v6 is installed before installing PERMANOVA+. PERMANOVA+ is an add-on product for installation on an individual PC, not a network server. You need to be logged on as an Administrator. The procedure is the same if you have an earlier version of PERMANOVA+ installed (e.g., a beta-test version). Note that this add-on package will not work with earlier versions of PRIMER (e.g., v5).

Insert the PERMANOVA+ CD in your CD drive. The install program will automatically run unless you have disabled AutoPlay on your drive. (If it does not run, open Windows Explorer, right click on the CD drive and select the ‘Install’ or ‘AutoPlay’ option.) The install program may first install or update the .Net framework. You will be asked for your serial number which is on the front of the PERMANOVA+ CD case. Setup may reboot your system during the installation and, as usual, it is advisable to close down all other programs before commencing installation.

This manual assumes that the user already has familiarity with the general features of PRIMER v6. Important information regarding the use of PRIMER v6 is given in the two manuals accompanying that software: a methods manual (Clarke & Warwick 2001) and a user manual/tutorial (Clarke & Gorley 2006). Information specific to PERMANOVA+ is contained in this manual and the software Help system. The latter is context sensitive: if you click on the Help button in a dialog box then you will get an appropriate help topic. You can also get into the help system by choosing the Help menu option or clicking on the help button (¶) on the Tool bar. You can then browse this system via the Contents or Index tabs in the Help window. If you are still having problems, contact us (see the top of this page), and we will be happy to help.
B. Introduction to the methods of PERMANOVA+

PERMANOVA+ is an add-on package which extends the resemblance-based methods of PRIMER to allow the analysis of multivariate (or univariate) data in the context of more complex sampling structures, experimental designs and models. The primary reasons that the methods currently available in the PRIMER package (such as ANOSIM and MDS) have become so widely used for the analysis of multivariate ecological (and other) data are their general robustness and flexibility. They are flexible due to their reliance, at heart, on a resemblance measure (a distance, dissimilarity or similarity), of which there are many to choose from, each with its own emphasis and properties. They are robust because, unlike traditional multivariate statistical methods, they make no explicit assumptions regarding the distributions of original variables, acting primarily only on the ranks of dissimilarities, and with any hypothesis-testing procedures (such as ANOSIM) using permutations to obtain \( P \)-values. The philosophy underlying the methods in PRIMER, as articulated by Clarke & Green (1988) and Clarke (1993), is to maintain this purely non-parametric rank-based approach throughout. This is generally highly appropriate, because most ecological data (being counts or other measures of abundances of species) tend to be overdispersed, with heavily right-skewed distributions (Taylor 1961, Seber 1982, McArdle et al. 1990) and a plethora of zeros (Welsh et al. 1996, Fletcher et al. 2005). Also, the number of variables (usually species or taxa) often far exceeds the number of sampling units (quadrats, cores, transects, etc.), making traditional statistical approaches either problematic or simply impossible (e.g., as in MANOVA). Even if the number of variables does not exceed the number of samples, traditional MANOVA (unlike univariate ANOVA) is not robust to violations of its assumptions, especially multivariate normality (e.g., Mardia 1971, Olsen 1974, Johnson & Field 1993).

By taking a purely non-parametric approach, however, one must also accept certain limitations. Perhaps the most important of these is the lack of any ability to partition the multivariate variation according to more complex experimental designs. More particularly, a purely non-parametric approach precludes us from: (i) partitioning variability according to one or more explanatory variables or factors; (ii) measuring or testing interactions among factors (which can only be defined by reference to main effects that have been modelled in some way); and (iii) developing explicitly quantitative models with any explanatory, discriminatory or predictive uses. Most ecological experiments, whether they be mensurative or manipulative (sensu Hurlbert 1984), do have complex sampling structures, with multiple hierarchical spatial and/or temporal scales. Few scientists embark on experiments having only one factor; experimental designs with more than one factor, including interactions, are far more common, being more efficient, informative and relevant (Underwood 1981, 1997). In addition, tests of interactions are used to examine the generality of phenomena (in time or space, e.g., Snedecor 1946, Beck 1997) and to assess environmental impact with rigour (e.g., Green 1979, Underwood 1991, 1992).

The purpose of the routines provided in the PERMANOVA+ add-on package is to maintain as much of the flexibility and robustness inherent in the methods offered by PRIMER as possible, yet to achieve the partitioning required to analyse more complex designs and to develop multivariate models. In essence, the methods in PERMANOVA+ allow multivariate data to be modelled, analysed and tested on the basis of any resemblance measure of choice, and all tests of hypotheses are done using permutation techniques\(^1\). The methods offered as part of the PERMANOVA+ add-on package do not in any way replace the existing methods in PRIMER, but rather, they provide a complementary set of tools for more complex designs and modelling. Unfortunately, you can’t get something for nothing, and so in order to take this step of partitioning, we must articulate what we mean by “variation” and, in so doing, give up the “purely non-parametric” label. In other words, we will allow the actual values in the resemblance matrix to take on meaning; they will not be replaced by their relative ranks; we will allow the chosen resemblance measure for a given analysis to dictate what we mean by “multivariate variability”. This is not too much of a tragedy, as we do still retain great flexibility (because we can still choose any resemblance measure we wish) and

\(^{1}\)Exceptions to this include: (i) the option to do Monte Carlo sampling from an asymptotic permutation distribution, in the event of there being too few possible permutations for a meaningful test in PERMANOVA and (ii) the option to use tabled \( F \) distributions in PERMDISP.
robustness (because permutation methods are being used). What it does mean, however, is that we will need to think carefully about the meaning of the resemblance measure we do choose to use (including any transformations) as this choice will play an even more important role in the analysis than is the case in other PRIMER routines.

As in PRIMER, the focus of PERMANOVA+ is generally on the analysis of ecological data, especially counts of species abundances. In addition, many of the data sets and examples are from ecology, particularly from marine or estuarine ecosystems, due to the origins of the methods arising from these contexts. However, the methods themselves are not limited, and may be applied to any situation where one wishes to analyse either univariate or multivariate data in response to either simple or complex experimental designs or models. The methods are particularly suited to the analysis of data which do not fulfil the assumptions of traditional statistical approaches because they have too many response variables, or because one or more response variables are not normally distributed. In general, the only requirement of the permutation-based techniques in any of these routines is that either the samples (observation units) themselves or the errors under a particular model be exchangeable under an appropriate null hypothesis.

A search of the ISI database of articles that have cited some of the core methods papers (e.g., Anderson 2001a, McArdle & Anderson 2001, Anderson & Willis 2003) gives a good indication of the types and variety of recent applications of these techniques. They include not only ecological examples from marine, freshwater and terrestrial environments, but also examples from diverse disciplines such as zoology (e.g., Pertoldi et al. 2006), soil science (e.g., Hoyle & Murphy 2006), genetics (e.g., Wessel & Schork 2006), psychology (e.g., Wheldon et al. 2007) and physiology (e.g., Hepner et al. 2002).

Individual details for each method are given in separate chapters, but as a general point of interest, when the methods are applied to a Euclidean distance matrix, they are equivalent to their traditional statistical counterparts (e.g., ANOVA, regression, PCA, RDA, discriminant analysis, etc.). This means that any data set for which a traditional statistical method would be appropriate can be analysed using the routines in PERMANOVA+, but with more robust statistical inferences provided by virtue of the use of permutations, rather than tables, to obtain P-values.

Virtually all of the methods provided in the PERMANOVA+ add-on package for PRIMER begin with the resemblance matrix. The first time the user opens up PRIMER after installation of the add-on, a new menu item will appear (Fig. B.1), called PERMANOVA+, from which all of the individual routines can be accessed once a resemblance matrix has been either imported or created within PRIMER.

The individual routines are:

- **PERMANOVA**, for the analysis of univariate or multivariate data in response to factors, groups or treatments in an experimental design;
- **PERMDISP**, to measure and test homogeneity of multivariate dispersions among a priori groups;
- **PCO**, to provide an unconstrained ordination of multivariate data on the basis of a chosen resemblance measure;
- **DISTLM**, for the analysis of univariate or multivariate data in response to continuous (or categorical) predictor variables (such as environmental variables), a distance-based regression approach, with various options for model selection;
- **dbRDA**, for the ordination and visualisation of fitted models (such as from DISTLM); and
- **CAP**, to use multivariate data to discriminate among a priori groups or to predict values along a gradient; also to do distance-based canonical correlation.

Although the methods offered by the PERMANOVA+ package may not be purely non-parametric, they still make no explicit assumptions regarding either the distribution of the original variables or of the resemblance values themselves, as a consequence of their use of permutation techniques. The methods may be considered, therefore, to be “semi-parametric” in some sense.
B. Introduction

Fig. B.1. The new menu item for the PERMANOVA+ add-on to PRIMER v6.

Fig. B.2 shows a schematic diagram and flowchart of the routines in the context of their use. Alongside each method (inside an oval shape) is also shown the corresponding traditional statistical method that would be obtained in the event that the resemblance matrix used to begin with contained Euclidean distances. The diagram does not show all of the many other routines available from within PRIMER, which are well described elsewhere (Clarke & Gorley 2006). Many of these may be used in combination with the routines offered in PERMANOVA+. For example, in the absence of *a priori* groups, PRIMER routines such as CLUSTER and the associated SIMPROF tests may be appropriate, and the use of exploratory tools, such as draftsman plots, to examine distributions and potential multicollinearity among predictor variables, are also highly recommended before proceeding with modelling (Fig. B.2).
Fig. B.2. Schematic flowchart showing the methods in PERMANOVA+ and their traditional counterparts (in ovals), obtained if Euclidean distances are used.
C. Changes from DOS to PERMANOVA+ for PRIMER

1) All of the original DOS routines have been fully re-written, translated from their original FORTRAN into the new Microsoft .NET environment, as used by PRIMER v6. This gives the software a fully modern Windows user interface.

2) The integration of the add-on package with the current PRIMER v6 is seamless and complete, with true multi-tasking, providing a greatly enhanced ability to both visualise and formally analyse multivariate data from within a single package.

3) There are no fixed size constraints on data matrices, the number of factors, or group sizes for any analysis. The limitations are imposed only by the total available memory.

4) Other major advantages of PRIMER v6 shared by the PERMANOVA+ add-on include: workspaces and explorer tree navigation; multiple input formats (including from Excel, or 3-column format); label matching (avoiding the need to worry about the order of data in worksheets, provided the labels used are consistent); data handling operations (sorting, ranking, merging, transforming, summing, etc.); easy cutting and pasting of text results or graphics; multiple output formats (.txt, .rtf, .xls, .jpg, .tif, .emf, etc.); factors on samples (or indicators on variables) can be read in from (or saved to) Excel or text formats along with the data; data can be averaged or summed according to factor levels or combinations of them; levels of multiple factors can be identified with symbols and labels on enhanced 2-d and 3-d plots. Importantly, the package offers a great range of pre-treatment transformations or standardisations, as well as the additional choice of more than 50 resemblance measures.

5) The new PERMANOVA routine can now be used to analyse any balanced or unbalanced experimental design, either with or without covariates, for fixed, random or mixed models, either with or without hierarchical nesting. There is no limit on the number of factors that can be analysed. A new algorithm ensures correct derivation of expected mean squares (EMS) and correct construction of pseudo-F ratios for each term in the model, with tests done using permutations.

6) A new ‘Design’ type of worksheet file (*.ppd) is introduced into PRIMER with this add-on package. Its sole function is to outline the correct experimental design for subsequent analysis by PERMANOVA. It is linked to the resemblance matrix to be analysed, and so utilises important information regarding the available factors and their levels accordingly. When created, the user specifies whether factors are fixed or random and whether they are nested within one or more other factors. A new feature is the ability to test the design file using dummy data. This will output EMS’s and will identify the numerator and denominator mean squares and degrees of freedom for each term in the model. This allows the user to check a specified design and to consider the potential power for individual terms in the model, even before data are collected.

7) Pair-wise tests in PERMANOVA are now done taking into account the position of the factor of interest within the full experimental design. In the DOS version, the data corresponding only to the two levels of the factor being compared were extracted and considered in isolation for each test. The Windows version, however, retains the full experimental design and treats the pair-wise test effectively as a 1df contrast between these two levels. This ensures that the correct denominator is used for all tests and that the structure of the full design, including any necessary conditioning, is maintained (which is especially important for unbalanced designs or when there are covariates).

8) Terms which traditionally have had “no test” in PERMANOVA, due to the complexity of the design (usually caused by multiple random effects), are now tested using correct linear combinations of appropriate mean squares.

9) Pooling or the exclusion of one or more individual terms in the model is now possible in PERMANOVA, as is the analysis of models lacking replication, such as randomised blocks, split-plots and repeated measures. The order in which terms are fitted can also be specified explicitly by the user (important in either unbalanced designs or designs with covariates where Type I SS are used).

10) The user can now choose the type of sums of squares to use (Type I, II or III SS) in PERMANOVA. These types are equivalent for balanced designs, but not for unbalanced designs or designs with covariates. An enhanced algorithm even allows for imbalance in the actual cell
structure of the model, which is especially useful in the context of asymmetrical experimental
designs (e.g., Underwood 1994, Glasby 1997).

11) The new PERMANOVA routine also allows the user to specify and test particular 1 df
contrasts of interest (such as a treatment versus a number of different control groups) as well as the
interaction of such contrasts with other factors.

12) The method used to fit covariables in the DOS version of PERMANOVA was a naïve approach
that did not adjust the EMS’s of the other terms in the ANOVA model accordingly. Thus, the
results obtained from the two versions will differ for a given (balanced) data set if there are
covariables (the new Windows version is correct).

13) In PERMDISP, the user now has the option of whether distances are calculated to centroids or
to spatial medians and also whether $P$-values are obtained using permutations or tabled values for
the $F$ ratio (the default is to use centroids and permutations, as recommended in Anderson 2006).
This speeds up the routine considerably, as clearly it is not necessary to perform all four possible
combinations of these approaches, as is done in the DOS version.

14) The PCO routine yields graphical output that has all of the facilities that similar graphical
objects (like PCA plots) in PRIMER possess, including choices for symbols and labels, the ability
to observe and spin 3-d plots, and to superimpose bubbles. A new graphical feature offered with
the PERMANOVA+ add-on is the ability to overlay vectors that correspond to either individual or
partial correlations between variables in a worksheet and ordination axes. These may be the
original variables, transformed variables, or some other variables of interest.

15) A new tool provided in PERMANOVA+ allows the user to calculate distance matrices among
centroids identified by factors (or cells consisting of combinations of factors) in the space of a
chosen resemblance measure. From this, centroids can then be further analysed or viewed in
ordinations (handy for complex designs), also enabling the visualisation of relative effect sizes.

16) The DOS version of DISTLM and its cousin, DISTLM_forward, have been dramatically
enhanced in the DISTLM routine of PERMANOVA+ for PRIMER. The DOS version only
performed forward selection on the basis of the simple $R^2$ criterion, whereas the new DISTLM is a
true multivariate regression modelling tool on the basis of a chosen resemblance measure. The user
may now choose among four model selection procedures (forward selection, backwards
elimination, step-wise selection, or the “best” of all possible combinations) on the basis of any of 4
model selection criteria ($R^2$, adjusted $R^2$, and multivariate analogues of AIC or BIC). In addition, if
one specifies “best” as the model selection procedure, then one may also choose how many of the
top solutions to view and also whether to view relatively brief or more detailed results.

17) The new DISTLM routine in PERMANOVA+ uses the existing utility in PRIMER of
specifying sets of variables using indicators, so that these may be kept together during model
selection (e.g., spatial, temporal and environmental variables, Anderson & Gribble 1998). Identifying
sets of variables is also useful for binary indicators or matrices that code for a
categorical variable or factor in model selection or fitting.

18) DISTLM also allows the user to explicitly identify particular predictor variables (or sets of
them) to be included or excluded in the model (or in the model selection activity), and optionally
can also specify the order in which the variables (or sets) are to be fit.

19) The relationship between patterns in a resemblance matrix and a set of predictor variables (such
as environmental variables) can be observed directly using a constrained ordination: distance-based
redundancy analysis (dbRDA). This is a completely new routine. Vector overlays of the predictor
variables are superimposed automatically. Like PCO, the dbRDA ordination routine also takes full
advantage of all of the tools offered in other PRIMER Windows graphics (e.g., choice of which
axes to view, in two or in three dimensions, ample choice of labels and symbols, etc.). DISTLM is
also directly linked to dbRDA, providing an instant constrained ordination, if desired, for any given
fitted model.

\footnote{Unfortunately, when a covariable is added to an ANOVA model, then the terms in the model are no longer
orthogonal (independent) of one another, even if the design is balanced.}
20) The CAP routine in PERMANOVA+ for PRIMER includes graphical output as well as the usual text output and diagnostics that would be familiar to users of the former CAP program in DOS. This graphical output (as for PCO or dbRDA) has all of the usual advantages of other PRIMER ordination graphics, including the new general vector overlay tool available in PERMANOVA+.

21) A brand new feature of the CAP routine is the ability to add new samples into the CAP model and “see where they fall”. This is highly useful for model validation and for classifying new samples whose group identity is unknown. The positions of the new samples are given in the canonical space in the text output file and visually in the CAP ordination graphic. If the CAP analysis focused on discriminating groups, then the new samples are also allocated to a new group based on their positions in the canonical space.
The typographic conventions for this manual follow those used by Clarke & Gorley (2006) for PRIMER v6, as follows:

Text in **bold** indicates the menu items that need to be selected,

> denotes cascading sub-menu items, tab choices, dialog boxes or sub-boxes,

● denotes a button entry in a dialog box (so-called ‘radio buttons’ – only one can be selected),

✓ indicates a tick in the specified box (so called ‘check boxes’ – either on or off),

text inside a cartouche is an instruction to select the suggested entry (e.g., filename) or actually to type it in, and

( ) & ( ) & ( ) indicate several steps that need to be carried out in the one box, where brackets are used naturally to split up the different components of the dialog.

For example:

**Analyse>Resemblance>(Analyse between●Samples) & (Measure●Bray-Curtis similarity) & (✓ Add dummy variable>Value: 1)**

is an instruction to select the main menu item Analyse, the sub-menu item Resemblance, and analyse between samples using Bray-Curtis similarity, adding a dummy species with value 1 for all samples, prior to computing similarities (see Clarke *et al.* 2006c). The dialog this corresponds to is shown below in Fig. D.1.

![Fig. D.1. Example of the directions: Analyse>Resemblance>(Analyse between●Samples) & (Measure●Bray-Curtis similarity) & (✓ Add dummy variable>Value: 1).](image)

The PRIMER CD contains a number of example data files needed for this manual. The installation will place a number of sub-directories (e.g., BorneoBirds, HoldNZ, FishNZ, VictAvi, MedMoll, etc.) into an ‘Examples add-on’ directory (the default installation location is C:\Program Files\PRIMER-6\PRIMER 6\Examples add-on). These files can also be read directly from the CD as you run the package. You might find it convenient, however, to copy them to a higher-level data
area on your hard disk. In what follows, it is assumed that they are in a top-level directory `C:\Examples add-on`. The various sub-directories contain data files (generally in .pri or .xls formats) that correspond to particular examples used in this manual. Examples from PRIMER v6 (located in the Examples v6 directory) will also be used and referred to from time to time. Details of these examples are provided in the User Manual/Tutorial for PRIMER v6 (Clarke & Gorley 2006).

At the end of the manual there is an index of the example data sets included, the name of the files in which they are each contained, the topics for which they are used to exemplify certain points, and page numbers identifying where they are referenced in the manual.

Each chapter is devoted to a particular routine and begins with a few key references followed by a general description of the method. The methods should be cited using these key references, as appropriate. The general citation for either the software itself or the manual is found on the inside front cover of this manual.

Descriptions of the methods and the examples provided are deliberately conceptual in nature and avoid, as far as possible, the use of mathematical notation and matrix algebra. A complete understanding of this matrix algebra, wherever it appears, is by no means necessary, however, for the successful use, understanding and interpretation of results of all of the routines provided by PERMANOVA+, in much the same way as it is not necessary to understand the finer mathematical details of non-metric MDS in order nevertheless to implement and interpret the output from the MDS routine in PRIMER. Rather, such information is provided for users with a mathematical bent who are interested in some of the underlying nuts and bolts, particularly with respect to the links, commonalities and differences among the different routines.

Cartouches in the margins identify new sections within each chapter, as listed in the table of contents. In addition to the index of occurrences of data sets, there is also an index of mathematical notation and symbols as well as a general index of topics at the end of the manual.
GUIDE TO SOFTWARE AND STATISTICAL METHODS

1. Permutational ANOVA and MANOVA (PERMANOVA)

Key references

General description
PERMANOVA is a routine for testing the simultaneous response of one or more variables to one or more factors in an analysis of variance (ANOVA) experimental design on the basis of any resemblance measure, using permutation methods. It is assumed that the user has relevant knowledge of multi-factorial ANOVA, which has the same basic logic in multivariate as in univariate analysis (see Underwood 1981, 1997 and Quinn & Keough 2002), and an understanding of what it means to test a multivariate hypothesis (see Clarke 1993). A more complete description of the method is given in Anderson (2001a) and McArdle & Anderson (2001). These papers merely elaborate the essential idea of partitioning for dissimilarity matrices which was originally (to our knowledge) presented by Brian McArdle (McArdle 1990, 1994) and which has also been articulated by Pillar & Orłóci (1996), Legendre & Anderson (1999) and Gower & Krzanowski (1999). In essence, the routine performs a partitioning of the total sum of squares according to the full experimental design specified by the user, including appropriate treatment of factors that are fixed or random, crossed or nested (hierarchical), and all interaction terms. The routine will correctly calculate an appropriate distance-based pseudo- statistic for each term in the model, based on the expectations of mean squares (EMS), in a fashion that is directly analogous to the construction of the statistic for multi-factorial univariate ANOVA models (Cornfield & Tukey 1956, Hartley 1967, Rao 1968). P-values are obtained using an appropriate permutation procedure for each term, and the user can specify whether permutation of raw data or residuals under either a full or reduced model are to be used (Anderson 2001b, Anderson & ter Braak 2003). Correct P-values may also be obtained through Monte Carlo random draws from the asymptotic permutation distribution (Anderson & Robinson 2003) in the event that too few permutations are available for a given test.

In addition to the main overall PERMANOVA partitioning and tests, the routine will also perform a posteriori pair-wise comparisons among levels of factors, including within individual levels of other factors (or cells) in the case of significant interaction terms. Other important features of the PERMANOVA routine include: (i) catering for unbalanced designs, including a choice regarding the type of sum of squares to be used for the partitioning; (ii) pooling or the exclusion of individual terms from a model; (iii) choice to include one or more quantitative covariates in the model; (iv) contrasts; (v) analysis of designs lacking replication; and (vi) analysis of asymmetric designs.

Before using PERMANOVA, the user should have a fairly solid grasp of the basic issues and logic in experimental design (some good reference textbooks on this topic include Mead 1988, Snedecor & Cochran 1989, Winer et al. 1991, Underwood 1997 and Quinn & Keough 2002). Know your hypotheses and know your design! It is also best to store and import the factors and their levels along with the data in order to avoid making mistakes. Otherwise, use the tools available in PRIMER to either create or import factors and their levels (see chapter 2 in Clarke & Gorley 2006).

Partitioning
We shall begin by considering the balanced one-way (single factor) ANOVA experimental design. A factor is defined as a categorical variable that identifies several groups, treatments or levels which we wish to compare. Imagine that we have one factor with a groups (or levels) and n observations (samples) per group for a total of N = a x n samples. For each sample, we have recorded the values for each of p different variables. Recall that in univariate analysis of variance, the total sum of squares (SSR, the sum of squared deviations of observations from the overall mean)

---

4 The total sum of squares is understood here to be defined by reference to the distance/dissimilarity measure of choice. It will only correspond to the traditional univariate total sum of squares when one variable is being analysed and Euclidean distance has been chosen as the basis of the analysis.

5 The word sample will be used throughout this manual in the manner that ecologists, and not statisticians, have come to understand the word. A sample shall mean a single unit used for sampling, such as a core, a transect, or a quadrat. This is consistent with the use of this word in PRIMER.
1. PERMANOVA

is partitioned into two parts that are meaningful for testing hypotheses about group differences: the within-group (or residual) sum of squares ($SS_{res}$, the sum of squared deviations of observations from their own group mean) and the among-group sum of squares ($SS_a$, the sum of squared deviations of group means from the overall mean). A directly analogous partitioning is done in multivariate space by PERMANOVA.

PERMANOVA may be thought of as a method that takes a geometrical approach to MANOVA (Edgington 1995). Let each of the $p$ variables be a dimension, and each of the $N$ samples be represented by points in the $p$-dimensional space according to the values they take for each variable along each dimension. Now, the simplest of all multivariate systems has only $p = 2$ variables. It is good to consider this situation, because it is easy to draw just 2 dimensions! Now, imagine that there were $n = 10$ replicate samples in each of $a = 3$ groups, as shown in Fig. 1.1.

![Fig. 1.1. Plot of a hypothetical data set with $p = 2$ variables (dimensions) and $n = 10$ replicate samples in each of $a = 3$ groups. The three groups are identified by different symbols in the plot.](image)

The whole set of $N = 30$ samples taken together create a data cloud which is centred on a point called the overall centroid. For a Euclidean system, this is obtained as the arithmetic average for each of the (in this case 2) variables. Similarly, each of the groups also has its own group centroid, located in the centre of each of the clouds of points identified for each group.

![Fig. 1.2. Plots of the hypothetical data set from Fig. 1.1 showing the geometric partitioning.](image)

Just as in univariate ANOVA, we can consider the distance of any given point (sample) from the overall centroid in this space as being made up of two parts: the distance from the point to its group centroid (Fig. 1.2a) plus the distance from the group centroid to the overall centroid (Fig. 1.2b). This is the essence of the geometric approach to MANOVA in Euclidean space. We can calculate the sums of squares as:
$SS_T = \text{the sum of squared distances from the samples to the overall centroid,}$

$SS_{Res} = \text{the sum of squared distances from the samples to their own group centroid, and}$

$SS_A = \text{the sum of squared distances from the group centroids to the overall centroid.}$

The well-known univariate ANOVA identity: $SS_T = SS_A + SS_{Res}$ also holds for this geometric conception of MANOVA in Euclidean space. Verdonschot & ter Braak (1994) and Legendre & Anderson (1999) also remark on how these sums of squares are equal to the sum of the individual univariate sums of squares for each of the separate variables if Euclidean distance is used.

This partitioning is fine and perfectly valid for Euclidean distances. But what happens if we wish to base the analysis on some other dissimilarity (or similarity) measure? This is important because Euclidean distance is generally regarded as inappropriate for analysing ecological species abundance data\(^6\), especially because of its lack of emphasis on species composition (e.g., Faith et al. 1987, Legendre & Legendre 1998, Clarke 1993, Clarke et al. 2006c). Other measures, such as Bray-Curtis (Bray & Curtis 1957), Kulczynski (see Faith et al. 1987), Jaccard (see Legendre & Legendre 1998), binomial deviance (Anderson & Millar 2004), Hellinger (Rao 1995) or a modified Gower measure (Anderson et al. 2006), may be more appropriate, in different situations, for analysing community data. Although analyses on the basis of some measures (such as chi-squared or Hellinger) can be achieved by applying Euclidean distances to data that have been transformed in a particular way (see Legendre & Gallagher 2001 for details), we wish to retain the full flexibility of methods like ANOSIM or MDS and allow the analysis to be based on any reasonable distance or dissimilarity measure of choice.

Unfortunately, for many resemblance measures favoured by ecologists, we cannot easily calculate distances to centroids. The reason for this is that the sample centroid (the point lying in the centre of the data cloud) in the space defined by these measures will not be the same as the vector of arithmetic averages for each variable. It is actually quite unclear just how one could go about calculating centroids for the majority of these non-Euclidean measures on the basis of the raw data alone. Thus, for these situations, we shall instead rely on what is known as Huygens’ theorem\(^7\) (Fig. 1.3). This theorem states (for Euclidean space) that the sum of squared distances from individual points to their group centroid is equal to the sum of squared inter-point distances, divided by the number of points in the group (e.g., Legendre & Anderson 1999, see Appendix B therein).

![Fig. 1.3. Schematic diagram of Huygens’ theorem.](image)

This theorem appears to be very simple, and it is. In fact, it is perfectly safe to calculate these two quantities using Euclidean distances and you are encouraged to try a simple example (say in two dimensions) to prove it to yourself. Importantly, what it means for us is that we can use the inter-

---

\(^6\) See Warton & Hudson (2004) for an alternative point of view and counter-argument to this assumption.

\(^7\) Christian Huygens (1629-1695) was a Dutch mathematician, astronomer and physicist, famous for, among other things, patenting the first pendulum clock.
1. PERMANOVA

Point distances (or dissimilarities) alone in order to calculate the sums of squares, without ever having to calculate the position of the centroids at all! Therefore, for the cases where we cannot calculate centroids directly (i.e., for most dissimilarity measures), we can use the inter-point values instead to do the partitioning. The only properties required of our dissimilarity measure in order to use this approach are that it be non-negative, symmetric (the distance or dissimilarity from point A to point B must be the same as from point B to point A), and that all self dissimilarities (from point A to itself) be zero. Virtually all measures that are worth using possess these three properties.

We can now consider the structure of a distance/dissimilarity matrix and how sums of squares for a one-way multivariate ANOVA partitioning would be calculated (Fig. 1.4).

If we let $d_{ij}$ be the dissimilarity (or distance) between sample $i$ and sample $j$, then the total sum of squares is the sum of the inter-point dissimilarities among all samples, divided by $N$:

$$SS_T = \frac{1}{N} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij}^2$$

and the residual (within-group) sum of squares (assuming, for now, a balanced design having an equal sample size of $n$ per group) is:

$$SS_{Res} = \frac{1}{n} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij}^2 \cdot \omega_{ij},$$

where $\omega_{ij}$ takes the value of 1 if samples $i$ and $j$ are in the same group, otherwise it takes the value of zero. This amounts to adding up the squares of all the dissimilarities between samples that occur within the same group. These quantities are shown schematically in Fig. 1.4. The among-group sum of squares can also be calculated directly or, more simply, as the difference: $SS_A = SS_T - SS_{Res}$.

Partitioning of distance matrices having Euclidean metric properties according to ANOVA experimental designs has been discussed previously by Edgington (1995), Pillar & Orlóci (1996) and Excoffier et al. (1992), whereas Gower & Krzanowski (1999) extended this idea to semi-metric dissimilarities having only the properties of symmetry (i.e., $d_{ji} = d_{ij}$) and that $d_{ij} \geq 0$ and $d_{ii} = 0$ for all samples.

A related point is that PERMANOVA (or any of the other methods that partition variability in the routines offered by the PERMANOVA+ add-on) does not suffer from the problems recently identified by Legendre et al. (2005) for the Mantel and partial Mantel test (Mantel 1967, Smouse et al.

---

*The mathematical and geometric properties of dissimilarity measures are reviewed by Gower & Legendre (1986).*
al. 1986). Confusion might arise because Legendre et al. (2005) used the words “variation partitioning on distance matrices” to describe the general Mantel approach. PERMANOVA, however, is not a Mantel test. The essential issue regarding the partial Mantel approach (known for some time to be potentially problematic, see Dutilleul et al. 2000, Legendre 2000, Raufaste & Rousset 2001 and Rousset 2002) is that it works on “unwound” distance matrices, where inter-point distance values are treated independently as a single vector. In contrast, the methods in the PERMANOVA+ add-on all work on the distance/dissimilarity matrix directly but, importantly, they retain its inherent structure; the values in the matrix are not unwound nor are they treated or modelled as independent of one another. The PERMANOVA+ methods are therefore directly akin to the so-called “canonical partitioning” methods referred to by Legendre et al. (2005), and are correct for partitioning and analysing the actual variability inherent in multivariate data clouds.

Once the partitioning has been done we are ready to calculate a test statistic associated with the general multivariate null hypothesis of no differences among the groups. For this, following R. A. Fisher’s lead, a pseudo-$F$ ratio is defined as:

\[ F = \frac{SS_A/(a-1)}{SS_{Res}/(N-a)} \]  

(1.3)

where \((a-1)\) are the degrees of freedom associated with the factor and \((N-a)\) are the residual degrees of freedom. It is clear here that, as the pseudo-$F$ statistic in (1.3) gets larger, the likelihood of the null hypothesis being true diminishes. Interestingly, if there is only one variable in the analysis and one has chosen to use Euclidean distance, then the resulting PERMANOVA $F$ ratio is exactly the same as the original $F$ statistic in traditional ANOVA\(^9\) (Fisher 1924). In general, however, the PERMANOVA $F$ ratio should be thought of as a “pseudo” $F$ statistic, because it does not have a known distribution under a true null hypothesis. There is only one situation for which this distribution is known and corresponds to Fisher’s traditional $F$ distribution, namely: (i) if the analysis is being done on a single response variable and (ii) the distance measure used was Euclidean distance and (iii) the single response variable is normally distributed. In all other cases (multiple variables, non-normal variables and/or non-Euclidean dissimilarities), all bets are off! Therefore, in general, we cannot rely on traditional tables of the $F$ distribution to obtain a $P$-value for a given multivariate data set.

Some other test statistics based on resemblance measures (and using randomization or permutation methods to obtain $P$-values, see the next section) have been suggested for analysing one-way ANOVA designs (e.g., such as the average between-group similarity divided by the average within-group similarity as outlined by Good (1982) and Smith et al. (1990), see also all of the good ideas in the book by Mielke & Berry (2001) and references therein). Unlike pseudo-$F$, however, these can be limited in that they may not necessarily yield straightforward extensions to multi-way designs.

An appropriate distribution for the pseudo-$F$ statistic under a true null hypothesis is obtained by using a permutation (or randomization) procedure (e.g., Edgington 1995, Manly 2006). The idea of a permutation test is this: if there is no effect of the factor, then it is equally likely that any of the individual factor labels could have been associated with any one of the samples. The null hypothesis suggests that we could have obtained the samples in any order by reference to the groups, if groups do not differ. So, another possible value of the test statistic under a true null hypothesis can be obtained by randomly shuffling the group labels onto different sample units. The random shuffling of labels is repeated a large number of times, and each time, a new value of pseudo-$F$, which we will call pseudo-$F^*$, is calculated (Fig. 1.5). Note that the samples themselves have not changed their positions in the multivariate space at all as a consequence of this procedure.

\(^9\) The original simple Mantel test (Mantel 1967) to relate two distance matrices, from which several of the PRIMER routines (RELATE, BIOENV, BEST, BVSTEP and 2-stage MDS) all drew some inspiration, is valid and does have utility in appropriate applications, as pointed out by Legendre et al. (2005).

\(^{10}\) In fact, a nice way to familiarise oneself with the routine is to do a traditional univariate ANOVA using some other package and compare this with the outcome from the analysis of that same variable based on Euclidean distances using PERMANOVA.
it is only that the particular labels associated with each sample have been randomly re-allocated to them (Fig. 1.5).

If the null hypothesis were true, then the pseudo-$F$ statistic actually obtained with the real ordering of the data relative to the groups will be similar to the values obtained under permutation. If, however, there is a group effect, then the value of pseudo-$F$ obtained with the real ordering will appear large relative to the distribution of values obtained under permutation.

![Diagram](image)

**Fig. 1.5.** After calculating an observed value of the test statistic, a value that might have been obtained under a true null hypothesis is calculated by permuting labels.

The frequency distribution of the values of pseudo-$F^x$ is discrete: that is, the number of possible ways that the data could have been re-ordered is finite. The probability associated with the test statistic under a true null hypothesis is calculated as the proportion of the pseudo-$F^x$ values that are greater than or equal to the value of pseudo-$F$ observed for the real data. Hence,

\[
P = \frac{(\text{No. of } F^x \geq F) + 1}{(\text{Total no. of } F^x) + 1}
\]

(1.4)

In this calculation, we include the observed value as a member of the distribution, appearing simply as “+1” in both the numerator and denominator of (1.4). This is because one of the possible random orderings of the data is the ordering we actually got and its inclusion makes the $P$-value err slightly on the conservative side (Hope 1968) as is desirable.

For multivariate data, the samples (either as whole rows or whole columns) of the data matrix (raw data) are simply permuted randomly among the groups. Note that permutation of the raw data for multivariate analysis does not mean that values in the data matrix are shuffled just anywhere. A whole sample (e.g., an entire row or column) is permuted as a unit; the exchangeable units are the labels associated with the sample vectors of the data matrix (e.g., Anderson 2001b). For the one-way test, enumeration of all possible permutations (re-orderings of the samples) gives a $P$-value that yields an exact test of the null hypothesis. An exact test is one where the probability of rejecting a true null hypothesis is exactly equal to the a priori chosen significance level ($\alpha$). Thus, if $\alpha$ were chosen to be 0.05, then the chance of a type I error for an exact test is indeed 5%.

In practice, the possible number of permutations is very large in most cases. A practical strategy, therefore, is to perform the test using a large random subset of values of pseudo-$F^x$, drawn randomly, independently and with equal probability from the distribution of pseudo-$F^x$ for all possible permutations. PERMANOVA does not systematically do all permutations, but rather draws a random subset of them, with the number to be done being chosen by the user. Such a test is still exact (Dwass 1957). However, separate runs will therefore result in slightly different $P$-values.

---

11 Equivalently, permutations can be achieved by the simultaneous re-ordering of the rows and columns of the resemblance matrix (e.g., see Fig. 2 in Anderson 2001b).
for a given test, but these differences will be very small for large numbers of permutations (e.g., in the 3rd decimal place for 9999 permutations) and should not affect interpretation.

Once again, as a rather nice intuitive bonus, PERMANOVA done on one response variable alone and using Euclidean distance yields Fisher’s traditional univariate F statistic. So, PERMANOVA can also be used to do univariate ANOVA but where P-values are obtained by permutation (e.g., Anderson & Millar 2004), thus avoiding the assumption of normality. Note also that if the univariate data do happen to conform to the traditional assumptions (normality, etc.), then the permutation P-value converges on the traditional normal-theory P-value in any event.

Recall that for traditional one-way ANOVA, the assumptions are that the errors are independent, that they are normally distributed with a mean of zero and a common variance, and that they are added to the treatment effects. In the case of a one-way analysis, the PERMANOVA test using permutations assumes only that the samples are exchangeable under a true null hypothesis12. The assumption of exchangeability is tantamount to assuming that the multivariate observations (samples) are independent and identically distributed (i.i.d.) under a true null hypothesis. Thus, although there are no explicit assumptions regarding the distributions of the original variables (they are certainly not assumed to be normally distributed), independence and homogeneity of dispersions (in the space of the resemblance measure) are directly implied by the permutation procedure. Clearly, if samples have very different dispersions in different groups, then they are not really exchangeable. Also, if the samples are unequally correlated with one another (e.g., temporally or spatially), then randomly shuffling them will destroy this kind of inherent structure. In contrast, it is not expected that the individual variables which have been measured on the same samples (in the multivariate case) are independent of one another and this is not assumed. When permutations are done, the values for different variables within a sample are kept together as a unit, so whatever correlation structure there might be among the variables is not altered under permutation.

**Fig. 1.6.** PERMANOVA will be sensitive to differences in dispersions (left) but not differences in correlation structure (right) among groups.

PERMANOVA, like ANOSIM (Clarke 1993), will be sensitive to differences in dispersion among groups (Fig. 1.6). Indeed, the construction of pseudo-F ratios in PERMANOVA uses pooled estimates of within-group variability, so homogeneity of multivariate dispersions is also implicit in the partitioning. A separate test for homogeneity of dispersions, using the PERMDISP routine (see chapter 2) can be done prior to performing PERMANOVA (or, indeed, to investigate the null hypothesis of homogeneity in its own right). However, we consider that a non-significant result from PERMDISP is not strictly necessary to achieve prior to using PERMANOVA. It is likely that PERMDISP will detect differences in dispersion that, in many cases, are not substantial enough to

---

12 Exchangeability of multivariate observations (samples) is assured if we have done a random allocation of sample units to groups or treatments a priori (Fisher 1935). For observational studies, where we cannot do this (i.e., the groups already occur in nature and we draw a random sample from them), we must assume exchangeability under a true null hypothesis (Kempthorne 1966).
“de-rail” (i.e. to inflate the error rates of) the PERMANOVA test\textsuperscript{13}. This is analogous to the situation in univariate analysis; traditional ANOVA is quite robust to many forms of heterogeneity, especially with large sample sizes (Box 1953). Instead, we can consider the homogeneity of dispersions to be included as part of the general null hypothesis of “no differences” among groups being tested by PERMANOVA (even though the focus of the PERMANOVA test is to detect location effects). If significant heterogeneity were detected by PERMDISP and differences among groups were also detected using PERMANOVA, then the latter could have been caused by differences in location, differences in dispersion, or some combination of the two. Thus, performing a test using PERMDISP, as well as examining the average within and between-group dissimilarities and the position of samples from different groups in unconstrained ordination plots (MDS or PCO), will help to uncover the nature of any differences among groups detected by PERMANOVA.

Unlike many of the traditional MANOVA test statistics (e.g., Mardia et al. 1979, Seber 1984), PERMANOVA will not be sensitive, however, to differences in correlation structure among groups (Fig. 1.6). See Krzanowski (1993) for a permutation test designed to compare correlation structures among variables across different groups.

Our first real example comes from a study by Gray et al. (1990), who studied changes in community structure of soft-sediment benthic macrofauna in relation to oil-drilling activity at the Ekofisk oil platform in the North Sea. These data consist of \( p = 174 \) species sampled by bottom grabs at each of \( N = 39 \) stations. The stations were placed roughly along five transects radiating out from the centre of the oil platform. Stations have been grouped with labels according to their distance from the oil platform as A (> 3.5 km), B (1 km – 3.5 km), C (250 m – 1 km) and D (< 250 m). The data are located in the file ekma.pri in the ‘Ekofisk’ folder of the ‘Examples v6’ directory.

\textsuperscript{13} Certainly a worthy topic for future study is to discover the conditions under which PERMANOVA will show inflated rates of either Type I or Type II error in the face of heterogeneity in the distributions of multivariate samples among groups.
Open the file by selecting **File > Open** from within PRIMER and using the browser. Next, select **Edit > Factors** to view the factor labels associated with each sample. See chapter 2 of Clarke & Gorley (2006) for detailed information concerning creating, importing and editing factors to identify sample groups within PRIMER. Of interest here is to test the null hypothesis of no differences among the communities inhabiting the benthic habitats in these four different groups.

First, we may wish to visualise the relationships among the samples in terms of a relevant resemblance measure, using ordination. A well-known robust procedure for doing this is non-metric multidimensional scaling (MDS, Shepard 1962, Kruskal 1964, Kruskal & Wish 1978, Minchin 1987). Produce a resemblance matrix among the samples by selecting **Analyse > Pre-treatment > Transform (overall)** > **Transformation**: fourth-root, followed by **Analyse > Resemblance** > (Analyse between•Samples) & (Measure•Bray-Curtis similarity). Next, produce an MDS plot by selecting **Analyse > MDS** and click ‘OK’ with all of the default options. Once the 2-dimensional graph is in view, show the samples according to their factor labels by selecting **Graph > Data labels & symbols** > (Labels > ✓Plot > ✓By factor Dist) and by removing the ✓ from the (Symbols > Plot) box (Fig. 1.7). The resulting ordination plot suggests that the communities are fairly distinct in the different distance groups, and also that they tend to occur along a gradient, from those lying closest to the platform (group D) to those lying furthest away (group A).

We shall formally test the hypothesis of no differences in community structure among the four groups (where, in this case, “differences in community structure” is defined by the Bray-Curtis measure on fourth-root transformed data) using PERMANOVA. For any analysis using PERMANOVA, there are necessarily two steps involved. First, one must create a design file, which provides all of the necessary information for the partitioning to be done according to the correct factors and experimental design. The design file is unique to the PERMANOVA+ add-on package, with its own special icon: 📝. It has a direct link to the factor information associated with the samples in the resemblance matrix. (The factor information associated with the raw data matrix is inherited by any resemblance matrix produced from the data.) Once an appropriate design file has been created, the second step is to run the PERMANOVA analysis itself on the resemblance matrix, providing the name of the design file. Although the above general description of the method has used a distance or dissimilarity matrix at its base, PERMANOVA will also perform a correct analysis when given a resemblance matrix of similarities. For example, in the present case, the Bray-Curtis dissimilarity is defined simply as 100 minus the Bray-Curtis similarity. This is done automatically and the user need not perform any extra steps.

To create a design file for the Ekofisk macrofauna example (Fig. 1.8), use the explorer tree in the left-hand panel and click on the resemblance matrix, then select **PERMANOVA+ > Create PERMANOVA design**. In the initial dialog box, entitled ‘PERMANOVA design properties’, specify (Title: Ekofisk oilfield macrofauna) & (Number of factors: 1). This will bring up a new design worksheet file with one row for each factor (so only one row in the present example) and four columns. The first column is used to indicate the name of each factor, the second column is for specifying any factors within which it may be nested, the third column is for specifying whether the factor is fixed or random and the fourth column is for specifying specific contrasts among levels of the factor. You can either begin by typing the name of the factor inside the cell of the first column, or by double clicking inside this cell and a list of all the factors associated with the resemblance matrix will be brought up automatically for you to choose from. For this data set, select the factor Dist, as shown (Fig. 1.8). This factor is not nested within any other factor (leave the cell in the second column blank), it is fixed (select ‘Fixed’ in the third cell, which is the default), and we do not wish at present to specify any specific contrasts (leave the cell in the fourth column blank).

Once the design file is created, you may change its name, edit its properties and so on, as for any other PRIMER file within the workspace. For example, select **Edit** from the main menu to see how rows (i.e., factors) may be inserted, moved or deleted. Rows of the design file can also be edited by selecting them directly, or by right-clicking on the design file itself, as you would for editing a

---

14 Here and throughout, we shall use the acronym MDS to denote non-metric (as opposed to metric) multi-dimensional scaling. More details about the method of MDS and its implementation in PRIMER can be found in chapter 5 of Clarke & Warwick (2001) and chapter 7 of Clarke & Gorley (2006).
1. PERMANOVA

worksheet (e.g., see chapter 1 of Clarke & Gorley 2006). Like other PRIMER files, the design file can also be saved separately in its own format (*.ppd). For the present example, change the name of the design file for the Ekofisk data from the default (Design1) to One-way by right-clicking on its name in the explorer tree and selecting Rename item. Save the entire workspace with the name ekofisk.pwk for continued analysis.

To run PERMANOVA on the Ekofisk data, click on the resemblance matrix and select PERMANOVA+ > PERMANOVA. In the PERMANOVA dialog box (Fig. 1.9), leave the defaults for all options, except (Num. permutations: 9999) & (Permutation method: Unrestricted permutation of raw data). Although any of the available methods of permutation offered by PERMANOVA are sound (see the section Methods of permutation below), permutation of raw data will provide an exact test for a simple one-way design. In addition, although the default for the number of permutations is 999, it is clearly desirable to perform as many permutations as reasonable time will allow. Power and precision increase with increases in the number of permutations (Hope 1968). Manly (2006, pp. 94-98) suggested that, to draw inferences at a significance level of 0.05, P-values should be calculated using at least 999 permutations, whereas 4999 permutations should be done to draw inferences at a level of 0.01. The ever-increasing speed of personal computers generally allows a large number to be chosen here for most designs with moderate sample sizes. Recall also that PERMANOVA obtains a random subset of all possible permutations, so will not necessarily reproduce exactly the same P-value for a given test if the analysis is run again. However, any such difference from one run to the next will be small (in the 3rd decimal place for 9999 permutations).

The results of PERMANOVA are shown in a new separate window with text-form information for this analysis (Fig. 1.10). The first part of the file provides information regarding the choices made, such as transformations, the resemblance measure and the method and number of permutations. There is also information regarding the experimental design and whether any terms were excluded (see the section Pooling or excluding terms below). The default ‘Type’ of sums of squares is Type III, which will be discussed in detail in the section Unbalanced designs below. These three types of
sums of squares are equivalent for one-way models or any balanced (equal replication) ANOVA designs, so need not concern us here.

The essential information of interest is provided in the ‘PERMANOVA table of results’, which contains the sources of variation in the model (‘Source’), the degrees of freedom (‘df’), sums of squares (‘SS’), mean squares (‘MS’), pseudo-F ratio (‘Pseudo-F’) and permutation P-value (‘P(perm)’). This table can be read and interpreted in a way that is directly analogous to a traditional ANOVA, but bear in mind that what is being partitioned here is multivariate variability based on the chosen resemblance measure, with P-values obtained using permutations.\(^\text{15}\)

The last column of the PERMANOVA table (‘Unique perms’) indicates how many unique values of the test statistic were obtained under permutation. Recall that PERMANOVA does not systematically do all permutations, but rather draws a random subset of them. In the example (Fig. 1.10), this value is very large (9862) and close to the number of random permutations that were chosen to be done by the user (9999). This means that only a few repeated values of pseudo-F\(^\pi\) were encountered under permutation, and the number of unique values is plenty enough to make reasonable inferences using the resulting permutation P-value, as shown. This information is important because, in some cases, the number of possible permutations is not large, and very few unique values of the test statistic are obtained. In such cases, a more meaningful (but approximate) P-value can be obtained by random sampling from the asymptotic permutation distribution instead (see the section Monte Carlo P-values below).

Underneath the PERMANOVA table of results are given further details regarding the analysis, including the expected mean squares (EMS) for each term in the model, the construction of the F ratio and the estimated sizes of components of variation (see the sections Components of variation,\(^\text{15}\)

\(^{15}\) For the one-way case, such as this, the PERMANOVA permutation P-values are exact; however, for multi-factor models, especially mixed models, analyses with covariates or unbalanced designs, the P-values provided are, necessarily, approximate, and rely on the exchangeability of residuals from the linear ANOVA model being fitted.
1. PERMANOVA

Expected mean squares. Constructing $F$ from EMS and Estimating components of variation below).

**Fig. 1.10.** Window of results from PERMANOVA on the Ekofisk macrofauna.

For the Ekofisk example, there is indeed a significant effect of the distance groupings (Fig. 1.10, pseudo-$F = 4.93$, $P = 0.0001$). A natural next question to ask is: wherein do these significant differences lie? That is, which groups differ significantly from which other groups?

Pair-wise comparisons among all pairs of levels of a given factor of interest are obtained by doing an additional separate run of the PERMANOVA routine. This is appropriate because which particular comparisons should be done, in most cases, is not known a priori, but instead will follow logically from the specific terms in the model found to be statistically significant in the main PERMANOVA analysis.

We could use the $F$ ratio for the pair-wise tests as well, but a more natural statistic to use here is a direct multivariate analogue to the univariate $t$ statistic. In traditional univariate analysis, an $F$ test to examine the effects of a factor having only two groups is equivalent to performing a two-sample (two-tailed) $t$ test. In fact, in this case the $t$ statistic for comparing two groups is simply equal to the square root of the $F$ ratio. Similarly, when PERMANOVA performs pair-wise tests, it takes each pair of levels to be compared, in turn and, treating it like a specific contrast, calculates pseudo-$t$ as the square root of pseudo-$F$. If a single response variable is being analysed and the resemblance measure chosen is Euclidean distance, then the $t$ statistics calculated by PERMANOVA for the pair-wise tests correspond exactly to Gosset’s original $t$ statistic (Student 1908). However, unlike traditional statistics packages, $P$-values for all pair-wise tests in PERMANOVA are obtained using permutations, not tables. This is necessary because, if the data are not normal, if more than one variable is being analysed and/or if the distance measure used is not Euclidean distance, then the distribution of this pseudo-$t$ statistic under a true null hypothesis is unknown, so (as with pseudo-$F$), we generate it using permutations.

To run pair-wise comparisons for the Ekofisk example, click on the window containing the resemblance matrix once again and select **PERMANOVA+ > PERMANOVA**. All of the choices in the dialog box that were made for the main test should still be there. (PRIMER v6 has a good memory!) Most of these will remain as before. The same design file is needed (One-way), but this time choose (Test > •Pair-wise test > For term: Dist > For pairs of levels of factor: Dist).
The file of results will contain the same preliminary information regarding the design and other choices as was seen for the main PERMANOVA test, but then the individual comparisons, using pseudo-\(t\) and with \(P\)-values by permutation, are given for each pair of levels of the factor. Generally speaking, as with the \(F\) ratio, the larger the \(t\) statistic, the greater the evidence against the null hypothesis of no difference in community structure between the two groups.

![PERMANOVA software interface](image)

**Fig. 1.11.** Pair-wise tests among groups for the Ekofisk macrofauna.

For the Ekofisk example, there is fairly strong evidence to suggest that all of the groups differ from one another \((P < 0.001\) for most comparisons, Fig. 1.11). However, the evidence against the null hypothesis for the comparison of group B with group C is weaker than the rest (Fig. 1.11, \(t = 1.24, P = 0.021\)).

A relevant point to note here is that no corrections for multiple comparisons have been made to any of these tests. It is a well-known phenomenon in statistics that the more tests you do, the greater your chance of rejecting one or more null hypotheses simply by chance. For example, if we were to perform 20 tests using an \textit{a priori} significance level of 0.05, we would expect to reject a true null hypothesis (i.e., to get a \(P\)-value smaller than 0.05) in one of those 20 tests by chance alone. However, the permutation \(P\)-values do provide an exact test of each individual null hypothesis of interest. In contrast, most \textit{ad hoc} experiment-wise corrections that could be used here (such as Bonferroni) are inexact and known to be overly conservative (e.g., Day & Quinn 1989). Thus, our philosophy (here and elsewhere) is to report the exact permutation \(P\)-values directly and to let the user decide whether or not to apply any additional corrections.

We recommend that one should consider the set of tests as a whole, in probabilistic terms. If there are \(a\) groups, then there will be \(a(a - 1)/2\) tests. For the Ekofisk example, this is \(4(4 - 1)/2 = 6\) tests. Clearly, in the present case, if all null hypotheses were true, it would be highly unlikely to get six out of six \(P\)-values less than 0.05 (as we have done) simply by chance! If, however, the number of tests were very large, with few small \(P\)-values encountered (e.g., if one obtained only one or two \(P\)-
1. PERMANOVA

values less than 0.05 out of 20 or so tests), then one might choose to apply a formal correction or, at the very least, exercise caution in interpreting the results.

The routine for pair-wise tests also provides a triangular matrix containing the average resemblances between samples that are either in the same group (along the diagonal) or in different groups (sub-diagonal elements, Fig. 1.11). This identifies the relative sizes of average similarities (or dissimilarities) between each pair of groups in the multivariate space, and also helps (along with the formal assessment provided in the PERMDISP routine, see chapter 2) to identify potential differences among the groups in terms of their within-group variability.

In some situations, there are not enough possible permutations to get a reasonable test. Consider the case of two groups, with two observations per group. There are a total of 4 observations, so the total number of possible re-orderings (permutations) of the 4 samples is 4! = 24. However, with a groups and n replicates per group, the number of distinct possible outcomes for the $F$ statistic in the one-way test is $(an)!/[a!(n!)^a]$ (e.g., Clarke 1993), which in this case is: $(4)!/[2!(2!)^2] = 3$ unique outcomes. This means that even if the observed value of pseudo-$F$ is quite large, the smallest possible $P$-value that can be obtained is $P = 0.333$. This is clearly insufficient to make statistical inferences at a significance level of 0.05.

An alternative is to use the result given in Anderson & Robinson (2003) regarding the asymptotic permutation of the numerator (or denominator) of the test statistic under permutation (see equations (1) and (4) on p. 305 of Anderson & Robinson 2003). It is demonstrated (under certain mild assumptions) that each of the sums of squares has, under permutation, an asymptotic distribution that is a linear form in chi-square variables, where the coefficients are actually the eigenvalues from a PCO of the resemblance matrix (see chapter 3). Thus, chi-square variables can be drawn randomly and independently, using Monte Carlo sampling, and these can be combined with the eigenvalues to construct the asymptotic permutation distribution for each of the numerator and denominator and, thus, for the entire pseudo-$F$ statistic, in the event that too few actual unique permutations exist.

A case in point where such an issue arises is given by a study of Victorian avifauna by Mac Nally & Timewell (2005). The data consist of counts of $p = 27$ nectarivorous bird species at each of eight sites of contrasting flowering intensity within the Rushworth State Forest in Victoria, Australia. One pair of sites had heavy flowering (‘good’ sites), another pair had intermediate flowering (‘medium’ sites), and a third pair had relatively little flowering (‘poor’ sites). Two sites near the good sites (‘adjacent’ sites), were also selected to explore possible “spill-over” effects. Sites were sampled using a strip transect method and data from four surveys were summed for each site (Mac Nally & Timewell 2005). The data are located in the file vic.pri in the ‘VictAvi’ folder of the ‘Examples add-on’ directory. An MDS ordination of the 8 sites on the basis of the binomial deviance dissimilarity measure (Anderson & Millar 2004) indicates potential differences among the four groups of sites in terms of their avifaunal community structure and an overall gradient from good to poor sites (Fig. 1.12).

We have $a = 4$ and $n = 2$, so there will be a total of $(8)!/[4!(2!)^4] = 105$ unique values of the pseudo-$F$ ratio under permutation for the main PERMANOVA test. This will provide some basis for making inferences from the $P$-value. For the pair-wise tests, however, there will be only 3 unique values under permutation, so obtaining Monte Carlo $P$-values would clearly be useful for these. To obtain Monte Carlo results, in addition to the permutation $P$-values, simply choose (√ Do Monte Carlo tests) in the PERMANOVA dialog.

For the Victorian avifauna, the PERMANOVA main test suggests that there are significant differences in bird communities among the four types of sites (Fig. 1.13). The permutation and

16 See Wheldon et al. (2007), in which an exact correction for family-wise error across a large set of non-independent simultaneous multivariate permutation tests was achieved using the permutation distributions.

17 The essential assumptions here are (a) that the observations of multivariate observations under permutation are independent and identically distributed, (b) that the distances are not just governed by a couple of very large ones (so that the central limit theorem can apply) and (c) that the distances are not too discontinuous – so a small change in the data would not produce a large change in the distance (sensible distance functions satisfy this). See Anderson & Robinson (2003) for details.
Monte Carlo $P$-values are quite similar in value (‘$P(\text{perm})$’ = 0.02 and ‘$P(\text{MC})$’ = 0.03) and note how the routine also faithfully shows that 105 unique values of pseudo-$F$ were obtained under permutation (‘Unique perms’).

**Victorian avifauna**

Resemblance: Binomial deviance

![MDS plot of Victorian avifauna at sites with different flowering intensities.](image)

**Fig. 1.12.** MDS plot of Victorian avifauna at sites with different flowering intensities.

**PERMANOVA**

Permational MANOVA

Resemblance: Binomial deviance

Group
- Poor
- Medium
- Good
- Adjacent

Stress: 0.05

![PERMANOVA results](image)

**Fig. 1.13.** PERMANOVA results for the main test and pair-wise tests for the Victorian avifauna, including Monte Carlo $P$-values.

The $P$-values for the pair-wise tests are virtually meaningless in this example, as there are only 3 unique possible values under permutation in each case. However, the Monte Carlo $P$-values are clearly much more valuable here and can be interpreted. Sites with medium flowering intensity appear to differ significantly from those with good flowering intensity (‘$P(\text{MC})$’ = 0.02), which is
also reflected in the relative size of the pseudo-\(t\) statistic (= 4.16). None of the other \(P\)-values were less than 0.05, although the comparison of poor with good sites came close (\(P\text{(MC)} = 0.06\)). One possible reason this latter comparison did not achieve a higher \(t\) statistic is because of the large within-group variation of the poor sites (average within-group distance = 66.0; Fig. 1.13).

If the number of unique permutations is large (say, 100 or more), then the permutation \(P\)-value should be preferred over the Monte Carlo \(P\)-value, as a general rule, because it will provide a more accurate test. The Monte Carlo \(P\)-values are based on asymptotic theory (albeit a very robust theory that generates distributions for test statistics which are specific to each dataset), so they rely on large-sample approximations. When there are a large number of possible permutations, then the Monte Carlo and permutation \(P\)-values should be very similar, essentially converging on the same answer. When, on the other hand, there are very few possible permutations, then the two \(P\)-values may be quite different (as seen in the Victorian avifauna example), in which case the Monte Carlo \(P\)-value should probably be used in preference. Although the Monte Carlo \(P\)-value is, at least, interpretable in cases such as these, it is only an approximation which relies on the central limit theorem, so clearly we shouldn’t get too carried away with our statistical inferences and interpretation of \(P\)-values from studies having such small sample sizes.\(^\text{18}\)

In multi-factor designs (discussed below), it is not always easy to calculate how many unique permutations there are for a given term in the model. It depends not only on the permutation method chosen, but also potentially depends on other factors in the model, how many levels they have and whether they may, under permutation, coincide with the term being tested in serendipitous ways. Therefore, for each test, PERMANOVA simply keeps track of how many unique values of the test statistic it encounters out of the total number of random permutations done. Armed with this information (‘Unique perms’), it is then possible to judge how useful the permutation \(P\)-value given actually is, under the circumstances.

The analysis of similarities (ANOSIM), described by Clarke (1993) is also available within PRIMER and can be used to analyse multivariate resemblances according to one-way and some limited two-way experimental designs.\(^\text{19}\) Not surprisingly, ANOSIM and PERMANOVA will tend to give very similar results for the one-way design on a given resemblance matrix. There are two essential differences, however, between ANOSIM and PERMANOVA. First, ANOSIM ranks the values in the resemblance matrix before proceeding with the analysis. The rationale behind the ranking procedure in ANOSIM is that the information of interest is the relationships among the dissimilarities (i.e., whether a given dissimilarity is larger or smaller than another) and not the values of the dissimilarities themselves. This is consistent with the philosophy of non-metric MDS ordination, which seeks to preserve only the rank order of the dissimilarities among samples. In contrast, PERMANOVA takes the point of view that the information of interest is in the dissimilarity values themselves, which describe a cloud of samples in multivariate space. This means that for PERMANOVA one must take special care to choose a measure of resemblance that is meaningful for the data and the goals of the analysis. For example, squared Euclidean distance may give different PERMANOVA results than Euclidean distance itself, whilst such a monotonic transform of the resemblances does not change the ranks and therefore cannot change ANOSIM.

The second essential difference is in the construction of the test statistic. The ANOSIM \(R\) statistic (Clarke 1993) is scaled to take a value between \(-1\) and \(+1\). This is a very useful feature, as it makes it possible to interpret the \(R\) statistic directly as an absolute measure of the strength of the difference between groups. \(R\) values are also directly comparable among different studies. In contrast, the value of pseudo-\(F\) (or pseudo-\(t\)) is, first of all, necessarily reliant on the degrees of freedom of the analysis, so cannot necessarily be compared in value across studies. A value of pseudo-\(F = 2.0\) (like its univariate analogue) will generally provide much stronger evidence against the null hypothesis if the residual degrees of freedom are 98 than if they are 5. Although values of pseudo-\(F\) may be comparable across different tests where the degrees of freedom are equal (for a

---

\(^\text{18}\) Unfortunately, it is precisely in these situations where sample sizes are small and we would like to use the Monte Carlo \(P\)-value that the asymptotic approximations assumed by this approach are actually the most precarious! A clear topic for further study is to discover under what conditions, more specifically, the Monte Carlo \(P\)-values may become unreliable.

\(^\text{19}\) See chapter 12 in Clarke & Gorley (2006) and chapter 6 in Clarke & Warwick (2001).
given dissimilarity measure and original number of variables, that is), it is also worth bearing in mind that the variability among groups (as measured by the numerator of the statistic) is always scaled against the variability within groups (as measured by the denominator). Thus, the within-group variability has an important role to play in the value of pseudo-F (or pseudo-t). An example is provided by the Victorian avifauna comparisons (Fig. 1.13), where, despite the pattern shown on the MDS plot (that samples from poor sites are farther away from the good sites than are the medium sites), pseudo-t is actually larger for the difference between good and medium sites than it is between good and poor sites, simply because the within-group variability between the poor sites is so high. ANOSIM, in contrast, yields an R statistic value of 1.0 (its maximum possible value) in both cases. In summary, while ANOSIM’s R can be interpreted directly as a measure of the size of the between-group differences, PERMANOVA’s pseudo-F (or pseudo-t) cannot necessarily be interpreted in this way. The sizes of effects in PERMANOVA are measured and compared in other ways: either by the average similarities (or dissimilarities) among pairs of groups (provided by the pair-wise routine) or by examining the estimated sizes of components of variation (see the section Estimating components of variation below). In addition, in PERMANOVA it is the P-values (either ‘P(perm)’ or, when necessary, ‘P(MC)’) which should be used as a measure of strength of evidence with respect to any particular null hypothesis. The Monte Carlo option available here also means that the power of the test need not especially rely on the number of possible permutations, as is the case for ANOSIM. Power in PERMANOVA will rely, however, on the number of replicates (more particularly, on the denominator degrees of freedom) available for the test.

Unlike ANOSIM, PERMANOVA achieves a partitioning of multivariate variability. As discussed in the introduction (section B of the Overview), this means PERMANOVA can be used to analyse much more complex experimental designs than ANOSIM. Although one could conceivably rank the dissimilarities before proceeding with a PERMANOVA analysis, this is not generally advisable when the goal is to achieve a partitioning of multivariate variability. The reason is that ranking dissimilarities loses information and therefore may result in less power. Another reason is that ranking the dissimilarities will tend to make the multivariate system highly non-metric, which can result in negative sums of squares and thus negative values of pseudo-F! The concept of negative variance which arises in non-metric or semi-metric geometric systems is discussed in more detail by Legendre & Legendre (1998) and McArdle & Anderson (2001) and in chapter 3 below on principal coordinates analysis (PCO). Suffice it for now to state simply that such results are confusing and difficult to interpret. They usually result from a poor choice of resemblance measure, or from ranking resemblances unnecessarily, so should be avoided if possible. The partitioning of variability described by the resemblance matrix on the basis of most reasonable dissimilarity measures (that have not been ranked) will generally produce a result where all of the SS (and pseudo-F ratios) are positive.

The primary advantage of PERMANOVA is its ability to analyse complex experimental designs. The partitioning inherent in the routine allows interaction terms in crossed designs to be estimated and tested explicitly. As an example, consider a manipulative experiment done to test the effects of shading and proximity to the seafloor on the development of subtidal epibiotic assemblages in Sydney Harbour, Australia (Glasby 1999). The data are located in the file sub.pri in the ‘SubEpi’ folder of the ‘Examples add-on’ directory. It was observed that assemblages colonising pilings at marinas were different from those colonising nearby natural sandstone rocky reefs. Glasby (1999) proposed that this difference might be due to the fact that assemblages on pilings are far from the seafloor, or it might be caused by them being shaded by the pier structure. The experiment he designed to test these ideas (Fig. 1.14a) therefore consisted of two factors:

Factor A: Position (fixed with a = 2 levels, either near (N) or far (F) from the sea floor).

Factor B: Shade (fixed with b = 3 levels, shaded by an opaque Perspex roof (S), open (O) or a procedural control (C), consisting of a clear Perspex roof).

---

20 This is analogous to the way that non-parametric univariate statistics are less powerful than their more traditional parametric counterparts when the assumptions of the latter are fulfilled. Interestingly enough, distance-based permutation tests (using Euclidean distance) can achieve even greater power than the traditional MANOVA test statistics in some situations, even when the assumptions of the traditional tests are true (Smith 1998, Mielke & Berry 2001, see also chapter 6 on CAP below).
The design was balanced, with \( n = 4 \) replicate sandstone plates \((15 \times 15 \text{ cm}^2)\) deployed in each of the \( a \times b = 6 \) combinations of the two factors for a total of \( N = a \times b \times n = 24 \) samples in the experiment. The six combinations of the treatment levels are also called cells (Fig. 1.14b). A balanced design is defined as a design that has a complete cell structure (i.e., where all combinations of the factors of interest are represented) and an equal number of replicate samples within every one of its cells. The percentage cover\(^{21}\) of each of 45 different taxa colonising the plates after a period of 33 weeks were measured. (There are \( p = 46 \) variables in the sub.pri worksheet; one of the variables, ‘bare’, is a measure of the unoccupied space on each plate).

This design is crossed because, for every level of factor A, we find all levels of factor B, and vice versa. A crossed design is also identifiable as one for which we could just as easily have swapped the order of the factors, either in the description or as drawn in a schematic diagram (Fig. 1.14c). A common mistake is to assume that one factor is nested within another (see Nested designs below), just because of the order in which the factors happen to be listed, or the perception that one factor physically occurs within another (e.g., treatments within sites or blocks), when, in fact, the order of the factors is readily swapped (e.g., because all treatments occur at all sites and the treatments have the same name and meaning at all sites).

Perhaps the most important additional feature of a crossed design is that the two factors can interact. By an interaction, we mean that the size and/or the direction of the effects for one factor are not consistently additive, but instead they depend on which level of the other factor you happen to be talking about. An interaction between two factors contributes an additional term to the model, so an additional sum of squares appears as a ‘Source’ in the PERMANOVA table. For the present experimental design, partitioning of the total variation therefore yields the following terms:

\[
\begin{align*}
SS_P &= \text{sum of squares due to the Position factor;} \\
SS_S &= \text{sum of squares due to the Shade factor;} \\
SS_{P \times S} &= \text{sum of squares due to the interaction between the two factors;} \text{ and} \\
SS_{Res} &= \text{residual sum of squares.}
\end{align*}
\]

Also, because the design is balanced, these individual terms are independent of one another and add up to the total sum of squares: \( SS_T = SS_P + SS_S + SS_{P \times S} + SS_{Res} \).

For these data, a traditional MANOVA analysis is out of the question. First, the variables clearly do not fulfill the assumptions (multivariate normality, homogeneity of variance-covariance matrices

\(^{21}\) Taxa that were present but occupied less than 1% cover were given an arbitrary value of 0.5%
among groups, etc.). Second, the number of variables \((p = 46)\) exceeds the total number of samples \((N = 24)\). Third, Euclidean (or Mahalanobis) distance is not an appropriate choice here, given the large number of zeros. Thus, we shall analyse the variability in this multivariate system on the basis of Bray-Curtis dissimilarities, with a fourth-root transformation of the densities in order to downweight the influence of the more abundant taxa. Open the file `sub.pri` and look at the factors by choosing `Edit > Factors`. Do an overall fourth-root transformation, then obtain a resemblance matrix based on Bray-Curtis.

Fig. 1.15. PERMANOVA analysis of the two-way crossed design for subtidal epibiota.

Analysis by PERMANOVA proceeds according to the two steps previously outlined for the one-way case: (i) create the design file and (ii) run the PERMANOVA routine. For the first step, choose `PERMANOVA+ > Create PERMANOVA design` > (Title: Subtidal epibiota) > (Number of factors: 2). The design file will begin by showing 2 blank rows. Double click in the first cell of each row and select the name of each factor in turn: Position in row 1 and Shade in row 2. Choose `Fixed` for both of the factors in column 3 and leave columns 2 and 4 of the design file blank (Fig. 1.15). Choose `File > Rename Design` and rename the design file `Two-way crossed`, for reference. Save the workspace as `sub.pwk`.

Unlike the PERMANOVA routine available in DOS, it is not necessary for the order of the factors in the design file to match the order in which the factors happen to appear in the data file. PERMANOVA+ for PRIMER uses label matching to identify factors and their levels and so the order of the factors with respect to one another as they are designated within either the data worksheet or the resemblance matrix has no consequence for the analysis. However, the order of the terms in the `design file` does have a consequence: by default the factors are fit in the order in which they are listed in the design file. For balanced experimental designs (like this one), changing...
the order of terms in the model has no effect on the results anyway. The order will matter, however, for unbalanced designs or designs with covariables if Type I sums of squares are used (see the section on Unbalanced designs below).

Next, with the resemblance matrix highlighted, choose PERMANOVA+ > PERMANOVA > (Design worksheet: Two-way crossed) & (Test •Main test) & (Sums of Squares •Type III (partial)) & (Permutation method •Permutation of residuals under a reduced model) & (Num. permutations: 9999). This is a more complex experimental design than the one-way case (indeed, most experimental designs involve more than one factor), so here we shall use the default method of permutation of residuals under a reduced model (see Methods of permutation for more details).

The results (Fig. 1.15) indicate that there is no statistically significant interaction in the effects of Position and Shade on variability in these assemblages, although the P-value is not large (P = 0.09). Each of the main factors do appear, however, to have strong effects (P < 0.001 for each test). These results are also reflected in the patterns seen in the MDS plot of the two factors, where labels are used to denote different levels of the Position factor and symbols are used to denote different levels of the Shade factor (Fig. 1.16).

![MDS plot of subtidal epibiota with labels for the two-way crossed design.](image)

There is a clear separation in the MDS plot between assemblages near (N) versus those far (F) from the seafloor, suggesting a strong effect of Position. There is also a tendency for the samples in the shaded treatments to occur in the lower left of the diagram, compared to those in either the procedural control or in open treatments, which occur more in the centre or upper right. Although the size of the shade effect appeared to be a bit larger for assemblages near to the seafloor than for those far from the seafloor (Fig 1.16), the direction was very similar and the test revealed no significant interaction.

What do we mean by an “interaction” between two factors in multivariate space? Recall that for a univariate analysis, a significant interaction means that the effects of one factor (if any) are not the same across levels of the other factor. An interaction can be caused by the size and/or the direction of the effect being different within different levels of the other factor. Similarly, in multivariate space, we can consider that a single factor has an effect on the combined set of response variables when it causes a shift in the location of the data cloud. If the magnitude and/or the direction of that effect is different between levels of the other factor, then the two factors interact.

---

22 To see this, run the PERMANOVA analysis where the first factor listed in the design file (in the first row) is Shade and the second factor in the design file is Position.
shift for a given factor (e.g., factor A) is different when considered separately within each level of the other factor (say, factor B), then we may consider that those two factors interact. 

To clarify ideas, consider a hypothetical two-way crossed experimental design with factor A having 2 levels (A1 and A2), factor B having 3 levels (B1, B2 and B3) and which can be drawn in two dimensions in Euclidean space (Fig. 1.17). We can conceive of a situation where the effect (shift in location) due to factor A (e.g., from triangles to circles) is of a similar size and in a similar direction, regardless of which level of factor B we happen to be in (Fig. 1.17a). In this case, there is no interaction. Similarly, we can conceive of a situation where there are apparently no effects of factor A at all, regardless of factor B, which would also result in there being no significant interaction (Fig. 1.17b). In contrast, we can conceive of situations where the sizes (or directions) of factor A’s effects are different within different levels of factor B (Fig. 1.17c, d). Note that, for univariate analysis, we consider changes in direction only along a single dimension for a given response variable, with an effect being either positive or negative. In contrast, there are many possible ways that changes in direction can happen in multivariate space.

The logic of the interpretation of a test for interaction in PERMANOVA follows, by direct analogy, the logic employed in univariate ANOVA. In particular, it is important to examine the test of the interaction term first in order to know how to proceed. As for univariate analysis, the presence of a significant interaction generally indicates that the test(s) of main effects (i.e., the test of each factor alone, ignoring the other factor) may not be meaningful. Appropriate logic dictates that the next step after obtaining a significant interaction is to do pair-wise comparisons for the factor of interest separately within each level of the other factor (i.e., to do pair-wise comparisons among levels of factor A within each level of factor B) and vice versa (if necessary). On the other hand, if the interaction term is not statistically significant, then this indicates effects (if any) of each factor do

---

23 It was noted earlier that PERMANOVA is not sensitive to differences in correlation structure among groups (Fig. 1.6). Similarly, the PERMANOVA test of interaction is focused more on the additivity of the sizes of the effects, rather than on their direction, per se. Thus, differences in effects which are directional but which do not affect the additivity of effect sizes may go undetected. Studies examining the power of PERMANOVA to detect different kinds of interactions are needed to clarify this issue further.
not depend on the other factor, and one may examine the individual pseudo-\(F\) tests of each factor alone (and do subsequent pair-wise comparisons, if appropriate), ignoring the other factor.

For the subtidal epibiotic assemblages, the lack of a significant interaction meant that we could logically proceed to investigate the main effects, which were each highly statistically significant. The next logical step is to consider pair-wise comparisons separately for each factor. For the factor of Position, there are only 2 levels, so the pseudo-\(F\) ratio demonstrates already that there is a significant difference between these two groups (near and far). (We might decide nevertheless to do pair-wise comparisons for this term, if we wish to know something more about the sizes of the average within-group or between-group dissimilarities, for example.) On the other hand, the factor of Shade has three levels and pair-wise comparisons are desirable to discern wherein the significant differences may lie (i.e., between which pairs of groups). Run the PERMANOVA routine again with all of the same choices in the dialog as for the main test, except this time choose (Test ▶ Pair-wise test ▶ For term: Shade ▶ For pairs of levels of factor: Shade) (Fig. 1.18). The results show that there is no significant difference between assemblages in the open and procedural control treatments (C, O), although both of these differed significantly from assemblages in the shade treatment (S) (Fig. 1.18).

![PERMANOVA Dialog](image1.png)

**Fig. 1.18.** PERMANOVA dialog and results for the comparison among the three shade treatments for subtidal epibiota.

In the event that the interaction term had been statistically significant, then the logical way to proceed would have been to do pair-wise tests among levels of the Shade factor separately for each of the near and far situations (i.e., within each level of factor Position). To do pair-wise comparisons like this, following on from an interaction term, one would choose (Test ▶ Pair-wise test ▶ For term: PositionxShade ▶ For pairs of levels of factor: Shade) in the PERMANOVA dialog. Similarly, one could also in that case consider doing comparisons among levels of the factor Position (near versus far) separately within each of the shade levels (S, O and C). For the latter, one would choose (Test ▶ Pair-wise test ▶ For term: PositionxShade ▶ For pairs of levels of factor: Position). In either case, one first chooses the interaction term of interest, followed by the particular factor involved in that interaction which is the focus for that set of pair-wise comparisons. In complex designs, there may be multiple sets of pair-wise comparisons that the user may wish to do in order to follow up the full model partitioning and results shown by the main test.

As a final note, keep in mind that the patterns in an MDS (or PCO) ordination plot may or may not show the reasons for an interaction, if it is present, because the full dimensionality of the multivariate cloud has been reduced (usually to 2 or 3 dimensions). Importantly, PERMANOVA
works on the underlying dissimilarities themselves for the test, so its results should be trusted over and above any patterns (or lack of patterns) apparent in the ordination. Usually an ordination will help, however, to interpret the PERMANOVA results, provided the stress is not too high.24

### Additivity

Central to an understanding of what an interaction means for linear models25 is the idea of additivity. Consider the example of a two-way crossed design for a univariate response variable, where the cell means and marginal means are as shown in Fig. 1.19a. Note that the marginal means are the means of the levels of each factor ignoring the other factor. In an additive model, the difference between two levels of a factor (say between B1 and B2) between individual cells (i.e., within each level of A, that is to say, within each column) are equal to the differences in the marginal means (i.e., the difference between the mean of B1 and B2 if factor A were to be ignored). This can be contrasted with the situation where the differences in cell means are quite different from the differences in marginal means (e.g., Fig. 1.19b), in which case, there is an interaction between the factors. So, this is another way to articulate what is meant by a significant interaction: effects of factors within levels of other factors are non-additive and thus do not match the corresponding shifts in marginal means. The interaction term, in fact, measures the deviation of the cell means we actually got from what we would expect them to be if they were to follow the marginal means, as would be the case if the effects of the two factors were purely additive.

![Fig. 1.19. Marginal and cell means for a univariate crossed design showing examples of (a) additive effects, (b) multiplicative effects and (c) additivity after log transformation of (b).](image)

Clearly, the additivity (or not) of the effects of factors is also going to depend on whether or not the data have been transformed (or standardised or ranked) prior to analysis. This is as true for multivariate data as it is for univariate data. For example, if a log (base 10) transformation is applied to the means shown in Fig. 1.19b, then we would have an additive model with no significant interaction (Fig. 1.19c). Such a situation typifies phenomena where the true effects are multiplicative, rather than being additive. In univariate analysis, transformations can often be used to remove significant interaction terms, yielding additivity (Tukey 1949, Box & Cox 1964, Kruskal 1965, Winsberg & Ramsay 1980).

For multivariate analysis of ecological data, however, transformations are usually applied neither to fulfill assumptions, nor in order to remove significant interactions, but rather as a method of changing the relative emphasis of the analysis on rare versus more abundant species (e.g., Clarke & Green 1988, Clarke & Warwick 2001). In PRIMER, a blanket transformation can be applied to all variables by choosing Analyse > Pre-treatment > Transform (overall) and then choosing from a range, in increasing severity, from no transformation, square root, fourth root or log(x+1) down to a reduction of the values to binary presence (1) or absence (0). An approach using an intermediate-level transformation (square root or fourth root) has been recommended as a way to reduce the contribution of highly abundant species in relation to less abundant ones in the calculation of the

---

24 **Stress** is a measure of how well inter-point distances in an MDS ordination represent the rank-ordered inter-sample dissimilarities in the original resemblance matrix. For a more complete description of the notion of stress, see previous references to MDS along with chapter 5 of Clarke & Warwick (2001) and chapter 7 of Clarke & Gorley (2006).

25 The ANOVA models analysed by PERMANOVA are linear only in the space of the multivariate cloud defined by the dissimilarity measure of choice; they are not linear in the space of the original variables (unless the resemblance measure chosen was Euclidean distance).
Bray-Curtis measure; rare species will contribute more, the more severe the transformation (Clarke & Green 1988, Clarke & Warwick 2001).

In addition to the transformation, additivity of effects in multivariate analysis is also going to depend on whether or not the dissimilarities are ranked before analysis (yet another reason why patterns in a non-metric MDS, which preserves ranks only, may not necessarily clearly reflect what is given in the PERMANOVA output). The choice of dissimilarity measure itself is also very important here. By performing the partitioning, PERMANOVA is effectively applying a linear model to a multivariate data cloud, as defined by these choices. So the presence of a significant interaction (or not) by PERMANOVA will naturally depend on them. Nevertheless, the choice of an appropriate dissimilarity measure (and also the choice of transformation, if any) should genuinely be driven by the biology and ecology (or other nature) of the system being studied and what is appropriate regarding your hypotheses, and not by reference to these statistical issues (unlike typical traditional univariate ANOVA).

As for the one-way case, the distribution of each of the pseudo-$F$ ratios in a multi-way design is generally unknown. Thus, a permutation test (or some other approach using re-sampling methods) is desirable. When there is more than one factor, situations commonly arise which prevent the possibility of obtaining an exact test of individual terms in the model using permutations. For example, there is no exact permutation test for an interaction (but see Pesarin (2001), who describes a synchronised permutation method for testing interactions). In addition, restricted permutation methods for testing main effects in ANOVA models generally have low power (Anderson & ter Braak 2003). However, several good approximate permutation methods can be used instead to get accurate $P$-values (e.g., Anderson & Legendre 1999). PERMANOVA provides three general options regarding the method of permutation to be used: (i) unrestricted permutation of raw data, (ii) permutation of residuals under a reduced model, or (iii) permutation of residuals under the full model. These methods and their properties are described in detail elsewhere (e.g., Anderson & Legendre 1999, Anderson 2001b, Anderson & Robinson 2001, Anderson & ter Braak 2003, Manly 2006). Although these methods do not give exact $P$-values for complex designs in all cases, they are asymptotically exact and give very reliable results. In practice, these three approaches will give very similar results, so there is (thankfully) no need to agonise much about making a choice here. All three of the methods are implemented in PERMANOVA so as to ensure that the correct exchangeable units (identified by the denominator of the pseudo-$F$ ratio) are used for each individual test (see Anderson & ter Braak (2003) for details). Some of the known properties of the three methods are outlined below.

- **(i) Unrestricted permutation of raw data.** This is a good approximate test proposed for complex ANOVA designs by Manly (1997). It will generally have type I error close to $\alpha$, although with larger sample sizes it tends to be more conservative (less powerful) than the tests that permute residuals (Anderson & ter Braak 2003).\(^{27}\) However, this method does not need large sample sizes to work well (Gonzalez & Manly 1998). It is also, computationally, the fastest option. The method does suffer from a few problems, however, if there happen to be outliers in covariables (if present, see Kennedy & Cade 1996), so should not be used for such cases.

- **(ii) Permutation of residuals under a reduced model.** This approach was first described for linear models by Freedman & Lane (1983) and is the default option in PERMANOVA because it has excellent empirical and theoretical properties. Empirically, it yields the best power and the most accurate type I error for multi-factorial designs in the widest set of circumstances (Anderson & Legendre 1999, Anderson & ter Braak 2003). Also, this method is theoretically the closest to the conceptually exact test (Anderson & Robinson 2001). The idea is to isolate the term of interest in the model for each test by fitting the other terms (the reduced model),

\(^{26}\) An asymptotically exact test is a test for which the type I error (probability of rejecting the null hypothesis when it is true) asymptotically approaches (converges on) the a priori chosen significance level ($\alpha$) with increases in the sample size ($N$).

\(^{27}\) Note that “less powerful” does not necessarily mean that using (i) will give you a smaller $P$-value than (ii) or (iii) for any particular data set. It means that, in repeated simulations, the empirical power (estimated probability of rejecting the null hypothesis when it is false) was, on average, smaller for method (i) than for either of the other two methods in most situations.
obtaining residuals from that reduced model, and permuting those. In other words, the entities that are exchangeable under the null hypothesis for a particular term are the errors (estimated by the residuals) obtained after removing the terms in the model that are not of interest for that test. The definition of the reduced model (and therefore the residuals arising from them) thus depends on which term is being tested.

(iii) Permutation of residuals under the full model. This method was described by ter Braak (1992). The idea is to obtain residuals of the full model by subtracting from each replicate the mean corresponding to its particular cell (combination of factor levels). These residuals are estimating the errors associated with each replicate. These are then permuted and the statistic is re-calculated for all terms using these residuals (as if they were the data) under permutation. This method mostly gives results highly comparable to method (ii). It relies somewhat more than (ii), however, on large within-cell sample sizes for precision. It has the advantage, however, of being faster than method (ii) for the analysis of the entire design, as the same residuals are permuted for all terms in the model under test.

In general, we recommend using method (ii), which is the default. Method (i), however, does provide an exact test for the one-way case, so should be used for one-way ANOVA models. Otherwise, note that methods (ii) and (iii) both require estimation of parameters (means) in order to calculate residuals as deviations from fitted values. When sample sizes are small, these estimates are not very precise (i.e., they may not be very close to their “true” values), so the residuals being permuted, in turn, may not be good representatives of the “true” errors (Anderson & Robinson 2001). Thus, in the case of relatively small sample sizes (say, \( n < 4 \) replicates per cell), method (i) is also recommended (provided there are no outliers in covariables, as mentioned above). Method (iii) is probably only advisable if you wish to use method (ii), but the time required is getting overly burdensome.

Recall that we assume for the analysis of a one-way design by PERMANOVA that the multivariate observations are independent and identically distributed (i.i.d.) under a true null hypothesis. When more complex (multi-way) designs are analysed, a few more assumptions are added, due to the way partitioning is done and the method of permutation used (generally, the method used is permutation of residuals under a reduced model). More specifically, for multi-way designs, PERMANOVA fits an additive linear model to the multivariate samples in the space of the chosen resemblance measure. This assumes that effects of factors and their interactions can be modeled meaningfully in this additive fashion, as opposed to using, say, a non-linear, multiplicative or other approach (e.g., see Millar et al. 2005). It also assumes that the errors (which may be estimated using the residuals after fitting a given PERMANOVA model) are i.i.d. across the full design in the space of the chosen resemblance measure.

At present, it is unknown to what extent and in what circumstances departures from these assumptions might affect error rates or interpretations of results from PERMANOVA. Plots showing distributions of residuals vs fitted values (e.g., from either PERMANOVA or DISTLM models, see chapter 4) for each of a series of PCO dimensions (see chapter 3) in order of decreasing importance, for example, and also multivariate ordination plots of residuals could provide helpful diagnostic tools. Further study is warranted to develop these tools and to investigate the effects of violations of the assumptions, especially for mixed models, models with covariates or unbalanced designs (which are all discussed in more detail later in this chapter) and even for simple designs that include interactions. In addition, the good asymptotic properties of the methods of permutation of residuals used by PERMANOVA (and DISTLM) require reasonable sample sizes within the cells. Small numbers of observations within cells will result in inter-correlations among residuals. Although exchangeability of errors (implying independence and homogeneity) and additive effects in the space of the resemblance measure are fairly modest assumptions, we nevertheless look forward to future studies where the full implications of the use of these models in different situations and with different resemblance measures may become clearer. In the meantime, simulations done with univariate data having highly non-normal error

---

28 For unbalanced designs, it also depends on which Type of SS is chosen for the test. For Type I SS, the order in which the terms are fitted will also matter here.
structures indicate that the permutation methods implemented by PERMANOVA are quite stable for reasonable sample sizes (as indicated above) and when observations are i.i.d.

In some cases, what is of interest in a particular experimental design is not necessarily the comparisons among all pairs of levels of some factor found to be significant, but rather to compare one or more groups (or levels) together versus one or more other groups. A comparison such as this is called a contrast and is usually logically formulated at the design stage (a priori). Pair-wise comparisons, on the other hand, are generally done after the analysis, so are also sometimes called unplanned or a posteriori comparisons. For example, in the study of the subtidal epibiotic assemblages, we may especially wish to compare a priori the assemblages in the shaded treatment with those occurring in either the open or the procedural control treatments. That is, we would like to contrast group (S) with groups (C, O), effectively treating the latter two treatments together as a single group. Another possible contrast of interest would be the comparison of the procedural control with the open group (i.e., C versus O), as this contrast would identify possible artefacts of the structure used to create shade.

PERMANOVA allows the user to specify particular contrasts of interest. Each of these has 1 degree of freedom and is used to further partition the sums of squares (SS) attributable to that factor. In addition, any interaction terms involving that factor are also partitioned according to the contrast. To see how this works for the subtidal data, click on the design file named Two-way crossed in the sub.pwk workspace and choose Tools>Duplicate, then choose File> Rename Design and name this new design file With Contrasts for reference. Next, in row 2 for the Shade factor, double click on the cell in the final column, headed ‘Contrasts’, which will bring up a new ‘Contrasts’ window (Fig. 1.20). Click on ‘Add’ in order to add a new row to the list of contrasts shown in the window. Double click within the cell in the first column (‘Name’) and type the name S-vs-(C,O). Double click in the cell in the second column (‘Contrast’). This brings up a dialog in which all of the levels available in that factor are listed. Identify the contrast by clicking on S, then on to place it on the left and then click on each of C and O in turn, each followed by , to place the latter two levels on the right, followed by ‘OK’. Next, add a second row in the contrasts window which specifies a contrast of the C and O treatments, to be called C-vs-O (Fig. 1.20).

Fig. 1.20. Dialog to create contrasts for the subtidal epibiotia.
When you are finished, the design file will show the names of the contrasts you have specified in the ‘Contrasts’ column for the Shade factor (Fig. 1.20).

Re-run the PERMANOVA analysis by selecting the resemblance matrix and choosing PERMANOVA+ > PERMANOVA > (Design worksheet: With Contrasts) & (Test •Main test) & (Sums of Squares •Type III (partial)) & (Permutation method •Permutation of residuals under a reduced model) & (Num. permutations: 9999). The contrasts associated with a particular factor are indented in the output file, to emphasise that these are a further partitioning of the SS associated with a given factor (or interactions involving that factor) (Fig. 1.21). These are therefore “extra” tests, in addition to the test of the factor as a whole.

In general, one may construct up to \(df_a\) orthogonal (independent) contrasts, where \(df_a\) is the number of degrees of freedom for the factor. For these situations (i.e., orthogonal a.k.a. independent contrasts), then the SS for the contrasts chosen will add up to the SS for the factor (that is, if the number of orthogonal contrasts is equal to \(df_a\)) and it is generally considered unnecessary to worry about doing any corrections for multiple tests. This is true of the example we have done. However, if the contrasts are not orthogonal, and especially if one has chosen to perform a great many contrasts that exceed \(df_a\) in number, then the potential inflation of overall type I error due to performing multiple tests may be an issue (see the section on Pair-wise comparisons above).

For our example (Fig. 1.21), the results reveal the strong effect of the shading treatment versus the other two treatments (note how S-vs-(C,O) has a pseudo-\(F\) ratio even larger than the pseudo-\(F\) for the Shade factor overall) and also show the lack of any apparent procedural artefact (C-vs-O, \(P = 0.28\)). In addition, the interaction term of Position \(\times\) S-vs-(C,O) is approaching statistical significance (\(P = 0.05\)), providing increased evidence that the effect of shading (which is tested more directly by this specific contrast than by the test comparing all three treatments) is different near the seafloor compared to far away from the seafloor (e.g., Glasby 1999).

A special situation where contrasts might be useful is in the comparison of, say, an impact location versus several control locations in an environmental impact study design (Underwood 1992). If there is no replication of the impact state, but there are several control locations, then this actually generates what is known as an asymmetrical design. For such designs, the correct SS can be obtained using contrasts, however, it is necessary to treat the model explicitly as an asymmetrical design in order to get correct pseudo-\(F\) ratios and \(P\)-values for all of the tests (see the specific section on Asymmetrical designs below).

All of the factors considered so far have been fixed, but factors can be either fixed or random. In univariate ANOVA the choice of whether a particular factor is fixed or random has important consequences for the assumptions underlying the model, the expected values of mean squares (thus, the construction of a given \(F\) ratio), particularly in more complex designs, the hypothesis tested by the \(F\) ratio and, perhaps most importantly, the extent and nature of the inferences. This is also true for PERMANOVA, which follows the analogous univariate ANOVA models in terms of the construction of pseudo-\(F\) ratios from expectations of mean squares (EMS).
For a fixed factor, there is a finite set of levels which have been explicitly chosen to represent particular states. These states (the levels) are generally explicit because they have been manipulated (e.g., shade, open and control), because they already exist in nature (e.g., male and female), or because they bear some meaningful relationship to other chosen levels (e.g., high, medium and low). All of the levels of a fixed factor (or at least all of the ones we are effectively interested in) occur in the experiment. So, for a fixed factor, individual levels have a meaning in themselves, and generally, if we were going to repeat the experiment, we would choose the same levels to investigate again (unless we were going to change the hypothesis). The effects of a fixed factor are deemed to be constant values for each nominated state (or level)\(^{29}\). Furthermore, the component of variation attributable to a fixed factor in a given model is considered in terms of the sum of squared fixed effects (divided by the appropriate degrees of freedom). Finally, when we perform the test of the fixed factor, the resulting \(P\)-value and any statistical inferences to be drawn from it apply only to those levels and to no others.

For a random factor, however, the particular levels included in the experiment are a random subset from a population of possible levels we could have included (e.g., sites, locations, blocks). We do not consider the individual levels (site 1, site 2, etc.) as representing any particular chosen state. The levels do not have any particular meaning in themselves; we are not interested in comparing, say, site 1 vs site 2, but rather multiple levels of the factor are included to provide us with a measure of the kind of variability that we might expect across the population of possible levels (e.g., among sites). Random factors, unlike fixed factors, actually contribute another source of random variance into the model (in addition to the error variance) and so the effects, rather than being fixed, are instead used to estimate the size of the variance component for that factor. Repeating the experiment would also probably not result in the same levels being chosen again. Importantly, when the \(F\) ratio is constructed and the \(P\)-value is calculated for a random factor, the statistical inference applies to the variance component for the whole population of possible levels (or effects) that could have been chosen, and not just to the levels included in the experiment.

The difference in the inference space for the fixed versus the random factor is important. For example, if one obtained a \(P\)-value less than \(\alpha = 0.05\) (the usual convention) for a fixed factor, one might state: “There were significant differences among (say) these three treatments in the structure of their assemblages”. In contrast, for a random factor, one might state: “There was significant variability among sites in the structure of the assemblages.” The fixed factor is more specific and the random factor is more general. Note also that the statement regarding the fixed factor often logically calls for more information, such as pair-wise comparisons between individual levels – which treatments were significantly different from one another and how? Whereas, the second statement does not require anything like this, because the individual levels (sites) are generally not of any further interest in and of themselves (we don’t care whether site 1 differs from site 3, etc.); it is enough to know that significant variability among sites is present. Thus, it is generally not logical to do pair-wise comparisons among levels of a random factor (although PERMANOVA will not prevent you from doing such tests if you insist on doing them)! A logical question to ask, instead, following the discovery of a significant random factor would be: how much of the overall variability is explained by that factor (e.g., sites)? For a random factor, we are therefore more interested to estimate the size of its component of variation and to compare this with other sources of variation in the model, including the residual (see Estimating components of variation below).

To clarify these ideas, consider an example of a two-way crossed experimental design used to study the effects of disturbance by soldier crabs on meiofauna at a sandflat in Eaglehawk Neck, Tasmania (Warwick et al. 1990a). The \(N = 16\) samples consist of two replicates within each combination of four blocks (areas across the sandflat) and two natural ‘treatments’ (either disturbed or undisturbed by soldier crab burrowing activity). There were \(p = 56\) taxa recorded in the study, consisting of 39 nematode taxa and 17 copepod taxa. We therefore have

**Factor A:** Treatment (fixed with \(a = 2\) levels, disturbed (D) or undisturbed (U) by crabs).

\(^{29}\) In univariate analysis, the effect for a given group is defined as the deviation of the group mean from the overall mean. Similarly, when using PERMANOVA for multivariate analysis, the effect for a given level is the deviation (distance) of the group’s centroid from the overall centroid in the multivariate space, as defined by the dissimilarity measure chosen.
Factor B: Block (random with \( b = 4 \) levels, labeled simply B1-B4).

The first factor is fixed, because these two treatment levels do have a particular meaning, representing particular states in nature (disturbed or undisturbed) that are of interest to us and that we explicitly wish to compare. In contrast, the second factor in the experiment, Blocks, is random, because we do not have any particular hypotheses concerning the states of, say B1 or B2, but rather, these are included in the experimental design in order to estimate variability across the sandflat at a relevant spatial scale (i.e., among blocks) and to avoid pseudo-replication (sensu Hurlbert 1984) in the analysis of treatment effects. A design which has both random and fixed factors is called a mixed model. This particular design (where a fixed factor is crossed with a random one), also allows us to test for generality or consistency in treatment effects across the sandflats.

The data are located in the file tas.pri in the ‘TasMei’ folder of the ‘Examples add-on’ directory. View the factors by choosing Edit>Factors. In order to obtain different symbols for each of the \( a \times b = 2 \times 4 = 8 \) cells in the design, create a new factor whose 8 levels are all combinations of Treatment \( \times \) Block by choosing ‘Combine’ in the Factors dialog box and then ➤ followed by ‘OK’. Create a resemblance matrix on the basis of Bray-Curtis similarities after square-root transformation. An MDS of these data (as shown in chapter 6 of Clarke & Warwick 2001) shows a clear effect of disturbance on these assemblages, with some variability among the blocks as well (Fig. 1.22).

Fig. 1.22. MDS of Tasmanian meiofauna showing variation among the blocks on the x-axis and separation of disturbed from undisturbed communities on the y-axis.

Set up the design file according to the correct experimental design (Fig. 1.23) and rename it as Mixed model, for reference. The next step is to run PERMANOVA on the resemblance matrix according to the mixed model design. Due to the small number of observations per cell, choose (Permutation method •Unrestricted permutation of raw data) & (Num. permutations: 9999). Save the workspace with the PERMANOVA results as tas.pwk.

The PERMANOVA results show that there is a statistically significant (though borderline) interaction term, indicating that the treatment effects vary from one block to the next \((P = 0.044, \text{ Fig. 1.23})\). The MDS plot shows that the direction of the treatment effects appears nevertheless to be fairly consistent across the blocks (at least insofar as this can be discerned using an ordination to represent the higher-dimensional cloud of points), so in this case the significant interaction may be caused by there being slight differences in the sizes of the treatment effects for different blocks.

For any given ANOVA design, PERMANOVA identifies a component of variation for each term in the model, denoted as ‘\( S(*) \)’ for the fixed terms and ‘\( V(*) \)’ for the random terms. This notation is used because, in the analogous univariate case, components of variation for a fixed factor are sums of squared fixed effects (divided by appropriate degrees of freedom), while components of...
variation due to random factors are actual *measures of variability* or variance components. This is appropriate, because the hypotheses are also different in these two cases. For fixed effects, the hypothesis only concerns *the effects of those levels* that were included in the experiment, whereas for the random factor, the hypothesis is about the *variability among a population of levels*, of which the levels in the experiment are a random representative sample. Note that any interaction involving a random factor will also be random. The residual component is also denoted by ‘V(Res)’ because it too is a measure of variability. Thus, for example, in the two-way mixed model design for the Tasmanian meiofauna, the total variation is partitioned according to four sources, as follows:

- **S(Tr):** sum of squared treatment effects (divided by degrees of freedom);
- **V(BI):** variation due to blocks;
- **V(TrxBI):** variation in treatment effects among blocks; and
- **V(Res):** residual variation.

Note that if one is using PERMANOVA to analyse a single variable with Euclidean distance, then the components of variation are indeed true variance components (for random factors) and true sums of squared fixed effects divided by degrees of freedom (for fixed factors). Here and in PERMANOVA, they are simply called *components of variation*, however, in order to cover the more general cases, including the analysis of many variables on the basis of non-Euclidean resemblance measures.

![Mixed model](image1)

**Fig. 1.23.** Design file and PERMANOVA results for the Tasmanian meiofauna, including details of the expected mean squares and construction of F ratios for each term in the mixed model.

An important consequence of the choice made for each factor as to whether it be fixed or random is identified by examining the *expected mean squares (EMS)* for each term in the resulting model. This is vitally important because the EMS’s are used to identify an appropriate denominator mean square that one must use for each particular term in the model in order to construct a correct pseudo-F ratio that will isolate that term of interest for the test. For univariate ANOVA, the
underlying theory for deriving expectations of sums of squares and mean squares is covered well elsewhere (e.g., Cornfield & Tukey 1956, Hartley 1967, Rao 1968, Winer et al. 1991, Searle et al. 1992). PERMANOVA actually uses the same “rules” for constructing these expectations, implementing these as a direct multivariate analogue to the univariate approach.

The default rules used by PERMANOVA assume that fixed effects sum to zero, following Cornfield & Tukey (1956) and Winer et al. (1992). Some constraint on fixed effects is necessary, due to the intrinsic over-parameterisation of the ANOVA model (e.g., Scheffé 1959), and the sum-to-zero constraint is a highly convenient one. The constraint chosen does not affect the sums of squares. It does, however, affect the EMS’s and thus the F ratios and P-values that are obtained for ANOVA mixed models. Some well-known computer packages for univariate statistics relax the sum-to-zero constraint for fixed effects in mixed interactions, including SPSS and the ‘proc GLM’ routine in SAS. To obtain EMS’s in accordance with these packages, remove the ✓ in front of the ‘Fixed effects sum to zero’ box in the PERMANOVA dialog. See Hartley & Searle (1969), Searle (1971), Hocking (1973), McLean et al. (1991) and Searle et al. (1992) for further discussion and debate regarding this issue.

Once the individual components of variation have been identified, then the expectations of the mean squares for each term in the model are determined precisely in terms of these components, and are provided in the output under the heading ‘Details of the expected mean squares (EMS) for the model’. Thus, we can see, for example (Fig. 1.23), that the expectation for the mean square calculated for the term ‘Block’ (abbreviated as ‘Bl’ in the output) is one times the residual variation plus four times the variation due to blocks (denoted as ‘1*V(Res) + 4*V(Bl)’ in the output). Each term in the model will have an expected mean square that consists of some linear combination of the components of variation in the model.

The determination of the EMS’s gives a direct indication of how the pseudo-F ratio should be constructed in order to isolate the term of interest to test a particular hypothesis (e.g., Table 1.1). Consider the test of the term ‘Block’ (‘Bl’) in the above mixed-model experimental design. The null hypothesis is that there is no significant variability among blocks. Another way of writing this, in the notation used by PERMANOVA, is H₀: V(Bl) = 0. To begin, we will construct an F ratio where the numerator is the mean square for the term of interest (e.g., to test blocks, then the numerator will be the mean square for blocks). Now, given the EMS for this numerator term of interest, we essentially need to find a denominator whose expectation would correspond to the numerator if the null hypothesis were true. In the case of the ‘Block’ term, if V(Bl) were equal to zero, then the EMS for blocks would just be 1*V(Res). The term whose EMS corresponds to 1*V(Res) is the residual. Therefore, the pseudo-F ratio for the test of the ‘Block’ term is the mean square for blocks divided by the residual mean square, viz: F_{Bl} = MS_{Bl} / MS_{Res}. By following a similar logic, we can see that the test of the interaction term (‘TrxBl’) is provided by F_{TrxBl} = MS_{TrxBl} / MS_{Res} (Table 1.1).

For the ‘Treatment’ term (‘Tr’), we see that its EMS is: 1*V(Res) + 2*V(TrxBl) + 8*S(Tr). The null hypothesis here is that there are no consistent treatment effects, or, equivalently, H₀: S(Tr) = 0. If the null hypothesis were true, then the EMS for treatments would be: 1*V(Res) + 2*V(TrxBl). The term whose EMS corresponds to this is the interaction term ‘TrxBl’. Therefore, the pseudo-F ratio for the test of no treatment effects is the mean square for treatments divided by the interaction mean square, i.e., F_{Tr} = MS_{Tr} / MS_{TrxBl}. Clearly, it would be incorrect to construct F_{Tr} = MS_{Tr} / MS_{Res}, because then, if V(TrxBl) were non-zero, we might easily reject H₀: S(Tr) = 0 even if it were true\(^\text{30}\).

In the PERMANOVA output, the terms which provide the numerator and denominator mean squares for each pseudo-F ratio are provided in the output under the heading ‘Construction of Pseudo-F ratio(s) from mean squares’ (Fig. 1.23). Also given here are the degrees of freedom associated with the numerator (‘Num.df’) and denominator (‘Den.df’) terms. Most of the time, the multipliers on these mean squares will simply be 1, and a single term can be found to provide an

\(^{30}\) Some might argue that the existence of an interaction necessarily implies the existence of at least some non-zero treatment effect(s), but we consider the null hypothesis for the test of the main effect in a mixed model such as this to include the concept of consistency in treatment effects, rather than simply the notion of whether there are any treatment effects at all. See the next section on Inference space and power.
appropriate denominator mean square for each of the relevant hypotheses in the model. In some cases, however, a linear combination of mean squares must be sought in order to construct correct pseudo-$F$ ratios to test particular terms (see the section Linear combinations of mean squares, below). PERMANOVA can deal with these situations and, in such cases, details of the linear combinations used are also provided.

Table 1.1. Null hypothesis, construction of pseudo-$F$, and ratio of expectations for each term in the mixed model for the study of Tasmanian meiofauna. Note how the construction of pseudo-$F$ isolates the component of interest under the null hypothesis (circled) so that, in each case, the numerator and denominator expectations will match one another if the null hypothesis were true.

<table>
<thead>
<tr>
<th>Term</th>
<th>Null hypothesis</th>
<th>Pseudo-$F$</th>
<th>Ratio of expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>$S(Tr) = 0$</td>
<td>$\frac{MS_{Tr}}{MS_{TrxBl}}$</td>
<td>$1<em>V(Res) + 2</em>V(TrxBl) + (8*S(Tr))$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$1<em>V(Res) + 2</em>V(TrxBl)$</td>
</tr>
<tr>
<td>Block</td>
<td>$V(Bl) = 0$</td>
<td>$\frac{MS_{Bl}}{MS_{Res}}$</td>
<td>$1<em>V(Res) + 4</em>V(Bl)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$1*V(Res)$</td>
</tr>
<tr>
<td>Treatment × Block</td>
<td>$V(TrxBl) = 0$</td>
<td>$\frac{MS_{TrxBl}}{MS_{Res}}$</td>
<td>$1<em>V(Res) + 2</em>V(TrxBl)$</td>
</tr>
</tbody>
</table>

The denominator mean square of the pseudo-$F$ ratio for any particular term in the analysis is important not just because it isolates the component of interest in the numerator for the test: it also identifies the exchangeable units needed to obtain a correct test by permutation (Anderson & ter Braak 2003). In the one-way case, it is clear that the units that are exchangeable under a true null hypothesis are the individual samples. These can be shuffled randomly among the groups (or, alternatively, the group labels can be randomly shuffled across all samples) if the groups have no effect and the null hypothesis is true. In fact, whenever a term has a pseudo-$F$ ratio with the residual mean square as its denominator, then the exchangeable units for the test are the individual samples themselves (regardless of which of the three methods of permutation offered by PERMANOVA is to be employed).

For more complex designs, the correct exchangeable units for a given test are identified by the term used as the denominator mean square of that particular term. This is sensible from the perspective that the denominator identifies what constitute the “errors” for a given null hypothesis. Thus, in the Tasmanian meiofauna example, the pseudo-$F$ ratio for the test of the main effect of treatments (‘Tr’) is $F_{Tr} = MS_{Tr} / MS_{TrxBl}$. As the denominator here is the interaction term ‘TrxBl’, the exchangeable units for this test are the 8 cells that correspond to the $2 \times 4$ combinations of treatments by blocks. The samples within each of those 8 cells will be kept together as a unit under permutation. This yields $8! / [(2!)^4 \times 4!] = 105$ unique values of the numerator and $8! / [4! \times 2!] = 840$ unique values for the denominator and, thus, 840 unique values of the whole pseudo-$F$ statistic under permutation (as shown in the ‘Unique perms’ column for the term ‘Tr’ in Fig. 1.23 above). For more details concerning exchangeable units for permutation tests in ANOVA designs, see Anderson & ter Braak (2003).

It is worthwhile pausing to consider how the above tests correspond to meaningful hypotheses for the mixed model. What is being examined by $F_{Tr}$ is the extent to which the sum of squared fixed effects can be detected as being non-zero over and above the potential variability in these effects among blocks, i.e., over and above the interaction variability (if present). Thus, in a crossed mixed model like this, $F_{TrxBl}$ first provides a test of generality (i.e., do the effects of treatments vary significantly among blocks?), whereas the test of the main effect of treatments ($F_{Tr}$) provides a test of the degree of consistency. In other words, even if there is variation in the effects of treatments (i.e. $V(TrxBl) \neq 0$), are these consistent enough in their size and direction that an overall effect can be detected over and above this? Given the pattern shown in the MDS plot, it is not surprising to
learn that, in this case, the main treatment effect is indeed discernible over and above the variation in its effects from block to block \((P = 0.0037)\).

We can contrast these results with what would have happened if we had done this analysis but treated the blocks as fixed instead of random (Fig. 1.24). The consequence of this choice is a change to the EMS for the ‘Treatment’ term, which is now \(1*V(\text{Res}) + 8*S(\text{Tr})\) and therefore contains no component of variation for the interaction. As a consequence, the pseudo-\(F\) ratio for treatment main effects in this fully fixed model is constructed as \(F_{\text{Tr}} = \frac{MS_{\text{Tr}}}{MS_{\text{Res}}}\) and, correspondingly, the denominator degrees of freedom for this test has increased from 3 to 8, the value of the pseudo-\(F\) ratio has changed from 4.67 to 8.08 and the \(P\)-value has decreased (cf. Figs. 1.23 & 1.24).

![PERMANOVA](image)

**Fig. 1.24.** Design file and analysis of Tasmanian meiofauna, treating the ‘Blocks’ as fixed.

On the face of it, we appear to have achieved a gain in power by using the fully fixed model in this case, as opposed to the mixed model that treated ‘Blocks’ as random. Power is the probability of rejecting the null hypothesis when it is false. When using a statistic like pseudo-\(F\) (or pseudo-t), power is generally increased by increases in the denominator degrees of freedom. Basically, the more information we have about a system, the easier it is to detect small effects. This choice of whether to treat a given factor as either fixed or random, however, doesn’t just affect the potential power of the test, it also rather dramatically affects the nature of our hypotheses and our inference space.

If we choose to treat ‘Blocks’ as random (Fig. 1.23), then: (i) the test of ‘TrxBI’ is a test of the generality of disturbance effects across blocks; (ii) the test of ‘BI’ is a test of the spatial variability among blocks; and (iii) the test of ‘Tr’ is a test of the consistency in treatment effects, over and above the potential variability in its effects among blocks. Importantly, the inference space for each test refers to the population of possible blocks from which we could have sampled. In contrast, if we choose to treat ‘Blocks’ as fixed (Fig. 1.24), then: (i) the test of ‘TrxBI’ is a test of whether disturbance effects differ among those four particular blocks; (ii) the test of ‘BI’ is a test of whether there are any differences among those four particular blocks; and (iii) the test of ‘Tr’ is a test of
treatment effects, ignoring any potential variation in its effects among blocks. Also, the inference space for each of these tests in the fully fixed model refers only to those four blocks included in our experiment and no others.

In the end, it is up to the user to decide which hypotheses are the most relevant in a particular situation. The choice of whether to treat a given factor as fixed or random will dictate the EMS, the pseudo-\(F\) ratio, the extent of the inferences and the power of the tests in the model. Models with mixed and random effects will tend to have less power than models with only fixed effects. (Consider: in the context of the example, it would generally be easier to reject the null hypothesis that treatments have no effect whatsoever than it would be to reject the null hypothesis that treatments have no effect given some measured spatial variability in those effects.) However, random and mixed models can provide a much broader and therefore usually a more meaningful inference space (e.g., extending to the wider population of possible blocks across the sampled study area, and not just to those that were included in the experiment). Such models will therefore tend to correspond to much more logical and ecologically relevant hypotheses in many situations.

As we have seen, PERMANOVA employs direct multivariate analogues to the univariate results for the derivation of the EMS and construction of the pseudo-\(F\) ratio, so all of the well-known issues regarding logical inferences in experimental design that would occur for the univariate case (e.g., Cornfield & Tukey 1956; Underwood 1981; Hurlbert 1984; Underwood 1997) necessarily need to be considered for any multivariate analysis to be done by PERMANOVA as well.

As a final note regarding inference, the traditional randomization test is well known to be conditional on the order statistics of the data (e.g. Fisher 1935). In other words, the results (\(P\)-values) depend on the realised data values. This fact led Pitman (1937a, b, c) and Edgington (1995) to argue that the inferences from any test done using a randomization procedure can only extend to the actual data themselves, and can never extend to a wider population\(^{31}\). In a similar vein, Manly (1997, section 7.6) stated that randomization tests must, by their very nature, only allow factors to be treated as fixed, because “Testing is conditional on the factor combinations used, irrespective of how these were chosen” (p. 142). PERMANOVA, however, uses permutation methods in order to obtain \(P\)-values, but it also (clearly) allows factors to be treated as random. How can this be?

Fisher (1935) considered the validity of a permutation test to be ensured by virtue of the \textit{a priori} random allocation of treatments to individual units in an experiment. That is, random allocation \textit{before} the experiment justifies randomization of labels to the data \textit{afterwards} in order to create alternative possible outcomes we could have observed. However, we shall consider that a permutation test gains its validity more generally (such as, for example, in observational studies, where no \textit{a priori} random allocation is possible), by virtue of (i) \textit{random sampling} and (ii) the assumption of \textit{exchangeability} under a true null hypothesis. As stated earlier (see the section Assumptions above), PERMANOVA assumes only the \textit{exchangeability} of appropriate units under a true null hypothesis. Random sampling of the levels of a random factor from a population of possible levels (like random sampling of individual samples), coupled with the assumption that “errors” (whether they be individual units or cells at a higher level in the design) are independent and identically distributed (“i.i.d.”) (e.g., Kempthorne 1952) ensures the validity of permutation tests for observational studies. For further discussion, see Kempthorne (1966), Kempthorne & Doerfler (1969) and Draper et al. (1993).

Thus, provided (i) the population from which levels have been chosen can be conceived of and articulated clearly, (ii) the exchangeability of levels can be asserted under a true null hypothesis \textit{and} (iii) random sampling has been used, then the permutation test is valid for random factors, with an inference space that logically extends to that population. In section 7.3 of the more recent edition of Manly’s book (2006), the validity of extending randomization procedures to all types of analysis of variance designs, including fixed and random factors, is also now acknowledged on the basis of this important notion of exchangeability (Anderson & ter Braak 2003).

\(^{31}\) To be fair, Edgington (1995) also suggested that making such wider inferences using the normal-theory based tests was almost always just as much a “leap of faith” as it would be for a randomization test, due to the unlikely nature of the assumptions required by the former.
Given the fact that so many important aspects of the results (pseudo-$F$ ratios, $P$-values, power, the inference space, etc.) depend so heavily on the experimental design (information given in the design file), one might wish to examine the qualities of various designs, even before embarking on the serious task of actually gathering the data. In PERMANOVA+, a special routine is provided that allows the user to explore different designs without actually analysing data. All that is required to perform a test of a given design is (i) a (dummy) data worksheet file possessing relevant factor information and (ii) a design file. The dummy data file with factor information is needed in order to identify the number of replicates per cell and how many levels there are for all of the factors.

For example, highlight the design file Mixed model in the tas.pwk file. Choose PERMANOVA+ > Test design > (Dummy data worksheet: tas) & (Sums of Squares • Type III (partial)). The results will include details of the expected mean squares (EMS) for each term in the model and the construction of the pseudo-$F$ ratio for each term, including the degrees of freedom for each test (Fig. 1.25). This allows the user to trial different experimental designs and to consider the best options, for relevant statistical inferences and for assessing power (on the basis of denominator degrees of freedom) for given hypotheses of interest. For example, in the output for the Tasmanian meiofauna, one can see that the construction for the pseudo-$F$ ratio for ‘Tr’ (the most important test of interest to the experimenter here) is $F_{Tr} = MS_{Tr} / MS_{Tr \times Bl}$. Therefore, a more powerful experimental design for the test of disturbance effects (the term ‘Tr’) would actually be obtained by increasing the number of blocks in the experiment (and thus, the degrees of freedom associated with the term ‘TrxBl’), rather than increasing the number of replicates per block (residual degrees of freedom), which would be unlikely to have any immediate effect on the test of $F_{Tr}$.

We have seen how a crossed design is identifiable by virtue of every level of one factor being present in every level of the other factor, and vice versa (e.g., Fig. 1.14). We can contrast this situation with a nested design. A factor is nested within another (upper-level) factor if its levels take on different identities within each level of that upper-level factor.
1. PERMANOVA

For example, consider a study by Anderson et al. (2005a) of the spatial variability in assemblages of invertebrates colonising holdfasts of the kelp Ecklonia radiata in northeastern New Zealand. An hierarchical sampling design was used (Fig. 1.26). Divers collected $n = 5$ individual holdfasts (separated by metres) within each of 2 areas (separated by tens of metres) within each of 2 sites (separated by hundreds of metres to kilometers), within each of 4 locations (separated by hundreds of kilometers) along the coast. The design was balanced and fully nested with three factors:

Factor A: Locations (random with $a = 4$ levels).

Factor B: Sites (random with $b = 2$ levels, nested in Locations).

Factor C: Areas (random with $c = 2$ levels, nested in Sites and Locations)

A design with a series of nested terms, like this one, is sometimes also called a fully hierarchical design. Such designs are especially useful for describing patterns and estimating variability at different temporal or spatial scales (Andrew & Mapstone 1987, Underwood et al. 2000).

Consideration of a schematic diagram for this design (Fig. 1.26) indicates directly how it differs from the crossed designs seen earlier (cf. Fig. 1.14). Unlike the crossed design, we cannot swap the order of the factors. The fact that sites are nested within locations means that we are obliged to consider them in this order: locations first (at the top of the diagram), then sites within locations, and so on. Furthermore, the sites at the first location have nothing to do with the sites at the second location. Individual site levels actually belong to particular location levels – an important hallmark of a nested factor. Finally, for a nested design, the fact that particular levels of factor B belong to particular levels of factor A indicates that the individual levels of factor B do not have a particular meaning in and of themselves. For example, the specific identity or meaning of site 1 at location 1 in the design is not the same as that of site 3 at location 2. Instead, it is clear that the levels of factor B are used to measure variability at the correct spatial (or temporal) scale in order to test the upper-level factor in the design (e.g., Hurlbert 1984). Thus, any nested factor is also, necessarily, random. An upper-level factor, however, may be either fixed or random (in the present example, the ‘Locations’ factor happens to be random).

The data from this example are located in the file hold.pri in the ‘HoldNZ’ folder of the ‘Examples add-on’ directory. There were $p = 351$ variables recorded from a total of $N = a \times b \times c \times n = 80$ holdfasts. Most of the variables were counts of abundances, but some species or taxa (primarily encrusting forms, such as sponges, bryozoans and ascidians) were recorded using a subjective ordinal rating (0 = absent, 1 = rare, 2 = present, 3 = common). For this example, we shall begin by concentrating on data obtained for molluscan only. Open the file and choose Select > Variables > (Indicator levels > Indicator name: Phylum), then click on the ‘Levels...’ button. Next, in the ‘Selection’ dialog, click on to move all of the phyla into the ‘Available’ box, then click on the name Mollusca, followed by to place it into the ‘Include’ box and click ‘OK’ (Fig. 1.27). Now with the molluscan species selected, choose Tools > Duplicate to obtain these in their own

---

32 A holdfast is a root-like structure at the base of the kelp that holds it to the substratum, which is usually a rocky reef. A great diversity of invertebrates inhabit the interstices of these complex structures.
separate worksheet and rename this **molluscs** for reference. For this new worksheet, select **Edit > Properties** and change the ‘Title’ to **NZ kelp holdfast molluscs**.

Fig. 1.27. Dialog for selection of a subset of variables (molluscs only) for the holdfast data.

Fig. 1.28. Creating the PERMANOVA design for the New Zealand holdfast data, including nesting.
The objective here is to partition the variability in the species composition of molluscs according to the three-factor hierarchical experimental design. With our focus, for the moment, on composition alone (presence/absence), we will base the analysis on the Jaccard measure, which (when expressed as a dissimilarity) is directly interpretable as the percentage of unshared species between two sample units. We wish to determine if there is significant variability among areas, among sites and among locations in the composition of the molluscan assemblage. Furthermore, if significant variability is detected at any of these levels, it would then be logical (and interesting) to estimate and compare the sizes of each component of variation, which correspond to these different spatial scales.

Fig. 1.29. Design file and PERMANOVA analysis of variability in mollusc composition (based on the Jaccard measure) from kelp holdfast assemblages.
same approach (Fig. 1.28)\(^3\). When you are finished specifying the design, rename the design file *Nested design* and save the workspace created so far as *hold.pwk*.

To run the analysis, highlight the resemblance matrix, choose **PERMANOVA+ > PERMANOVA > (Design worksheet: Nested design) & (Num. permutations: 9999)**, leaving all other options as the defaults. The results show significant variability at each level in the design (Fig. 1.29). The EMS’s reveal the rationale for constructing correct pseudo-\(F\) ratios for each term in the model (Fig. 1.29).

The EMS’s also yield another important insight: they provide a direct method to get *unbiased estimates* of each of the components of variation in the model. PERMANOVA estimates these components using mean squares, in a directly analogous fashion to the unbiased univariate ANOVA estimators of variance components (e.g., Searle *et al.* 1992). In essence, this is achieved by setting the mean squares equal to their expectations and solving for the component of interest.

For example, by setting \(MS_{Ar}\) and \(MS_{Res}\) equal to their respective expectations (placing “hats” on the parameters to indicate that we are now talking about estimates of these things, rather than their true parameter values), we have:

\[
\hat{MS}_{Ar} = 1\hat{\mu}(\text{Res}) + 5\hat{\mu}(\text{Ar(Si(Lo)))}
\]
\[
\hat{MS}_{Res} = 1\hat{\mu}(\text{Res})
\]

Thus,

\[
\hat{\mu}(\text{Res}) = MS_{Res} / 1
\]
\[
\hat{\mu}(\text{Ar(Si(Lo)))} = (MS_{Ar} – MS_{Res}) / 5
\]

From the output, we therefore can calculate these estimates directly from the mean squares. The estimated component of variation for the residual is \(MS_{Res} = 2525.7\) and for areas this is \((MS_{Ar} – MS_{Res}) / 5 = (3111.8 – 2525.7) / 5 = 117.2\). Similar logic, when applied to the other terms in the analysis yields:

\[
\hat{\mu}(\text{Si(Lo)}) = (MS_{Si} – MS_{Ar}) / 10 = 110.9
\]
\[
\hat{\mu}(\text{Lo}) = (MS_{Lo} – MSSi) / 20 = 381.7
\]

These estimates are all calculated automatically by the program and included in the output file in the column labeled ‘Estimate’ under the heading entitled ‘Components of variation’ (Fig. 1.29). For the species composition of molluscs in these kelp holdfast assemblages, the greatest component of variation occurred at the smallest spatial scale (the residual), followed by locations, and then areas and sites, with the latter two being comparable in size (Fig. 1.29).

An important point here is that these estimates are not actual “variance components” in the traditional sense unless one is analysing a single variable and the resemblance measure used is Euclidean distance. In addition, these are obviously not the same as variance-covariance matrices used in traditional multivariate statistics either (e.g. Mardia *et al.* 1979, Seber 1982), because they do not include any estimation of covariance structure at all. Rather, they are interpretable geometrically as measures of variability from a partitioning on the basis of the dissimilarity (or similarity) measure chosen.

These estimates (like their univariate counterparts of variance components) will be in terms of the squared units of the dissimilarity measure chosen. Thus, in order to put these back onto the original units, PERMANOVA also calculates their square root (provided in the column labeled ‘Sq.root’ in the results file). These values are akin to a standard deviation in a traditional univariate analysis. Thus, if the value of the dissimilarity measure used has a direct interpretation (such as the Jaccard or Bray-Curtis measures, which are both percentages), then these can be examined and interpreted as well. For example, the greatest variation in molluscan composition is at the level of individual replicate holdfasts, which (according to the square root of the estimated component of variation due

---

\(^3\) **PERMANOVA** will do the correct analysis here if we choose to nest Areas within Sites only (because Sites have already been specified as being nested in Locations), as shown, or if we choose to specify explicitly that Areas are nested in Sites and also in Locations.
to the residual of 50.3) may share only around 50% of their species, even though they may be separated by just a few metres. Over and above this, holdfasts in different areas may be an additional 10-11% dissimilar in their composition, on average, and so on.

Although the above design included all random factors, for which a discussion of estimating components of variation is a fairly natural one, we can also estimate the components of variation due to fixed effects. Recall that these are not measures of variance per se, but rather are sums of squared fixed effects divided by appropriate degrees of freedom (see the section on Components of variation above). However, if we are interested in comparing the amount of variation that is attributable to different terms in the model, estimates of components for fixed and/or random factors are useful and are directly comparable. In fact, it is indeed these estimates of components of variation that should be used as a correct basis for comparing the relative importance of different terms in the model towards explaining overall variation (Underwood & Petraitis 1993). In contrast, the raw sums of squares (whether alone or as a percentage of the total sum of squares) are not directly comparable, because different terms generally have different degrees of freedom (e.g., it would clearly be inappropriate to compare the percentage of the total sum of squares explained by a factor having only 1 degree of freedom versus some other factor that had 5 degrees of freedom).

The only potentially unsettling consequence of using analogues of the ANOVA estimators to estimate components of variation is the fact that these estimates (even in the univariate case on the basis of Euclidean distance) can sometimes turn out to be negative (Thompson 1962, Searle et al. 1992). This is clearly illogical and is generally accompanied by there being little or no evidence against the null hypothesis for the term in question that its component is equal to zero (i.e., a large P-value). Although there are other methods available for estimating variance components (i.e., ML, REML, Bayesian, etc., see Searle et al. 1992), the ANOVA estimators do have the attractive quality of being unbiased. The best solution to this issue is often to re-analyse the data after removing that term from the model (e.g., Thompson & Moore 1963, Fletcher & Underwood 2002). This leads naturally to a consideration of how to remove terms from a model, also referred to in some cases as pooling (see the following section).

For a given design file, PERMANOVA, by default, will do a partitioning according to all terms that are directly implied by the experimental design. For multi-factor designs, PERMANOVA will assume that all factors are crossed with one another, unless nesting is specified explicitly in column 2 of the design file. Although a factor cannot interact with a factor within which it is nested, factors that are crossed with one another necessarily generate interaction terms, and this full model (having all possible interactions) is generated by default.

In some cases, however, one may wish to remove one or more terms from a given model. There are various reasons for wishing to remove individual terms, including:

(i) lack of evidence against the null hypothesis of that term’s component being equal to zero;
(ii) a negative estimate of that term’s component of variation;
(iii) previous studies have determined that term’s component to be zero or negligible;
(iv) hypotheses of interest require tests of models that exclude one or more particular terms.

The user should be aware, however, that removing a term from a model equates with the assertion that its component of variation (that is, either $S(*)$ or $V(*)$ for a fixed or a random term, respectively, as the case may be) is equal to zero. By asserting that a component is equal to zero, one effectively combines, or pools, that term’s contribution (and its associated degrees of freedom) with some other term in the model. In the dialog of the PERMANOVA routine, we make a distinction between two different ways of removing a term:

- **Excluding a term from the model** – in which case the term is completely excluded and is not considered as ever having been part of the model in any form. Regardless of where the term occurs in the structure of the experimental design, excluding a term in this way is equivalent to pooling the df and SS for that term with the residual df and SS;
- **Pooling a term** – in which case the df and SS for that term is pooled with the term (or terms) which have equivalent EMS’s after that term’s component of variation is set to zero. For

---

34 An unbiased estimator is one whose expectation is equal to the parameter it is trying to estimate.
example, in a fully hierarchical design, this would correspond to a term being pooled with the term occurring immediately below it within the structure of the design.

Complete exclusion of a term might be done, for example, in cases where we wish to construct a particular model that fully ignores those terms (e.g., in designs that lack replication, see the section on Split-plot designs below). This is done by clicking on the ‘Terms…’ button in the PERMANOVA dialog. More generally, however, the removal of terms should be done with correct and appropriate pooling, where the component of variation for that term is set to zero and the EMS’s for the other terms in the model are re-evaluated. Pooling like this should be done, for example, to sequentially remove terms from a model having negative estimates for components of variation (e.g., Fletcher & Underwood 2002), or to remove terms having large P-values. In PERMANOVA, this is done by clicking on the ‘Pool…’ button in the PERMANOVA dialog.

With respect to pooling on the basis of the reason given in (i) above, the assertion that a given term’s component is equal to zero should be made with some caution. Although a P-value > 0.05 (under the usual scientific convention) may not provide sufficient evidence to reject the null hypothesis (H₀), failing to reject H₀ is nevertheless logically very different from asserting that H₀ is true (Popper 1959, 1963)! There are differences of opinion regarding how large the P-value for a given term should be before the assertion of H₀ to remove that term might be justified (e.g., Hines 1996, Janky 2000). “To pool or not to pool” is a decision left to the user, but we note that many practicing scientists use the rule-of-thumb suggested by Winer et al. (1991) and Underwood (1997) that the P-value should exceed 0.25 before removing (i.e., pooling) any given term.

Pooling a single term has important consequences for the construction of pseudo-F ratios, P-values and the estimation of components for the remaining terms. Thus, it is generally unwise to remove more than one term at a time (unless there are sufficient a priori reasons for doing so). The general rule suggested by Thompson & Moore (1963) and Fletcher & Underwood (2002) is, when faced with more than one term which might be removed from a given model, remove only one term at a time, beginning with the term having the smallest mean square, and at each step re-assess whether more terms should be removed or not.

![Fig. 1.30. Calculating the average taxonomic distinctness of molluscs for New Zealand holdfast assemblages.](image-url)
As an example of pooling, consider the New Zealand holdfast assemblages discussed in the previous two sections. Here, we shall focus on the analysis of a single variable – the average taxonomic distinctness (AvTD, Δ+, Warwick & Clarke 1995) for molluscs in holdfasts. Open up the file hold.pwk and, from within this workspace, open up the aggregation file for the molluscs, called Mollusca.agg. Next, highlight the molluscs worksheet (see the section Nested design above for details on obtaining this worksheet). Use the built-in tool in PRIMER to obtain the average taxonomic distinctness in each sample as a worksheet: select Analyse > Diverse and remove the √ in front of all of the default options except for (√AvTD: Δ+) shown under the ‘Taxdisc’ tab & (√Results to worksheet) (Fig. 1.30). (Note that this also requires clicking on each of the ‘Other’, ‘Shannon’ and ‘Simpson’ tabs, in turn, in order to remove the √ for those options as well). Rename the resulting worksheet delta+ (which should contain only one variable, ‘Delta+’) and select Edit > Properties to give it the title: NZ holdfast AvTD for molluscs.

Next, from the delta+ worksheet, calculate a Euclidean distance resemblance matrix: Analyse > Resemblance > (Analyse between Samples) & (Measure Euclidean distance). Do the analysis on the basis of the nested design (see the design file in Fig. 1.29) using the PERMANOVA routine with (Design worksheet: Nested design) & (Num. permutations: 9999) and all other choices as per the defaults. This analysis will result in an ANOVA partitioning (yielding values for df, SS, MS, F ratios and estimates of variance components) equivalent to that obtained using classical univariate ANOVA. (The data in this worksheet can be exported and analysed using a different statistical package to confirm this). The only difference will lie in the P-values, which of course are obtained using permutations in PERMANOVA, as opposed to using the traditional tables of the F distribution which rely on the assumption of normality.

The results suggest that there is no significant variability in AvTD for molluscs among different sites (Fig. 1.31). Furthermore, the P-value for ‘Si(Lo)’ is quite large (P > 0.90) and the estimate of the variance component for ‘Si(Lo)’ is negative. Thus, using the rationale of either (i) or (ii) above, we may remove this term from the model by pooling it.

Fig. 1.31. Results of PERMANOVA for the full 3-factor nested model on AvTD of molluscs.

This is achieved relatively easily, by re-running the PERMANOVA routine on the basis of the same design file, but this time, click on the ‘Pool…’ button (Fig. 1.32). A new dialog box appears entitled ‘Selection’ with a list of all of the terms in the model. The user may choose which terms in
the model to pool. For the present case, we wish to pool the ‘Si(Lo)’ term. In the ‘Available’ box, click on the term Si(Lo) followed by to move this into the ‘Include’ box, then ‘OK’.

To understand how pooling is done in the PERMANOVA routine, consider the details of the EMS in the present example for the full model before pooling:

<table>
<thead>
<tr>
<th>Source</th>
<th>EMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lo</td>
<td>1<em>V(Res) + 5</em>V(Ar(Si(Lo))) + 10<em>V(Si(Lo)) + 20</em>V(Lo)</td>
</tr>
<tr>
<td>Si(Lo)</td>
<td>1<em>V(Res) + 5</em>V(Ar(Si(Lo))) + 10*V(Si(Lo))</td>
</tr>
<tr>
<td>Ar(Si(Lo))</td>
<td>1<em>V(Res) + 5</em>V(Ar(Si(Lo)))</td>
</tr>
<tr>
<td>Res</td>
<td>1*V(Res)</td>
</tr>
</tbody>
</table>

By pooling the term ‘Si(Lo)’, we are explicitly asserting that the component V(Si(Lo)) = 0. Setting this component deliberately to zero wherever it appears yields the following:

<table>
<thead>
<tr>
<th>Source</th>
<th>EMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lo</td>
<td>1<em>V(Res) + 5</em>V(Ar(Si(Lo))) + 20*V(Lo)</td>
</tr>
<tr>
<td>Si(Lo)</td>
<td>1<em>V(Res) + 5</em>V(Ar(Si(Lo)))</td>
</tr>
<tr>
<td>Ar(Si(Lo))</td>
<td>1<em>V(Res) + 5</em>V(Ar(Si(Lo)))</td>
</tr>
<tr>
<td>Res</td>
<td>1*V(Res)</td>
</tr>
</tbody>
</table>

Thus, the mean square for the term ‘Si(Lo)’ and the mean square for the term ‘Ar(Si(Lo))’ are now estimating the same thing, i.e. they have the same expectation. This means that their SS and df can be added together to obtain a pooled MS, as follows:

\[ MS_{\text{Pooled}} = \frac{SS_{\text{Si(Lo)}} + SS_{\text{Ar(Si(Lo))}}}{df_{\text{Si(Lo)}} + df_{\text{Ar(Si(Lo))}}} \] (1.5)

The PERMANOVA output file from the analysis after pooling identifies which terms were pooled, under the heading ‘Pooled terms’ and also identifies the terms whose SS and df were combined as a consequence of this (Fig. 1.32). The new pooled term is given a unique name – it is simply called ‘Pooled’ in the present case. In the event of there being more terms to pool, potentially more than one pool of terms may occur in a single analysis. This is also catered for by the PERMANOVA routine, if required, with each pool identified by its component terms and given its own unique name. The EMS’s after pooling in the present case are:

<table>
<thead>
<tr>
<th>Source</th>
<th>EMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lo</td>
<td>1<em>V(Res) + 5</em>V(Ar(Si(Lo))) + 20*V(Lo)</td>
</tr>
<tr>
<td>Pooled</td>
<td>1<em>V(Res) + 5</em>V(Ar(Si(Lo)))</td>
</tr>
<tr>
<td>Res</td>
<td>1*V(Res)</td>
</tr>
</tbody>
</table>

and the associated degrees of freedom for the pooled term are:

\[ df_{\text{Pooled}} = df_{\text{Si(Lo)}} + df_{\text{Ar(Si(Lo))}} = 4 + 8 = 12 \] (1.6)

Pooling the ‘Si(Lo)’ term has resulted in all of the remaining estimated components of variation in the model being non-negative. It has also, however, substantially changed the \( F \)-ratios and tests for the remaining terms. The term ‘Lo’ is now not statistically significant (pseudo-\( F = 1.16, P = 0.36 \), Fig. 1.32), whereas before, when tested using the mean square for ‘Si(Lo)’ as the denominator, it was (pseudo-\( F = 19.1, P = 0.01 \), Fig. 1.31). This is an important point. Although pooling may result in an increase in power, caused by an increase in the denominator df for the test of a given term (here, ‘Den.df’ for the test of ‘Lo’ has gone up from 4 to 12), this is also often off-set by an increase in the denominator MS for the test after pooling (cf. \( MS_{\text{Si(Lo)}} = 0.68 \), whereas \( MS_{\text{Pooled}} = 11.2 \)), which reduces the value of pseudo-\( F \) for the test. It is usually not possible to tell \textit{a priori} just how the tests for other terms in the model will be affected by pooling. Clearly, however, estimated components of variation, pseudo-\( F \) ratios and P-values will be affected by pooling. Thus, as stated previously, pooling should only be done one term at a time. From the present analysis, after pooling, there is apparently no statistically significant variability in the AvTD of molluscs among holdfasts at any of the spatial scales examined (Locations, Sites or Areas).

Note that these results (Fig. 1.32), obtained by removing the ‘Si(Lo)’ term using the ‘Pool…’ button in the PERMANOVA dialog, are \textit{not} the same as the results that would have been obtained if we had simply excluded ‘Si(Lo)’ from the model entirely, using the ‘Terms…’ button. In that case, the term would have been considered to be non-existent. It would not have been included in the partitioning at all and its SS and df would therefore have ended up as part of the residual.
1. PERMANOVA

is clearly illogical and undesirable in the present case. Removing the ‘Si(Lo)’ term should result in an increase to the \(df\) associated with the denominator MS being used to test the ‘Lo’ term (i.e., its SS and \(df\) should be combined with the ‘Ar(Si(Lo))’ term), and not merely ignored and added to the residual variation. Another possibility would be to re-cast the model with two factors: ‘Lo’ and ‘Ar(Lo)’, but we would need to be careful and make sure that the areas within any location (even those from different sites) each had a unique name in the specification of the ‘Areas’ factor levels.

Fig. 1.32. Dialog to pool the ‘Si(Lo)’ term and results for the PERMANOVA analysis of the AvTD of molluscs after pooling.

If pooling of a given term would result in its being combined with the residual in any event (which of course, occurs legitimately in some cases), then the results using these two approaches will be equivalent. Otherwise, however, they will not. We consider that removal of terms using the ‘Pool…’ button will be most appropriate for the majority of situations. The use of the ‘Terms…’ button to exclude terms entirely may be useful, however, to craft specific simplified models in the event that replication is lacking (see the section Designs that lack replication). The dialog available under the ‘Terms…’ button can also be used to change the order in which individual terms are fitted. Although the order in which terms are fitted is of no consequence for balanced ANOVA designs, it is important for analyses of unbalanced designs or designs including covariates using Type I (sequential) SS (see the sections Unbalanced designs and Designs with covariates).

A topic related to the issue of pooling is the issue of designs that lack replication. Familiar examples are some of the classical experimental designs, primarily from the agricultural literature, such as randomised blocks, split plots or latin squares (e.g. Mead 1988). For these designs the essential issue is that there is no replication of samples within cells, but rather there is only 1 sample per cell. This means that it is not possible to distinguish between variation among samples and variation due to the highest-order (most complex) interaction term. Thus, in order to proceed, the experimenter has either to assume (i) that the highest-order interaction term is zero, or (ii) that the so-called “residual” mean square in the model actually has expectation \(V(\text{Res}) + V(\text{highest-order interaction})\). Note that, for the latter assumption to work, at least one of the factors involved in the highest-order interaction has to be random. See Gates (1995) and chapter 10 of Quinn & Keough (2002) for further discussion of these issues.
From a practical perspective, for PERMANOVA to proceed with the analysis (regardless of which of the above two perspectives one chooses to take), the highest-order interaction term needs to be excluded from the analysis (see the section Pooling or excluded terms, above). This can either be done manually, or if the PERMANOVA routine detects that there is no within-cell replication, then it will issue a warning. If you choose to proceed by clicking ‘OK’, it will automatically exclude the highest-order interaction term from the model. If you receive this warning and you know that you do have within-cell replication, then there is a very good chance that you have mis-labeled your factor levels somehow.

An example of a two-way crossed design without replication is provided in a study by Winsor & Clarke (1940) to investigate the catch of various groups of plankton by two nets hauled horizontally, with one net being 2 metres below the other. Ten hauls were made with the pair of nets at depths of 29 and 31 meters, respectively. The experimental design is:

- **Factor A:** Position (fixed with $a = 2$ levels, either upper (U) or lower (L) depths).
- **Factor B:** Haul (random with $b = 10$ levels, labeled simply 1-10).

There is only 1 value per combination of treatments, with no replication, so $N = a \times b = 20$. This is effectively a randomised block design, where the hauls are “blocks”. The variables recorded correspond to five different groups of plankton. Standard deviations in the various groups were roughly proportional to the means, so data were transformed and are provided as logarithms of the catch numbers for each of the plankton groups. These data are located in the plank.pri file in the ‘Plankton’ folder of the ‘Examples add-on’ directory, and were provided by Snedecor (1946).

**Fig. 1.33.** PCA of the study of plankton from ten hauls (numbered) at either 29 m depth (upper) or 31 m depth (lower).

Examination of the data (already log-transformed) reveals no zeros and that their distributions are fairly even, with no extreme values or outliers. The variables are also on similar scales and are measured in the same units; therefore, an analysis based directly on Euclidean distances would be reasonable here – prior normalisation is not necessary. For data like these, an appropriate ordination method is principal components analysis (PCA). The first two principal components

---

35 For example, you may have given the levels of factor B the names b1, b2, b3 within the first level of factor A, but then called them B1, B2, B3 within the second level of factor A, and PRIMER will not interpret these names as being the same.

36 PRIMER’s Analyse>Draftsman plot routine is very useful for visually examining the distributions and joint distributions of variables in a worksheet.

37 For more details regarding this method and its implementation in PRIMER, see chapter 4 of Clarke & Warwick (2001) and chapter 10 in Clarke & Gorley (2006).
1. PERMANOVA

explained 83.6% of the total variance in the five variables (Fig. 1.33). Variability among the hauls is apparent in the diagram, but a clear difference in the plankton numbers due to the position of the nets (upper versus lower), if any, is not obvious.

The PERMANOVA analysis of these data on the basis of a Euclidean distance matrix has detected significant variability among the hauls, but also has detected a significant effect of the position of the net (Fig. 1.34). Notice that the output has identified the excluded term: ‘PositionxHaul’, as an important reminder that the analysis without replication is not without an additional necessary assumption in this regard.

It might seem surprising that the analysis has detected any effect of ‘Position’ at all, given the pattern seen in the PCA (Fig. 1.33). Looks can be deceiving, however. Close inspection of the plot reveals that, within almost all of the individual hauls (i.e., 1, 2, 3, 5, 7, 9, 10), the symbol for the ‘upper’ group lies to the right of the symbol for the ‘lower’ group. Only hauls 4, 6 and 8 do not conform to this pattern. We can perhaps understand the nature of this overall effect of Position by examining the averages for the ‘upper’ and ‘lower’ nets for each of the variables across all of the hauls. With the Plankton worksheet highlighted, select Tools > Average > (Samples • Averages for factor: Position) & (Variables • No averaging). The resulting worksheet shows that the average log(abundance) for all five of the plankton variables was larger for the nets towed at the shallower depth (the ‘upper’ nets) (Fig. 1.34).

![Fig. 1.34. Design file and PERMANOVA analysis for the two-way study that lacks replication within cells.](image)

The interaction term is, by necessity, excluded from the analysis, as it is already confounded with the residual variance. Also shown are the averages per depth for each of the 5 variables in the plankton study.

Another important point here is to recognise that, had we treated the above design as if the hauls were the replicates, and ignored the variation among hauls, then we would not have detected any effect of position at all. You can check this fact by running PERMANOVA on the data using a one-way design with the factor ‘Position’ only (the result is non-significant, with $P > 0.25$). Thus,
1. PERMANOVA

despite the fact that there is a consistent shift in the plankton assemblage between the upper and lower nets within each haul, the variation from haul to haul would have masked this entirely and we would have failed to detect it (as we at first did when contemplating the PCA plot), if we had not included the factor ‘Haul’ in our analysis. The advantages of “blocking” to achieve greater power to detect treatment effects have been known for a very long time (e.g., Fisher 1935, Snedecor 1946, Mead 1988), but this example shows that the phenomenon can occur equally strikingly in the analysis of multivariate data.

The analysis of some other designs that lack replication have other issues, on top of the one already noted regarding the highest-order interaction being inextricably confounded with the residual. For example, the experimental design known as the *latin square* consists of a random allocation of *t* treatments to a *t* × *t* matrix of sample units, with the added constraint that there be one of every treatment in every row and one of every treatment in every column of this array. The usual model fitted to such a design partitions the sum of squares according to the following sources: rows (R), columns (C) and treatments (T). None of the potential interaction terms (RxC, RxF, CxT, RxCxF) are traditionally included in these models, because none of them can be readily unconfounded from the residual. PERMANOVA does not have separate subroutines for treating these kinds of special designs, but will not give sensible results unless the terms which cannot be estimated are first removed from the model. For these more complex designs lacking replication, *it is up to the user to know if such interactions need to be excluded, to understand the consequences of the assumptions underlying these models if they are to be used, and to exclude the relevant terms manually, using the ‘Terms…’ button in the PERMANOVA dialog.*

Another special case of a design lacking appropriate replication is known as a *split-plot design*. These designs usually arise in an agricultural context, where the experimenter has applied the treatment levels for a factor (say, factor A) randomly to whole plots (usually within blocks of some kind) at a large scale, but then, within each of these whole plots, treatments for another factor (say, factor B) are applied randomly to smaller units. Thus, the whole plots are each “split” into smaller units. Although there are variations on this theme, the traditional split-plot design lacks replicates of the whole plots at the larger spatial scale (i.e., it is a randomised block design for factor A, with whole plots acting as the ‘error’). There is also commonly a lack of replication at the smaller spatial scale (i.e., the number of levels of factor B is equal to the number of sample units within each whole plot and these levels are allocated randomly and separately within each whole plot).

**Fig. 1.35.** Schematic diagram of the Woodstock split-plot design examining the potential effects of fire frequency (no burning, burning every 2 years, every 4 years or every 8 years) and the effect of grazers (by fencing some sub-plots (F), while leaving others unfenced (U)) on plant assemblages.

---

38 (or, at least, not the results that are traditionally provided for such designs)
A proposed rationale for using a split-plot design is that factors may occur naturally at different scales (e.g., Mead 1988). Another proposed rationale is that one may already know that factor A has important effects, and one may be willing to sacrifice information on factor A to get more precise results for factor B and the interaction A×B. Although neither of these actually provides a solid argument for ignoring the need for appropriate replication in experiments, split-plot designs do still occur from time to time in biological research and can be analysed (after some thought and with care) using PERMANOVA. For more information regarding the assumptions and potential disadvantages of split-plot designs, see Mead (1988) and Underwood (1997).

To analyse a split-plot design in PERMANOVA, effectively two analyses must be done: one at the ‘whole-plot’ level and one at the ‘sub-plot’ level. Results of these two analyses can then be combined to form the traditional partitioning that is usually presented for such designs. An example of a split-plot design is provided by a study of the effect of fire disturbance and fencing (to exclude grazers) on the composition of plant assemblages on the central western slopes of New South Wales in south-eastern Australia (Prober et al. 2007). The experimental design (shown schematically in Fig. 1.35) included the following factors:

- **Blocks** (random with \( r = 4 \) levels).
- **Factor A**: Fire frequency (fixed with \( a = 4 \) levels: 0 yrs, 2 yrs, 4 yrs or 8 yrs).
- **Whole-plots** (random and nested within Blocks and Fire frequency, unreplicated).
- **Factor B**: Fencing (fixed with \( b = 2 \) levels: fenced and unfenced)
- **Sub-plots** (random and nested within all of the above, unreplicated).

The two fencing treatments were randomly allocated to two sub-plots (measuring 5 m × 5 m) within each fire treatment (whole plots) and there is one of each fire treatment (4 whole plots) randomised within each block. Relative abundances (cover) of higher plant species within each sub-plot were estimated using a point-intercept technique (an 8 mm dowel placed vertically at each of 50 points on a grid across each plot). Although the design was set up at each of two locations (Woodstock and Monteagle) and data were obtained over a number of years (Prober et al. 2007), we consider here only data from the Woodstock location collected in 2003. We also will exclude two species: *Poa sieberiana* and *Themeda australis*, the dominant grasses, from the analysis. These have already been analysed separately in detail (see Prober et al. 2007) and our focus here instead will be on the more subtle potential responses of subsidiary forbs and exotic species.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block</td>
<td>((r - 1) = 3)</td>
</tr>
<tr>
<td>Fire</td>
<td>((a - 1) = 3)</td>
</tr>
<tr>
<td>Whole-plot error</td>
<td>((r - 1)(a - 1) = 9)</td>
</tr>
<tr>
<td>Whole-plot total</td>
<td>((ra - 1) = 15)</td>
</tr>
<tr>
<td>Fence</td>
<td>((b - 1) = 1)</td>
</tr>
<tr>
<td>Fire × Fence</td>
<td>((a - 1)(b - 1) = 3)</td>
</tr>
<tr>
<td>Sub-plot error</td>
<td>(a(b - 1)(r - 1) = 12)</td>
</tr>
<tr>
<td>Total</td>
<td>(abr - 1 = 31)</td>
</tr>
</tbody>
</table>

Partitioning for a split-plot design is traditionally done according to Table 1.2. The upper-level factors (Blocks and Fire frequency) are tested against the whole-plot error, while the lower-level factors (Fencing and the Fire × Fencing interaction term) are tested against the sub-plot error. To do this partitioning and the necessary tests using PERMANOVA, we first focus on the top-half of Table 1.2 only. This calls for a randomised block design, but where the whole plots are effectively treated as the sample units. So, first we need to obtain distances among centroids\(^{39}\) for the whole plots. Open the file `wsk.pri` (in the ‘Woodstock’ folder of the ‘Examples add-on’ directory).

\(^{39}\) Importantly, these centroids are not calculated on the original data, they are calculated on the full set of PCO axes obtained from the resemblance matrix, in order to preserve the resemblance measure chosen as the basis of the analysis. For more details, see chapter 3.
containing the abundances (cover measures) for \( p = 117 \) plant species. First select all of the variables except the variables numbered 50 and 63 in the dataset (‘Poa sieb’ and ‘Themaus’, respectively). Calculate a Bray-Curtis resemblance matrix after square-root transforming the selected data and choose PERMANOVA+ > Distances among centroids… > Grouping factor: WholePlot, then click ‘OK’. This yields a new matrix of Bray-Curtis resemblances among the 16 whole plots (4 fire treatments \( \times 4 \) blocks). Next, run a two-way randomised block design of Block and Fire on the resemblance matrix among centroids (‘Resem2’). As there is no replication of the whole plots, you will either have to remove the ‘Block \( \times \) Fire’ interaction term manually (by clicking on the ‘Terms…’ button in the PERMANOVA dialog), or let PERMANOVA do that for you. This analysis will give results for the top half (the first four terms) of the table (Fig. 1.36).

Fig. 1.36. Step one in the PERMANOVA analysis of the Woodstock split-plot design: a two-way randomised block design for whole-plots.

Now, for the lower half of the table, we will want to fit the ‘Fence’ and ‘Fire \( \times \) Fence’ terms, given the whole-plots, in order to get the correct sub-plot error. Go back to the original resemblance matrix among all 32 samples (‘Resem1’) and set up a PERMANOVA design file with three factors: WholePlot, Fire frequency and Fencing (Fig. 1.37). As usual, PERMANOVA attempts to construct and fit a full model, including all interaction terms implied by the structure that is specified. In the present case, it is not possible to fit all interaction terms, because of the lack of replication. Here, it is not just the lowest level in the analysis that is unreplicated, but we also lack replication at a higher level in the design. PERMANOVA will not automatically exclude the terms that would normally be excluded from a split-plot analysis, so these must be excluded manually by clicking on the ‘Terms…’ button in the PERMANOVA dialog and choosing only those terms in the model that we wish to fit (Fig. 1.37). For this design, there are a number of terms that are purposefully excluded: namely, any interactions involving whole plots as well as the main effect of Fire. (This is on top of the fact that we have also chosen to ignore any possible interactions involving Blocks.) It is important to recognise these assumptions underlying any split-plot analysis, and the requirement to manually remove terms is a good reminder of what is going on here. Of course, if we had replicated whole-plots and replicated sub-plots, we would be in a position to analyse the whole design, including all interactions, in a single PERMANOVA analysis.
Fig. 1.37. Step two in the PERMANOVA analysis of the Woodstock split-plot design: a three-way analysis of sub-plots, excluding certain terms, but including whole-plots.

Once we have the results from both “halves” of the split-plot analysis, we can use these to construct the complete table (Table 1.3). When combining the results of more than one analysis from a single set of data, such as this, it is a good idea to check that the partitioning has been done correctly by making sure that the sum of the individual SS add up to the total SS. This is true for the present example (Table 1.3) and will be true in general, at least for a correct partitioning of any balanced design. The analysis suggests that the fencing treatments had no significant effect on these assemblages, but fire frequency did. There was no evidence that these two factors interact significantly with one another. Spatial variation among blocks and among whole plots was also substantial (see the relative sizes of components of variation in the output file).

Table 1.3. Results from the PERMANOVA analysis of plant assemblages in response to fire frequency and fencing (removal of grazers) in the Woodstock split-plot experiment.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>Pseudo-F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block</td>
<td>3</td>
<td>4946</td>
<td>1648.6</td>
<td>2.839</td>
<td>0.0003</td>
</tr>
<tr>
<td>Fire</td>
<td>3</td>
<td>3622</td>
<td>1207.4</td>
<td>2.079</td>
<td>0.0069</td>
</tr>
<tr>
<td>Whole-plot</td>
<td>9</td>
<td>5226</td>
<td>580.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole-plot error</td>
<td>15</td>
<td>13793</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fence</td>
<td>1</td>
<td>840</td>
<td>840.5</td>
<td>1.449</td>
<td>0.1625</td>
</tr>
<tr>
<td>Fire × Fence</td>
<td>3</td>
<td>2252</td>
<td>750.7</td>
<td>1.294</td>
<td>0.1407</td>
</tr>
<tr>
<td>Sub-plot error</td>
<td>12</td>
<td>6963</td>
<td>580.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>31</td>
<td>37642</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

While randomised blocks, latin squares and split-plot designs lack spatial replication, a special case of a design lacking temporal replication (and which occurs quite a lot in ecological sampling) is the repeated measures design (e.g., Gurevitch & Chester 1986, Green 1993). In essence, these designs...
consist of individual sampling units (usually belonging to various treatments, etc.) which are repeatedly examined at several different time points. Such designs receive special attention for two reasons: (i) they do not have replication within cells, so suffer from the same issues discussed in the previous two sections and (ii) they require some additional assumptions because samples are generally considered to be non-independent through time. The first issue – lack of replication – is dealt with easily enough by PERMANOVA, as discussed above. The program (after issuing the usual warning) will simply exclude the highest-order interaction term (i.e. which will include “Time” as one of its members) and the analysis proceeds from there in the usual way.

The issue of non-independence is, however, another matter. In a traditional repeated measures analysis of univariate data, the partitioning of the total sum of squares is done in the usual way, treating “Time” as a fixed factor. What is taken on as an additional assumption, however, is something known as sphericity. Sphericity is an assumption about the nature of the correlations through time for the sample units, which must be similar for the different treatments. Although a formal test of sphericity is provided by Mauchley (1940) (see Winer et al. 1991 for an example), this approach is unfortunately rather highly susceptible to deviations from normality (Huynh & Mandeville 1979). Huynh & Feldt (1970) have demonstrated that a necessary and sufficient condition is to check the equality of the variances of the differences between levels of the repeated measures factor (“Time”) across treatments (or combinations of treatments). Thus, one calculates the differences between each pair of time points, obtains the variances of these difference values for each treatment (or treatment combination) and then checks for equality of these variances (e.g., Quinn & Keough 2002). If this assumption is violated, then the $F$ ratios obtained from the analysis are no longer distributed like traditional $F$ distributions under true null hypotheses. In this case, for traditional univariate analysis, a number of possible corrections to the degrees of freedom can be done to get a correct test (e.g., Box 1954, Geisser & Greenhouse 1958, Huynh & Feldt 1976).

When dealing with multivariate response data, one might consider doing an analogous test of sphericity (of some sort) by calculating the dissimilarities (or distances) between levels of the repeated measures factor across treatments. The variances of these dissimilarities could then be compared among treatments (or treatment combinations) using, for example, PERMDISP (see chapter 2), or even using a traditional test for homogeneity of variances among groups.

Recall, however, that PERMANOVA uses permutation procedures in order to generate a correct distribution of each pseudo-$F$ statistic under a (relevant) true null hypothesis. So the only essential assumption associated with the use of PERMANOVA (whether there be repeated measures or otherwise) is the exchangeability of samples (or of appropriate residuals). It must be admitted that the correlation structure among samples through time, if any, will be effectively ignored under permutation. Thus, differences in correlation structure through time among treatments (i.e. lack of sphericity) may, therefore, produce a statistically significant result. However, we consider that differences in correlation structure through time are indicative of (at least one type of) a treatment effect, so should warrant closer inspection by the investigator in any event.

Clearly, the degree to which correlation structure (in space or in time, as in repeated measures) can affect the results of permutation tests for repeated measures designs (or any other designs for that matter) warrants further study. If statistically significant results are obtained in a repeated measures analysis, the user may wish to accompany the PERMANOVA with a separate test for sphericity (in the case of univariate data) or its analogue (using dissimilarities rather than differences) for multivariate responses, in order to shed further light on the meaning of the results and appropriate inferences.

An example helps to clarify these ideas. The data on Victorian avifauna, previously examined in the section Monte-Carlo $P$-values above, actually consisted of a repeated measures design, as described by Mac Nally & Timewell (2005). The previous analyses (Figs. 1.12 and 1.13) were based on data summed across four different observation times. However, these data are also available at the level of individual surveys done at each time in the file vicisurv.pr in the folder ‘VictAvi’ of the ‘Examples add-on’ directory. Open up the data and calculate the dissimilarity matrix among samples on the basis of the binomial deviance dissimilarity measure; re-name this matrix BinomDev.
An MDS ordination on the basis of this matrix suggests a strong effect of flowering intensity on the bird communities, with ‘good’ sites appearing to the left of the diagram, ‘poor’ or ‘medium’ sites occurring to the right, and ‘adjacent’ samples being more variable than the other treatments through time (Fig. 1.38). The repeated measures experimental design here is:

- Factor A: Treatment (fixed with $a = 4$ levels: poor, medium, good or adjacent).
- Factor B: Site (random, nested in Treatments with $b = 2$ levels, labeled simply S1-S8).
- Factor C: Time (fixed with $c = 4$ levels, labeled 1-4).

After creating the design file, re-name it “Repeated measures” for reference. Note that there is no replication within cells, as there is only one sample per site at each time. We therefore need to exclude the highest-order interaction, ‘Site(Treatment) × Time’, from the model. See Fig. 1.39 and refer to the above section Pooling or excluding terms, if necessary, for details on how to do this. Analysis by PERMANOVA (with 9999 permutations of residuals under a reduced model) reveals significant treatment effects on the bird assemblages (Fig. 1.39), and shows that these effects are consistent through time (note that $P > 0.6$ for ‘TrxTi’). Pair-wise comparisons (not shown here, but you can do them easily) reveal that this effect is largely due to there being significant differences between the good sites vs the others.

Having identified significant treatment effects, we may wish to examine the multivariate analogue to the test of sphericity for univariate data. Namely, we can calculate the dissimilarities between each pair of time points within each treatment. Then, we can compare the time-point differences for equality of variances. This is purely optional and is not a requirement of the repeated measures analysis when performed using PERMANOVA, but it may nevertheless shed some further light on whether the significant treatment effects detected were due only to differences in location in multivariate space (as would appear to be the case from the diagram) or whether sites might also vary in the nature of their non-independence among time points.

To do this, we first need to extract the relevant dissimilarities between time-points for each sampling unit (in this case, from each site) from the full dissimilarity matrix. These are shown in Table 1.4. For example, the binomial deviance dissimilarity between times 1 and 2 for site 1 (a ‘poor’ site) is 17.85, and so on. These multivariate dissimilarity values can take on the same role as the differences among time points which are usually examined in a univariate repeated measures analysis.

---

$^a$We have chosen to treat “Time” as a fixed factor here, because this is traditionally how it is treated in repeated measures experimental designs. However, the user can of course choose to treat this factor as random if this is more in line with relevant hypotheses of interest.
1. PERMANOVA

The estimated variances in these dissimilarities (given in the last row of Table 1.4) appear to be similar among the six paired groups (i.e., the six columns in Table 1.4). Levene’s test (using either means $F_{5,42} = 0.59, P > 0.71$ or medians $F_{5,42} = 0.32, P > 0.90$), supports the assumption of homogeneity of these variances\textsuperscript{42}. This means we can fairly safely infer in this case that the treatment effects we detected were not caused by differences in dissimilarity structure within samples through time among the sites\textsuperscript{43}.

![Fig. 1.39. PERMANOVA analysis of a repeated measures experimental design for Victorian avifauna.](image)

Before leaving the topic of repeated measures, there is one other method of analysis worth mentioning. With univariate data, one has the option of treating the individual values at each time point as separate variables, then doing a multivariate analysis among treatments. Note that this approach will only work when there is one response variable measured at different times. It is to be distinguished from the situation we have above, where there are already multivariate data (i.e., abundances of 27 species of birds) obtained at each time point. The multivariate approach to a repeated measures analysis of a single response variable is, however, straightforward to do using PERMANOVA – one simply organises the data so that measures at different time points are the

\textsuperscript{41} Bear in mind that the univariate differences, however, can show direction by being either positive or negative, which will affect calculated variances. In contrast, dissimilarities are always positive, so cannot show direction, \textit{per se}. Therefore, this is not at all intended to be a strict test for sphericity, merely to provide a means of examining the null hypothesis of no difference in dissimilarity structure among time points \textit{for individual samples} across the different treatments. See also the approach used by Clarke \textit{et al.} (2006b).

\textsuperscript{42} Note the use of a traditional univariate Levene’s test here for testing homogeneity of variances. We could also have done this test using a permutational approach in the routine PERMDISP on the basis of a Euclidean distance matrix for the univariate variable produced in Table 1.1. Indeed, if the tables are used, then this is equivalent to doing a traditional Levene’s test. See the chapter on PERMDISP for more details.

\textsuperscript{43} Winer \textit{et al.} (1991) discuss how correlations may need to be examined at several different levels in more complex multi-factorial repeated measures designs (see chapter 7 therein). We do not pursue this further here.
“variables” and uses Euclidean distance as the basis of the analysis. Although taking a multivariate approach with univariate data (treating the time points as variables) completely avoids having to consider any notions of sphericity, it does not provide any test of the factor ‘Time’ or any tests of the interactions of ‘Time’ with (most) other factors, which would, of course, be provided by a traditional repeated measures partitioning. Green (1993) discusses other pros and cons with using the multivariate versus the univariate approach to a repeated measures analysis of a single variable (see Fig. 6 therein).

Table 1.4. Binomial deviance dissimilarities between time points for each site and their estimated variances.

<table>
<thead>
<tr>
<th>Site</th>
<th>T1-T2</th>
<th>T1-T3</th>
<th>T1-T4</th>
<th>T2-T3</th>
<th>T2-T4</th>
<th>T3-T4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor</td>
<td>17.85</td>
<td>13.84</td>
<td>9.85</td>
<td>17.08</td>
<td>15.58</td>
<td>9.79</td>
</tr>
<tr>
<td></td>
<td>3.90</td>
<td>6.02</td>
<td>4.48</td>
<td>7.56</td>
<td>8.15</td>
<td>5.24</td>
</tr>
<tr>
<td>Medium</td>
<td>6.65</td>
<td>15.00</td>
<td>18.44</td>
<td>22.51</td>
<td>29.11</td>
<td>25.19</td>
</tr>
<tr>
<td></td>
<td>8.17</td>
<td>14.32</td>
<td>11.87</td>
<td>21.98</td>
<td>18.76</td>
<td>3.79</td>
</tr>
<tr>
<td>Good</td>
<td>3.13</td>
<td>7.15</td>
<td>7.88</td>
<td>10.56</td>
<td>11.48</td>
<td>16.80</td>
</tr>
<tr>
<td></td>
<td>8.07</td>
<td>16.25</td>
<td>17.42</td>
<td>16.92</td>
<td>18.86</td>
<td>3.41</td>
</tr>
<tr>
<td>Adjacent</td>
<td>12.14</td>
<td>3.11</td>
<td>24.97</td>
<td>7.88</td>
<td>25.51</td>
<td>21.00</td>
</tr>
<tr>
<td></td>
<td>38.89</td>
<td>18.95</td>
<td>31.37</td>
<td>26.61</td>
<td>19.49</td>
<td>9.96</td>
</tr>
<tr>
<td>Var</td>
<td>137.06</td>
<td>31.75</td>
<td>82.53</td>
<td>51.18</td>
<td>46.83</td>
<td>67.67</td>
</tr>
</tbody>
</table>

For repeated measures designs and in other cases where there is known to be correlation structure (non-independence) among replicates, there may be other ways to analyze the data. First, one might consider rephrasing hypotheses in terms of dissimilarities between particular pairs of correlated objects, then analyzing those dissimilarities in a univariate analysis. For example, Faith et al. (1991) tested the null hypothesis of no difference in the average dissimilarity between an impact and a control site from before to after the onset of a disturbance (with individual time points treated as replicates). Similarly, if one has a baseline or control against which other treatments are to be compared through time, one might consider using principal response curves (PRC, van den Brink & ter Braak 1999), which is simply a special form of redundancy analysis (RDA).

Distance-based redundancy analysis (dbRDA) can also be used to model changes in the community through time (or space) explicitly as a linear, quadratic or other polynomial traveling through the multivariate data cloud (e.g. Makarenkov & Legendre 2002), rather than treating time (or space) as an ANOVA factor (see the chapters on DISTLM and dbRDA below for more details). Larger-scale monitoring programs, which generally have many sites sampled repeatedly at many times, can also be analyzed using multivariate control charts (Anderson & Thompson 2004). This approach is designed to detect when (and where) individual sites deviate significantly from what would be expected, given natural temporal variability.

For the one-way case (such as the Ekofisk oil-field data seen in the section One-way example above), the consequences of an unbalanced design are not problematic. We can perform PERMANOVA in the usual way, with the usual partitioning of sums of squared dissimilarities. Only two consequences of unequal replication are apparent for one-way designs. First, the multiplier on the EMS for the factor of interest is no longer necessarily a whole number, as it was for the balanced case. By scrolling down the PERMANOVA results window produced in the analysis of the Ekofisk oil-field data (shown only partially in Fig. 1.10 above), we can see the multiplier for the ‘S(Di)’ component of variation in the EMS for Distance is 9.57, whereas for any of the balanced designs, the multipliers for any component in any EMS are whole numbers (e.g., see the EMS in Figs. 1.15, 1.23 or 1.29 above, all of which are balanced designs). For one-way

Unbalanced designs
designs, this does not change, however, the use of the residual MS in the denominator of the pseudo-$F$ ratio for the test of “No differences among the groups”.

The second consequence of an unbalanced design is apparent when we consider the permutations. We still randomly allocate observation units across the groups (levels of the factor), while maintaining the existing group differences in sample sizes. Each individual unit no longer has an equal chance of falling into any particular group, but instead will have a greater chance of falling into a group that has a larger sample size. However, we can still proceed easily on the basis that all possible re-arrangements of the samples by reference to the existing (albeit unbalanced) experimental design are equally likely.

The more important issues facing experimenters with unbalanced designs occur when there is more than one factor in the design. In that case, the consequences are: (i) the multipliers on individual components of variation in the EMS’s are not necessarily whole numbers and these multipliers can differ for the same component when it appears in the EMS’s of different terms in the model; (ii) the main effects of factors and the interaction terms are no longer independent of one another. The latter is perhaps the most important conceptual issue in the analysis of unbalanced, as opposed to balanced, designs. This means that, like in multiple regression (see chapter 4), the order in which we choose to fit the terms matters.

![Venn diagrams showing the difference between (a) a balanced and (b) an unbalanced two-way crossed design.](image)

Begin by considering a two-way crossed ANOVA design, with factors A, B and their interaction AxB. If the design is balanced, then the individual amounts of variation in the response data cloud explained by each of the terms in the model are completely independent of one another. This can be visualised using a Venn diagram (Fig. 1.40), where the total variation in the system ($SS_T$) is represented by a large circle and the residual variation, $SS_{Res}$, is the area left over after removing all of the portions explained by the model. For a balanced design, the individual terms in the model explain separate independent portions of the total variation (Fig. 1.40a), whereas for an unbalanced design, there will be some overlap among the terms regarding the individual portions of variation that they explain (Fig. 1.40b).

**Types of SS**

When the design is unbalanced, there will be a number of different ways to do the partitioning, which will depend to some extent on our hypotheses and how we wish to treat the potential overlap among the terms. The different ways of doing the partitioning are called “Types” of sums of squares. More particularly, there are (at least) four types, known (perhaps unhelpfully) as Type I, II, III and IV. This terminology was initially coined by the developers of the SAS computer program (e.g., SAS Institute 1999), and is now in common usage. All of these types of SS produce identical results for balanced designs. Furthermore, Types II, III and IV will be identical for models with no interactions and Types III and IV will be identical if all cells are filled (i.e. if all cells have $n \geq 1$). Searle (1987) provides an excellent text regarding models and hypotheses for unbalanced designs, including a comparison of the types of SS, which are briefly described below:
1. PERMANOVA

- **Type I SS.** This approach may be described as *sequential*. Here, each term is fitted after taking into account (i.e., conditioning upon or treating as covariates) all previous terms in the model. The order of the terms listed in the design file for the analysis therefore matters. For example, in the two-way crossed design, if the order of the terms listed in the analysis were A, B and A×B, then the SS would calculate the SS for A (ignoring other terms), the SS for B (given that A is already in the model) and then the SS for A×B (given that A and B are both already in the model). This is shown diagrammatically in Fig. 1.41a. You would get a different partitioning for this same design, however, if you chose to fit factor B first and then A given B (Fig. 1.41b). Another thing to note about Type I SS is that the sum of the individual SS will add up to the total SS. The sequential analysis offered by the Type I approach may be appropriate for fully nested hierarchical models, for which there exists a natural ordering of the terms. In other cases, Type I SS may be used to explore the amount of overlap in the explained variability among terms, by changing the order of the terms in the analysis and seeing how this affects the results.

- **Type II SS.** This approach might be described as a *conditional* analysis. The Type II SS for a given term is defined as the reduction in the residual SS due to adding the term *after* all other terms have been included in the model *except* any terms that contain the effect being tested. In other words, main effects are not conditioned upon interaction terms that involve them. In the two-way crossed design, this amounts to fitting A given B, B given A and A×B given both A and B (Fig. 1.41c). Note that, given the explicit definition, the order in which the individual terms are fitted will not matter here. However, the individual SS in the analysis will not necessarily add up to the total SS. There will potentially be some “bits missing” as a consequence of this partitioning. (In the two-way case, the “bit missing” is the overlap of factors A and B in their explained variation).

- **Type III SS.** This approach might be described as a fully *partial* analysis. Every term in the model is fitted only after taking into account *all* other terms in the full model. Thus, for our example, we can see this amounts to fitting A given both B and A×B, B given both A and A×B, and A×B given both A and B (Fig. 1.41c). Like for Type II SS, the nature of the definition here ensures that the order in which terms are fit will not matter. However, also like Type II SS, the sum of the individual SS will not add up to the total SS, and there will be more “bits missing” (ignored regions of overlap) for the Type III case. However, complete *orthogonality* (independence) of all of the hypotheses is ensured using this method.

---

**Fig. 1.41.** Schematic Venn diagrams demonstrating the conceptual differences in Types of SS for a two-way crossed unbalanced design.
PERMANOVA

Type IV SS. This approach was developed by the SAS Institute (e.g., 1999) to deal with cases of a particular kind of imbalance known as “some cells empty”, in which data from some of the cells (combinations of treatments) are actually missing entirely (i.e. \( n = 0 \) for those cells). These situations can be contrasted with situations known as “all cells filled”, which may have either unequal or equal (balanced designs) replication per cell. According to Searle (1987), Type IV SS are for testing hypotheses determined rather arbitrarily by the SAS GLM routine, which depends not only on which cells have data in them, but also on the order in which levels of the factors happen to have been listed. As such, Type IV SS is not generally recommended and is not discussed further here. The situation of “some cells empty” is quite adequately dealt with using Type III SS.

The PERMANOVA dialog box offers the user the option of using Type I, II or III SS. The default in PERMANOVA is to use Type III SS. This is primarily because most editors of journals (at least, most ecological journals) have now come to expect Type III SS to be used for unbalanced designs, simply because these will tend to be the most conservative of the three. However, there is no particular reason not to use the other types, especially if one of these is better suited to particular hypotheses of interest. For example, as already noted, a sequential analysis (Type I SS) would be quite sensible to use for an hierarchical nested design. Indeed, many statisticians would consider Type I SS to be the most sensible general approach, as no components of variation are left out (i.e. there are no “bits missing”). Type I SS also allows clarification of the relative sizes of overlapping regions, when terms are fitted in different orders. In most cases, provided the degree of imbalance in the design is modest (due, for example, to just a few missing observations here and there), the overall conclusions of the study will be little affected by this choice.

A case in point is provided by an analysis of bird assemblages from Borneo, Indonesia in response to a two-way unbalanced design, as described by Cleary et al. (2005). A total of \( N = 37 \) sites were sampled within the Kayu Mas logging concession, close to Sangai, Central Kalimantan. The sites were cross-classified according to two factors: Logging (fixed with \( a = 3 \) levels: unlogged primary forest, forest logged in 1993/94 and forest logged in 1989/90) and Slope (fixed with \( b = 3 \) levels: lower, middle and upper). There were different numbers of sites \( (n) \) within each of the \( a \times b = 3 \times 3 = 9 \) cells in the design, as shown in Fig. 1.42. Within each site, spot-mapping (using calls and visual observations) was used to sample birds along each of two parallel 300 m linear transects, 50 m apart at each site. There were \( p = 177 \) bird species recorded in all and the data are located in the file born.pri, located in the ‘BorneoBirds’ folder of the ‘Examples add-on’ directory.

An MDS plot of the bird communities, on the basis of log\((x+1)\)-transformed data and Bray-Curtis similarities (Cleary et al. 2005) shows a clear effect of logging, and suggests some effects of slope as well, although these are less clear (Fig. 1.43). For illustration, four different analyses of the data were done using PERMANOVA (Fig. 1.44). First an analysis using Type I SS was done, fitting the factor of “Logging” first. Next, an analysis using Type I SS was done again, but this time fitting the factor of “Slope” first. Analyses were then done using each of Type II and Type III SS, in turn. Note that the SS for either factor using Type II SS corresponds to what is obtained if that factor is
fitted second in a sequential (Type I) analysis. The Type III SS are different from all others, except for the interaction term, which in all cases was conditioned upon both of the main effects. Note also that the multipliers on individual components in each EMS differ for the different types of SS as well. This, in turn, means that the estimates of components of variation will also differ (Fig. 1.44).

**Fig. 1.43.** MDS ordination of bird assemblages from Borneo in all combinations of logging (P = primary forest, L89 = logged in 1989/90, L93 = logged in 1993/94) and slope (L = lower, M = middle and U = upper).

In summary, the analysis of an unbalanced design using different Types of SS affects the values of (i) the SS themselves (and thus the values of MS and pseudo-\(F\)) for individual terms in the model; (ii) the EMS for each term and multipliers of individual components of variation; (iii) the estimates of the sizes of components of variation. Despite all of this, for the present example, the same general conclusions would be obtained, regardless of which Type of SS we had decided to use (Fig. 1.44). There are significant differences among bird assemblages in forests having different logging histories and the slope of the site also has a significant effect. These factors did not interact with one another and logging effects were much larger than slope effects (see Fig. 1.43 and also the estimated components of variation). The extent to which different types of SS will give comparable results will depend on just how unbalanced the design is – greater imbalances will generally lead to greater overlapping regions and thus potentially greater discrepancies. Perhaps the most important point is to recognise how different choices for the type of SS in unbalanced designs correspond conceptually to different underlying hypotheses (see Table 1.5, Fig. 1.41 and Searle 1987).

**Table 1.5.** Tests done using different types of SS in a two-way crossed unbalanced design (cf. Fig. 1.41). The vertical line is to be read as “given”, thus “A | B” should be read as “factor A given factor B”. A comma should be read as “and.”

<table>
<thead>
<tr>
<th>Type of SS</th>
<th>(SS_A)</th>
<th>(SS_B)</th>
<th>(SS_{A \times B})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I SS (fitting A first)</td>
<td>A</td>
<td>B</td>
<td>A×B</td>
</tr>
<tr>
<td>Type I SS (fitting B first)</td>
<td>A</td>
<td>B</td>
<td>A×B</td>
</tr>
<tr>
<td>Type II SS</td>
<td>A</td>
<td>B</td>
<td>A×B</td>
</tr>
<tr>
<td>Type III SS</td>
<td>A</td>
<td>B, A×B</td>
<td>B</td>
</tr>
</tbody>
</table>
A topic that is related (perhaps surprisingly) to the topic of unbalanced designs is the analysis of covariance, or ANCOVA. There are some situations where the experimenter, faced with the analysis of a set of data in response to an ANOVA-type of experimental design, would like to take into account one or more quantitative variables or covariates. The essential idea here is that the response variable(s) may be known already to have a relationship with (or to be affected by) some (more or less continuous) quantitative variable. What is of interest then is to analyse the response data cloud given this known existing relationship. That is, one may wish to perform the ANOVA partitioning only after including (i.e., fitting, conditioning upon or taking into account) the covariate(s) in the model. Unfortunately, even if the design is completely balanced, so that terms are independent of one another, this is almost certainly not the case if a covariate is added to the...
1. PERMANOVA

model. Thus, if one or more covariates are included in an ANOVA design (to yield what is commonly referred to as an ANCOVA model), the SS for individual terms in the model are not independent of one another. This means that the Type of SS must be chosen carefully for these situations, just as for an unbalanced case.

Generally, the covariate is to be fit first, with the design factors to be considered given the covariate, so there is a logical sequential order of terms. Thus, Type I SS usually makes the most sense here. However, using Type I SS does mean that the order of the fit of the design factors relative to each other will also matter, so this should be kept in mind. Thus, if Type I SS are to be used, the experimenter might also decide to re-run the analysis, swapping the order of factors in the ANOVA part of the model in order to check whether this affects any essential interpretations of the results. The order in which the individual terms are fit by PERMANOVA can be changed either by changing the order of the rows in the design file or by clicking on the ‘Terms...’ button in the PERMANOVA dialog to yield the ‘Ordered Selection’ window (e.g., Fig. 1.32), then using the up and down arrows to move the relative positions of individual terms that will be included in the sequential fit.

An example of a design which might include a covariate in the model is provided by the holdfast invertebrate dataset, seen in the section Nested design above. It is known that the community structure of organisms inhabiting holdfasts is affected by the volume of the holdfast habitat itself (e.g., Smith et al. 1996). This stands to reason, given the well-known ecological phenomenon of the species-area relationship (e.g., Arrhenius 1921, Connor & McCoy 1979). Larger holdfasts have not only a larger colonisable area (which might alone be dealt with adequately through a standardisation by total sample abundance), but they also have greater complexity, usually with a greater number of interstices and root-like structures (called haptera) as well. The volume of each holdfast in the study by Anderson et al. (2005a) was measured using water displacement, and should really be included as a covariate in the analysis.

![Graph1](image1.png)

**Fig. 1.45.** Draftsman plot of environmental variables for kelp holdfasts.

Open up the file hold.pri in the folder ‘HoldNZ’ of the ‘Examples add-on’ directory. Here, we shall analyse only those species (or taxa) that occurred as counts of abundances. We will not include the encrusting organisms that were recorded only as an ordinal measure from 0-3. Choose **Select > Variables > Indicator levels > Indicator name: Ordinal > Levels… > (Available Y) & (Include N).** Choose **Tools > Duplicate** and re-name the data file produced as hold.abund. We shall base the analysis on a dissimilarity measure that is a modification of the Gower measure, recently described by Anderson et al. (2006). This measure is interpretable as the average order-of-magnitude

44 The reason non-independence is introduced becomes clearer perhaps when we consider that if the covariate is fitted first and different ranges of the continuous covariate occur in different cells, then it cannot be independent of existing factors.
difference in abundance per species, where each change in order of magnitude (i.e., 1, 10, 100, 1000 on a log₁₀ scale) is given the same weight as a change in composition from 0 to 1 (see Anderson et al. 2006 for more details). Choose Analyse > Resemblance > (Analyse between •Samples) & (Measure •More (tab)), click on the tab labeled ‘More’ and choose (•Others: Modified Gower > Modified Gower log base: 10). Save the current workspace as hold_cov.pwk.

Environmental data, including depth (in m), density of kelp plants where the holdfast was collected (per m²) and a measure of volume for each holdfast (in ml), are located in the file holdenv.pri, also located in the ‘HoldNZ’ folder. Open up this file within the hold_cov workspace just created and choose Analyse > Draftsman plot to examine the distributions of these environmental variables, focusing on the variable of ‘Volume’ in particular (Fig. 1.45). The distributions of environmental or other quantitative variables to be included as covariates in PERMANOVA should be investigated before proceeding. There are two important reasons for doing this:

- The model fit by PERMANOVA is linear with respect to the covariate. That is, the relationship between the covariate and the multivariate community structure as represented in the space defined by the dissimilarity measure chosen is linear. Importantly, the relationship between the covariate and the original species (or other response) variables is emphatically not linear unless Euclidean distance is used as the basis of the analysis. So, the experimenter should look (as far as possible) for outliers and other oddities in the multivariate space of the dissimilarity measure chosen (i.e., using ordination) and should also examine the distribution of the covariate as part of a general diagnostic process before proceeding.

- Permutation of raw data will have inflated type I error if there are outliers in the covariates (Kennedy & Cade 1996, Anderson & Legendre 1999, Anderson & Robinson 2001). Thus, permutation of raw data is not allowed as an option if covariates are included in the PERMANOVA model.

If the distribution of the covariate is skewed or bimodal, then the user may wish either to transform the variable, or to split the dataset according to any clear modalities observed. See chapter 10 of Clarke & Gorley (2006) regarding the diagnostic process for interpreting and using draftsman plots for environmental variables. The main point is that covariates used in PERMANOVA should show approximately symmetric distributions that are roughly normal, with no extreme outliers. As pointed out by Clarke & Gorley (2006), it is not necessary to agonise over this issue! Normality is by no means an assumption of the analysis, but outliers can have a strong influence on results. A draftsman plot of the holdfast environmental variables indicates that the distribution of volumes is fairly even, with no obvious skewness or outliers (Fig. 1.45), so no transformation is necessary.

An MDS plot of the holdfast fauna with volume superimposed (as bubbles) shows a clear relationship between volume and community structure (Fig. 1.46); change in community structure
from left to right across the diagram is associated with increasing volume of the holdfast. Thus, it
would make sense to examine the variability in community structure at different spatial scales over
and above this existing relationship with volume.

Fig. 1.47. PERMANOVA dialog for the analysis of holdfast fauna in response to the fully nested design and
including volume as a covariate.

To proceed with the analysis, start by selecting the holdenv worksheet, highlight the single variable
of Volume, choose Select >Highlighted followed by Tools >Duplicate, then re-name the resulting
worksheet vol. It is important that the variable(s) to be used as covariate(s) in the model be
contained in a single separate datasheet. It is also important that the labels of the samples in this
data sheet match those that are contained in the resemblance matrix, although the particular order
of the samples need not be the same in the two files. (PERMANOVA, like other routines in
PRIMER, uses a label matching procedure to ensure correct analysis). Next, go to the Modified
Gower resemblance matrix calculated from the hold.abund data sheet and create a PERMANOVA
design file with three factors: Location, Site and Area, according to the fully nested design outlined
in the section Nested design above (e.g., Fig. 1.29). Re-name this design file Nested design. Click
on the dissimilarity matrix again and choose PERMANOVA+ > PERMANOVA > (Design
worksheet: Nested design) & (Covariable worksheet: vol) & (Sums of Squares •Type I SS
(sequential)) & (Num. permutations: 9999) & (Permutation method •Permutation of residuals
under a reduced model) & (✓Use short names), as shown in Fig. 1.47.

The results of this analysis show that there is a strong and significant effect of the covariate, i.e., a
significant relationship between holdfast volume and community structure as measured by the
Modified Gower dissimilarity measure (Fig. 1.48). This is not surprising, given the pattern seen in
the MDS plot (Fig. 1.46). Nevertheless, even given the variation among holdfast communities due
to volume, significant variability is still detected among the assemblages at each spatial scale in the
design: among locations (100’s of km’s), sites (100’s of m’s) and areas (10’s of m’s) (Fig. 1.48).

It is possible, in such an analysis, to include interactions between the covariate and each of the
other terms in the experimental design. A significant interaction between a factor (such as
1. Permanova

Locations) and a covariate (such as volume) indicates that the nature of the relationship between the covariate and the multivariate responses differs within different levels of the factor (i.e., that the relationship between volume and holdfast community structure differs at different locations, and so on for other factors in the model). In the case of univariate data with a single factor and one covariate, as analysed in a traditional ANCOVA model, the test of the interaction term between the covariate and the factor is also known as a test for homogeneity of slopes. Sometimes ANCOVA is presented as a model without the interaction term and in these cases it is stated that homogeneity of slopes is an assumption of the analysis. However, if one includes the interaction term in the model, then clearly this type of homogeneity is no longer an assumption, as the model explicitly allows for different slopes for different levels of the factor. See Winer et al. (1991) or Quinn & Keough (2002) for more complete discussions and examples of traditional ANCOVA models.

In PERMANOVA, tick the box labeled ‘Include interactions’ in the dialog under the specification of the covariable worksheet in order to include these terms in the model. Note that, after ticking this box, you can also still click on the ‘Terms…’ button and change the order in which the terms are fitted and/or exclude particular terms from the model if you wish, whether these be interactions with the covariate or some other terms. An analysis of the holdfast data where interaction terms are included suggests there is actually no compelling reason to include any of these interactions in the model in this particular case ($P > 0.18$ for all interactions with the covariate, Fig. 1.48).

We re-iterate that the most important thing to remember about analyses involving covariates is that the individual terms are not independent of one another, so the Type of SS chosen for the analysis will affect the results, just as it does for unbalanced designs. What is more, the underlying complexity of this particular analysis (even without including the interaction terms), which is a mixed model having both random factors and a (fixed) covariate, can be appreciated by considering the EMS’s and the rather impressive gymnastics the program must go through in order to produce tests of individual terms using pseudo-$F$ (Fig. 1.49). For more details on how these pseudo-$F$ ratios were constructed, see the next section, Linear combinations of MS.

---

**Fig. 1.48.** PERMANOVA results of the analysis of holdfast fauna, including volume as a covariate, either (a) without interactions or (b) including interactions.

---

45 This complexity was, unfortunately, not recognised by the authors of the original paper (Anderson et al. 2005a). The analyses they presented took a “naïve” view and conditioned on the covariate without considering the effects this would have on the EMS’s and pseudo-$F$ ratios for the other factors in the model. Making mistakes is, however, part of doing science and we would be nowhere if we did not make mistakes, acknowledge our errors and learn from them!
Several aspects of the above analysis demonstrate its affinity with unbalanced designs. Note that: (i) the multipliers for components of variation in each EMS are not whole numbers; and (ii) the multipliers for a given component of variation are not the same in different EMS’s. As a consequence of these two things, for many of the terms in the model, there is no other single term that, alone, can provide a MS which can act as a denominator to yield a correct pseudo-$F$ ratio. What this generally means for such cases is that some linear combination of mean squares must be used in order to construct a test of the given null hypothesis of interest. Depending on how these linear combinations are constructed, this can also mean that even the degrees of freedom used for the tests are not whole numbers (Fig. 1.49)!

The good news is that the PERMANOVA routine is actually equal to this (rather horrendous) task; (i) it determines the correct EMS’s for every term; (ii) it calculates the correct linear combinations of mean squares required to construct pseudo-$F$ ratios to test each hypothesis; and (iii) it determines the correct distributions of each pseudo-$F$ ratio under each relevant null hypothesis using permutations, producing accurate $P$-values. The routine is also not bothered by non-integer degrees of freedom. The exchangeable units for a given test (e.g., Anderson & ter Braak 2003) are chosen using what might be called a “highest-order term” approach\(^46\). That is, consider the linear combination of mean squares required for a given test. Of the terms giving rise to those mean squares, the term in the denominator which is of highest order (excluding continuous covariates and the residual) is used to identify exchangeable units for that particular test. For example, the terms whose mean squares are included in the denominator for the pseudo-$F$ ratio test of Location in Fig. 1.49 are: $\text{Si(Lo)}$, $\text{Ar(Si(Lo))}$ and Res. The highest-order term (excluding the residual) is $\text{Ar(Si(Lo))}$. Therefore, all replicates within an area will be kept together as a group under

\(^{46}\) The order of a term is defined here as follows: a main effect (e.g., $A$, $B$, …) is of first order, a two-way interaction (e.g., $A\times B$) is of second order, a three-way interaction (e.g., $A\times B\times C$) is of third order, and so on. Also, a nested term, such as $B(A)$, is of second order, while $C(A\times B)$ or $C(B(A))$ would both be of third order, etc.
permutation and the 16 different areas will be the units permuted for the test of Locations in this design. Importantly, after each permutation, the full pseudo-\(F\) ratio is constructed according to the requirements of the numerator and denominator, even if either one or both of these are linear combinations of mean squares.

The need to construct linear combinations of mean squares happens not just in highly complex models with covariates, as described above. It is also (much more commonly) required for certain terms even in many balanced designs, especially multi-way designs involving more than one random factor\(^47\). For example, consider a study of temperate rocky reef fish assemblages as described by Anderson & Millar (2004). The study consisted of surveys of fish biodiversity, where abundances of fish species were counted in 10 transects (25 m × 5 m) sampled by SCUBA divers from each of 4 sites in each of 2 habitats at each of 4 locations along the northeast coast of New Zealand. These surveys have been done each year in the austral summer. The data provided in the file fishNZ.pri (located in the folder ‘FishNZ’ in the ‘Examples add-on’ directory) are sums across the 10 transects within each site for each of \(p = 58\) fish species from surveys conducted in each of two years: 2004 and 2005\(^48\). The experimental design here is:

- **Factor A**: Year (random with \(a = 2\) levels: 4 = 2004 and 5 = 2005).
- **Factor B**: Location (random with \(b = 4\) levels: B = Berghan Point, H = Home Point, L = Leigh and A = Hahei).
- **Factor C**: Habitat (fixed with \(c = 2\) levels: b = urchin-grazed ‘barrens’ and \(k = \) kelp forest).

It is of interest to test the null hypothesis of no difference between the two habitats in fish assemblages. It is also (secondarily) of interest to test and quantify the variability among years and among locations in fish community structure. An analysis of the data according to the above experimental design using PERMANOVA has been done on the basis of the scaled binomial deviance dissimilarity measure (see Anderson & Millar 2004 for a description of this measure), yielding the results shown in Fig. 1.50.

What is of immediate interest to us here is the test for the main effect of ‘Habitat’ or ‘Ha’ in the output\(^49\). The EMS for this factor is:

\[
E(MS_{Ha}) = 1*V(Res) + 4*V(YexLoxHa) + 8*V(LoxHa) + 16*V(YexHa) + 32*S(Ha)
\]

Now, to construct pseudo-\(F\), we need to find a denominator mean square whose expectation is equal to the above when the null hypothesis that \(S(Ha) = 0\) is true. That is, we need a denominator whose expectation is:

\[
E(denom) = E(MS_{Ha}) - 32*S(Ha)
\]

\[
= 1*V(Res) + 4*V(YexLoxHa) + 8*V(LoxHa) + 16*V(YexHa)
\]

However, there is clearly no single term that, alone, can perform this duty, because we require both the \(LoxHa\) and the \(YeXHa\) components of variation to appear here. We can see, however, that the term we seek can be obtained by constructing a linear combination of mean squares:

\[
E(denom) = E(MS_{LoxHa}) + E(MS_{YexHa}) - E(MS_{YexLoxHa})
\]

It is desirable, however, not to include mean squares negatively, as this could generate negative pseudo-\(F\) ratios, which are not really sensible (e.g., Searle et al. 1992). Thus, re-arranging the above (so that all mean square terms appear positively), we have:

\[
E(denom) + E(MS_{YexLoxHa}) = E(MS_{LoxHa}) + E(MS_{YeXHa})
\]

or

---

\(^{47}\) The need to use linear combinations of mean squares to construct appropriate pseudo-\(F\) statistics will also arise much more commonly (or for more of the terms in the model) if the constraint that fixed effects should sum to zero across levels of random factors in mixed interactions is not applied (i.e., if one chooses to remove the \(\checkmark\) in front of the option ‘Fixed effects sum to zero’ in the PERMANOVA dialog).

\(^{48}\) Note that these are not the same years as those analysed by Anderson & Millar (2004), but are data from more recent surveys.

\(^{49}\) See the above section Inference space and power regarding the logic and hypotheses underlying tests of fixed main effects even in the presence of potentially non-zero interactions with random factors.
1. PERMANOVA

\[ E(\text{MS}_{\text{Ha}}) + E(\text{MS}_{\text{YexLoxHa}}) = E(\text{MS}_{\text{LoxHa}}) + E(\text{MS}_{\text{YeXHa}}) + 32*\text{S(Ha)} \]

Accordingly, if we wish to construct a pseudo-\(F\) ratio where the numerator and denominator will have the same expectation if \(H_0: \text{S(Ha)} = 0\) is true, and which gets large with increases in the size of \(\text{S(Ha)}\) alone then we can use:

\[ F_{\text{Ha}} = \frac{\text{MS}_{\text{Ha}} + \text{MS}_{\text{YexLoxHa}}}{\text{MS}_{\text{LoxHa}} + \text{MS}_{\text{YeXHa}}} \]

This is precisely the pseudo-\(F\) ratio constructed by the PERMANOVA program, as stated in the results in the line for ‘Ha’ under the heading ‘Construction of Pseudo-\(F\) ratio(s) from mean squares’ (Fig. 1.50).

![Fig. 1.50. Analysis of New Zealand temperate reef fish according to the three-way mixed model. Note the linear combination of mean squares needed to test the habitat main effect: ‘Ha’.

The exchangeable units for this test, using the “highest-order term” approach, are obtained by considering all of the terms involved in the construction of pseudo-\(F\) (apart from the term being tested, which is ‘Ha’ here) and determining the one with the highest order. The terms involved are: \(\text{YxLoxHa, YeXHa and LoxHa}\). The term with the highest order is \(\text{YeXLo} \times \text{Ha}\), so the exchangeable units for this test are the \(a \times b \times c = 2 \times 4 \times 2 = 16\) cells. So, the \(n = 4\) sites occurring within each of those cells will be permuted together as a unit and the full pseudo-\(F\) will be re-constructed after each permutation in order to test the term ‘Ha’ in this case\(^{50}\).

The attentive user will notice the large \(P\)-values (> 0.25) associated with a couple of the terms in the model, namely the ‘Ye\(\times\)Ha’ and ‘Lo\(\times\)Ha’ interaction terms. If possible, it is desirable to pool

\[^{50}\text{Although some preliminary simulation work has indicated that this “highest-order term” approach works well in trial cases in terms of maintaining rates of type I error at chosen significance levels, a more complete study of this rather complex issue of appropriate exchangeable units for \(F\) ratios involving linear combinations of mean squares would be welcome.}\]
PERMANOVA of the New Zealand fish assemblages after pooling of terms, showing the dialog used at the second step in the pooling procedure, when the term ‘Lo×Ha’ was pooled along with ‘Ye×Ha’, which had previously been pooled at step one.

In traditional univariate ANOVA, the potential use of pooling for these kinds of situations was very important. This is because the construction and subsequent testing of traditional $F$ ratios using linear combinations of mean squares in the numerator and denominator is fraught with difficulties, primarily because ratios of linear combinations of mean squares (called “quasi” $F$ ratios by Quinn & Keough 2002, see also Blackwell et al. 1991) no longer have (known) $F$ distributions under a true null hypothesis, even when the usual assumptions of normality, homogeneity, etc. are fulfilled (see also Searle et al. 1992). Complicated approximations have therefore been suggested in order to obtain $P$-values for these cases (e.g., Satterthwaite 1946, Gaylor & Hopper 1969), in the event that pooling was not possible.

One of the most important advantages of the PERMANOVA routine is that it uses permutation tests to obtain $P$-values. Thus, as long as: (i) the test statistic is constructed correctly in the sense that it isolates the term of interest under the null hypothesis; and (ii) the permutations are done so as to create alternative realisations under a true null hypothesis by permuting appropriate exchangeable units, then the calculations of the $P$-values are correct and can be used for valid inference. This is true whether or not the corresponding traditional univariate test would be able to be done at all using more traditional theoretical approaches. The more general unified approach of...
1. PERMANOVA

the new PERMANOVA software caters even for these situations (such as the need for linear combinations of mean squares), where the traditional tests would be very difficult (sometimes even impossible) to formulate. Also, of course, PERMANOVA can be implemented on distance or dissimilarity matrices which have been calculated from either univariate or multivariate data.

Although a previous section has been devoted to the analysis of unbalanced designs, there are some special cases of designs having missing cells which deserve extra attention. Such designs are commonly referred to as asymmetrical designs, and consist essentially of there being different numbers of levels of a nested factor within each different level of an upper-level factor. Important examples include the asymmetrical designs that can occur in studies of environmental impact. Here, there might only be a single site that is impacted, whereas there might be multiple control (or unimpacted) sites (e.g., Underwood 1992, 1994, Glasby 1997). The reason for asymmetrical designs arising frequently in the analysis of environmental impacts is that it is generally highly unlikely that an impact site of a particular type (e.g., an oil spill, a sewage outfall, the building of a particular development, etc.) will be replicated, whereas there is often no reason not to include multiple replicate control sites (at a given spatial scale) against which changes at the (purportedly) impacted site might be measured (e.g., Underwood 1992, 1994). On the face of it, the experimenter might consider that such a design presents a severe case of imbalance, where not all cells are filled. In actual fact, this is not really the case, because it is only the number of levels of the nested factor that are unequal, and actually all of the terms in the partitioning will be independent of one another, just as they would be in a balanced design.

An example of an asymmetrical design is provided by a study of subtidal molluscan assemblages in response to a sewage outfall in the Mediterranean (Terlizzi et al. 2005). The study area is located along the south-western coast of Apulia (Ionian Sea, south-east Italy). Sampling was undertaken in November 2002 at the outfall location and at two control or reference locations. Control locations were chosen at random from a set of eight possible locations separated by at least 2.5 km and providing comparable environmental conditions to those occurring at the outfall (in terms of slope, wave exposure and type of substrate). They were also chosen to be located on either side of the outfall, to avoid spatial pseudo-replication. At each of the three locations, three sites, separated by 80 - 100 m were randomly chosen. At each site, assemblages were sampled at a depth of 3 - 4 m on sloping rocky surfaces and \( n = 9 \) random replicates were collected (each replicate consisted of scrapings from an area measuring 20 cm \( \times \) 20 cm), yielding a total of \( N = 81 \) samples. The experimental design is:

Factor A: Impact versus Control (‘IvC’, fixed with \( a = 2 \) levels: I = impact and C = control).

Factor B: Location (‘Loc’, random, nested in IvC with \( b = 2 \) levels nested in C and 1 level nested in I, labeled as numbers 1, 2, 3).

Factor C: Site (random, nested in Loc(IvC) with \( c = 3 \) levels, labeled as numbers 1-9).

A schematic diagram (Fig. 1.52) helps to clarify this design, and why it is considered asymmetrical.

![Fig. 1.52. Schematic diagram of the asymmetrical design for Mediterranean molluscs.](image-url)
The data for this design are located in the file medmoll.pri in the ‘MedMoll’ folder of the ‘Examples add-on’ directory. As in Terlizzi et al. (2005), to visualise patterns among sites, we may obtain an MDS plot of the averages of samples at the site level. Go to the worksheet containing the raw data and choose **Tools > Average > (Samples ● Averages for factor: Site)**, then **Analyse > Resemblance > (Analyse between: ● Samples) & (Measure ● Bray-Curtis)**, followed by **Analyse > MDS**.

**Fig. 1.53.** MDS of site averages for Mediterranean molluscs, where I = impact and C = control locations.

There is apparent separation between the (averaged) assemblages at sites from the impact location compared to the controls on the MDS plot (Fig. 1.53), but to test this, we shall proceed with a formal PERMANOVA analysis. The difference, if any, between impact and controls must be compared with the estimated variation among control locations; Next, calculate a Bray-Curtis resemblance matrix directly from the original medmoll.pri raw data sheet (i.e., not the averaged data), then create a PERMANOVA design file according to the above experimental design, rename the design file Asymmetric and proceed to run the analysis by choosing the following in the PERMANOVA dialog: (Design worksheet: Asymmetric) & (Test: ● Main test) & (Sums of Squares ● Type III (partial)) & (Num. permutations: 9999) & (Permutation method ● Permutation of residuals under a reduced model) & (✓ Do Monte Carlo tests) & (✓ Fixed effects sum to zero). For clarity in viewing these results, un-check (i.e. remove the ✓ from) the option to ‘Use short names’.

The results indicate that there is significant variability among sites, but variation among control locations is not detected over and above this site-level variability (i.e. the term ‘Loc(IvC)’ is not statistically significant \( P > 0.12 \), Fig. 1.54). There is also apparently a significant difference in the structure of molluscan assemblages at the impact location compared to the control locations (\( P(\text{MC}) = 0.036 \)). Note that, in the absence of any replication of outfall (impact) locations, the only basis upon which location-level variability may be measured is among the (in this case only two) control locations. We should refrain from going overboard in the extent of our inferences here – it is inappropriate to place strong importance on an approximate MC \( P \)-value which relies on asymptotic theory (i.e. its accuracy gets better as sample size increases) yet was obtained using only two control locations and hence only 1 df in the denominator. Furthermore, of course, in the absence of any data from before the outfall was built, it is not possible to infer that the difference between the impact and the controls detected here was necessarily *caused* by the sewage outfall. It is also impossible to know whether this difference is something that has persisted or will persist in time. What we can say is that the molluscan assemblages at the outfall at the time the data were collected were indeed distinct from those found at the two control locations in the area sampled at that time, as observed in the MDS plot (Fig. 1.53).

Asymmetrical designs such as this one have often previously been analysed and presented by partitioning overall location effects into two additive pieces: (i) the SS due to the contrast of the impact vs the controls and (ii) the SS due to the variability among the controls (e.g., Underwood
1. PERMANOVA

1994, Glasby 1997). The reason for this has largely been because of the need for experimenters to utilise software to analyse these designs which did not allow for different numbers of levels of nested factors in the model. PERMANOVA, however, allows the direct analysis of each of the relevant terms of an asymmetrical design such as this, without having to run more than one analysis, and without any other special manipulations or calculations.

It is not necessary for the user to specify contrasts in PERMANOVA in order to analyse an asymmetrical design. It is essential to recognise the difference between the design outlined above, which is the correct one, and the following design, which is not:

Factor A: Locations (fixed or random?) with \( a = 3 \) levels and special interest in the contrast of level 3 (impact) versus levels 1 and 2 (controls),

Factor B: Sites (random, \( b = 3 \) levels, nested in Locations)

Be warned! If an asymmetrical design is analysed using contrasts in PERMANOVA, then although the SS of the partitioning will be correct, the \( F \) ratios and \( P \)-values for some of the terms will almost certainly be incorrect! Note that the correct denominator MS for the test of ‘IvC’ must be at the right spatial scale, i.e. at the scale of locations, even though, in the present design (with only one impact location), our only measure of location-level variability comes from the control locations. Recall that a contrast, as a one-degree-of-freedom component partitioned from some main effect, will effectively use the same denominator as that used by the main effect (e.g., see the section Contrasts above), which is not the logical choice in the present context. A clear hint to the problem underlying this approach is apparent as soon as we try to decide whether the ‘Locations’ factor should be fixed or random. It cannot be random, because the contrast of ‘IvC’ is clearly a fixed contrast of two states we are interested in. However, neither can it be fixed, because the control locations were chosen randomly. They are intended to represent a population of possible control locations and their individual levels are not of any interest in and of themselves. Once again, the rationale for analysing asymmetrical designs using contrasts in the past (and
subsequently constructing the correct $F$ tests by hand, see for example, Glasby 1997) was because software was not widely available to provide a direct analysis of the true design.

Although an asymmetrical design might appear to be unbalanced (and it is, in the sense that the amount of information, or number of levels, used to measure variability at the scale of the nested factor is different within different levels of the upper-level factor), it does not suffer from the issue of non-independence which was described in the section on Unbalanced designs above. In fact, all of the terms in an asymmetrical design such as this are completely orthogonal (independent) of one another. So, it does not matter which Type of SS the user chooses, nor in which order the terms are fitted — the same results will be obtained. (This is easily verified by choosing to re-run the above analysis using, for example, Type I SS instead.)

Some further comments are appropriate here regarding experimental designs to detect environmental impact (Green 1979, Underwood 1991, 1992, 1994). These designs generally include measurements of a response variable of interest before and after a potential impact from one or more control site(s) and from the purportedly impacted site(s). These are referred to as “BACI” designs51 as an acronym for the two levels in each of the two major factors of the design: the temporal factor (“before”/“after”) and the spatial factor (“control”/“impact”). Importantly, a significant interaction between these two factors would (potentially) lead to inferences regarding significant impact, so the ability formally to test the interaction term(s) in such models is paramount here. With regard to extending non-parametric multivariate hypothesis-testing methods to handle such designs, Clarke (1993) conceded: “This would appear to defy development within the similarity-based framework… and must be accepted as a limitation of the current methodology, though there is clearly scope for further study here” (p. 138).

Indeed, further study led to the development of PERMANOVA, which allows (under slightly less general conditions, e.g., the approach is no longer fully non-parametric) tests of interaction terms for any multi-factorial model on the basis of any resemblance measure of choice, with $P$-values obtained by appropriate permutation techniques. Its new implementation as an add-on to PRIMER also now allows appropriate analyses of asymmetrical designs (i.e., in the case of there being multiple control sites, but only one impact site – see the section Asymmetrical designs above), as required. In essence, as PERMANOVA can be used to analyse multivariate responses to any ANOVA model, it therefore can be used readily for the analysis of assemblages in response to either BACI or beyond BACI experimental designs in environmental impact studies.

A further contribution of Underwood (1991, 1992, 1994) in the area of experimental designs to detect environmental impacts (in addition to proposals to extend the basic model to include multiple sites and times of sampling and thus avoid pseudo-replication) was to propose the use of two-tailed $F$-tests to detect potential impacts on variability in the response variable, and not just to detect changes in means. Recently, Terlizzi et al. (2007) have implemented these ideas for multivariate data (albeit not in the context of environmental impact, but to investigate patterns along a depth gradient). More particularly, they used bootstrapping to place confidence intervals on differences in the sizes of multivariate components of variation. Although the current PERMANOVA+ software does not implement this approach, direct measures of the components of variation for each term in the model are provided as part of the PERMANOVA output. Thus, the sizes of multivariate (or univariate) components of variation for different sub-sets of the data may be calculated directly in this way.

Another approach to addressing hypotheses concerning variability in assemblage structure is to use the PERMDISP routine (see chapter 2), which can be applied either to individual replicates within groups (or cells), or can be applied to centroids (calculated from PCO axes when non-Euclidean measures are being used, see chapter 3) to compare dispersions for higher-level factors in more complex designs.

---

51 Much to the delight of Italian ecologists!
1. PERMANOVA
PERMDISP is a routine for testing the homogeneity of multivariate dispersions on the basis of any resemblance measure. The test is a dissimilarity-based multivariate extension of Levene’s test (1960), following the ideas of van Valen (1978), O’Brien (1992) and Manly (1994), who used Euclidean distances. In essence, the test uses the ANOVA $F$ statistic to compare (among different groups) the distances from observations to their group centroid. The user has the choice of whether to perform the test on the basis of distances to centroids or distances to spatial medians (Gower 1974). The user also has the choice either to use traditional tables or to use permutation of appropriate residuals (either least-squares residuals if centroids are used or least-absolute-deviation residuals if spatial medians are used) in order to obtain $P$-values.

There are various reasons why one might wish to perform an explicit test of the null hypothesis of no differences in the within-group multivariate dispersion among groups. First, such a test provides a useful logical complement to the test for location differences that are provided by PERMANOVA. As pointed out in chapter 1, PERMANOVA, like ANOSIM, is sensitive to differences in multivariate dispersion among groups (e.g., see Fig. 1.6). In fact, the logic of the partitions of variability for all of the ANOVA designs in chapter 1 dictates that the multivariate dispersions of the residuals should be homogeneous, as should also be the dispersions of levels for random factors in nested designs and mixed models. A significant result for a given factor from PERMANOVA could signify that the groups differ in their location (Fig. 2.1b), in their dispersion (Fig. 2.1c) or some combination of the two (Fig. 2.1d). An analysis using PERMDISP focuses only on dispersion effects, so teases out this part of the null hypothesis (equal dispersions) on its own, perhaps as a prelude to a PERMANOVA analysis, analogous to a univariate test for homogeneity prior to fitting ANOVA models.

Of course, the other reason for applying a test of dispersion is because specific hypotheses of interest may demand such an analysis in its own right. There are many situations in ecology for which changes in the variability of assemblages are of direct importance. For example, increases or decreases in the multivariate dispersion of ecological data has been identified as a potentially

---

**Fig. 2.1.** Schematic diagram showing two groups of samples in a bivariate system (two dimensions) that (a) do not differ in either location or dispersion, (b) differ only in their location in multivariate space, (c) differ only in their relative dispersions and (d) differ in both their location and in their relative dispersion.

Levene (1960) proposed doing an analysis of variance (ANOVA) on the absolute values of deviations of observations from their group mean. A multivariate analogue was described by van Valen (1978) and given in Manly (1994), based on distances. Let $y_j$ be a vector of $p$ response variables for the $j$th observation in the $i$th group and let $\mathbf{y}_i$ be the centroid vector for group $i$, then the distance between each observation and its group centroid is a single value that can be denoted by $z_{ij}$. This is visualised easily in the case of two variables (two dimensions), as shown in Fig. 2.2 for the case of 2 groups. Once a value of $z$ is calculated for each observation, this is treated as a univariate variable and the test consists of doing a traditional ANOVA comparing the values of $z$ among the groups. The central idea here is that if the groups differ in their within-group dispersions, then the values of $z$ (on average) will differ among the groups. Note that if Euclidean distance is used and there is only one variable, then the PERMDISP test is equivalent to the traditional univariate Levene’s test.

An example is provided by Manly (1994) who analysed the morphometric characteristics of sparrows from data due to Bumpus (1898). It was hypothesised that stabilising selection acting on organisms should reduce the multivariate heterogeneity of their morphometric characteristics. The data consist of $p = 5$ morphometric variables (total length, alar extent, length of beak/head, length of humerus and length of keel of sternum, all in mm) from each of $N = 49$ female sparrows. Twenty-one of the sparrows died in a storm and twenty-eight survived. Of central interest is the question: are the morphologies of the sparrows that died more variable than for those that survived? That is, are sparrows with morphologies distant from the “norm” or “average” sparrow more likely to have died?

The data are located in the file spar.pri in the ‘BumpSpar’ folder of the ‘Examples add-on’ directory. Observations on the raw data (e.g., using a draftsman plot) reveal that the variables are quite well-behaved (approximately normal), as is usual with morphometric data, and that they are correlated with one another. A PCA (principal components analysis) would be a good idea for ordination (see chapter 4 of Clarke & Warwick 2001 for details). These variables are, however, on quite different measurement scales and so should be normalised before further analysis. Choose Analyse > Pre-treatment > Normalise variables. Next, for the normalised data, choose Analyse > PCA. The first two principal components account for about 83% of the total variation across the 5 variables and are shown in Fig. 2.3. There appears to be slightly greater variation in morphological characteristics (i.e. a slightly greater spread of points) for the sparrows that died compared to the
ones that survived, but the extent of such differences, if any, can be difficult to discern in a reduced-space ordination plot.

![Bumpus’ sparrow data](image)

**Fig. 2.3.** PCA ordination of Bumpus’ sparrows on the basis of 5 normalised morphometric variables.

To test the null hypothesis of no difference in dispersion between the two groups (Fig. 2.4), begin by selecting the normalised data and then choose **Analyse > Resemblance >** (Analyse between •Samples) & **(Measure •Euclidean distance).** From the resemblance matrix, choose **PERMANOVA+ > PERMDISP >** (Group factor: Group) & (Distances are to •Centroids) & (P-
values are from Tables). There is some evidence to suggest a difference in dispersion between the two groups, although the $P$-value obtained is borderline ($F = 3.87$, $P = 0.055$, Fig. 2.4).

The output provides not only the value of the $F$ ratio and its associated $P$-value, but also gives the mean and standard error for the $z$ values in each group. These values are interpretable on the scale of the original resemblance measure chosen. Thus, in the present case, the average distance-to-centroid in the Euclidean space of the normalised morphometric variables is 1.73 for the sparrows that survived and 2.23 for the sparrows that died. The mean distance-to-centroid (± 1 standard error) might be usefully plotted in the case of there being many groups to compare (e.g., Anderson 2006). The individual $z$ values obtained for each observation (from which the mean and standard errors are calculated) will also be provided in the results file if one chooses (✓Output individual deviation values) in the PERMDISP dialog box (Fig. 2.4).

Of course, in many applications that we will encounter (especially in the case of community data), the Euclidean distance may not be the most appropriate measure for the analysis. What we require is a test of homogeneity of dispersions that will allow any resemblance measure to be used. There are (at least) two potential problems to overcome in order to proceed. First, there is the issue that the centroids in the space of the dissimilarity (or similarity) measure chosen are generally not the same as the vector of arithmetic averages taken for each of the original variables (which is what we would normally calculate as being in the “centre” of the cloud in Euclidean space). Therefore, it is not appropriate to calculate (say) the Bray-Curtis dissimilarity between an observation and its group centroid when the centroid has simply been calculated as an arithmetic average. What is needed, instead, is the “centre” of the cloud of points for each group in Bray-Curtis (rather than Euclidean) space (or in whatever space has been defined by the dissimilarity measure of choice for the particular analysis).

The solution to this is to place the points into a Euclidean space in such a way so as to preserve all of the original inter-point dissimilarities. This is achieved through the use of a method called principal coordinates analysis (PCO, see chapter 3), described by Torgerson (1958) and Gower (1966). In essence, PCO generates a new set of variables (principal coordinate axes) in Euclidean space from the dissimilarity matrix (Legendre & Legendre 1998). Usually, PCO is used for ordination (see chapter 3), and in this case only the first two or three PCO axes are drawn and examined. However, the full analysis actually generates a larger number of PCO axes (usually up to $N - 1$, where $N$ is the total number of samples). All of the PCO axes taken together can be used to re-create the full set of inter-point resemblances, but in Euclidean space. More particularly (and here is what interests us for the moment), if we calculate the Euclidean distance between, say, sample 1 and sample 2, using all of the PCO axes, then this will be equal to the dissimilarity calculated between those same two sample points using the original variables. Thus, any calculation that would be appropriate in a Euclidean context (such as calculating centroids as arithmetic averages), can be achieved in a non-Euclidean (dissimilarity) space by performing the operation on PCO axes from the resemblance matrix.

This means that we can proceed with the following steps in order to obtain the correct values for $z$ in more general (non-Euclidean distance) cases: (i) calculate inter-point dissimilarities, using the resemblance measure of choice; (ii) do a PCO of this dissimilarity matrix; (iii) calculate centroids (arithmetic averages) of groups using the full set of PCO axes; and (iv) calculate the Euclidean distance from each point to its group centroid using the PCO axes. These then correspond to the dissimilarity from each point to its group centroid in the space of the chosen resemblance measure.

The PERMDISP routine indeed follows these four steps (see Anderson 2006 for more details) in

---

52 The distances obtained from the PCO axes must be calculated by first keeping the axes that correspond to the negative and the positive eigenvalues separate. These two parts are then brought together to calculate the final distance by taking the square root of the difference in two terms: the positive sum of squares and the absolute value of the negative sum of squares. This will result in a real (non-imaginary) value provided the positive portion exceeds the negative portion. See Anderson (2006) and chapter 3 of this manual for more details.
order to calculate appropriate z values for the test of homogeneity of dispersions on the basis of any dissimilarity measure.\textsuperscript{53}

The other hurdle that must be cleared is to recognise that, in line with the philosophy of all of the routines in the PERMANOVA+ add-on, we have no particular reason to assume that the distribution of the z’s will be necessarily normal. Yet, to use tabled P-values for Fisher’s F distribution requires this assumption to be made. Of course, we wish to cater for any situation where the z’s may be non-normal. This can be achieved by using a permutation procedure.

Note that, unlike the PERMANOVA one-way analysis, it does not make sense to simply permute the labels randomly among the groups to test the null hypothesis of homogeneity. This is because such permutations will cause potential differences in location among the groups to suddenly be included as part of the “within-group” measures. However, differences in location do not form part of the null hypothesis and so it is not logical to be mixing samples together from groups having different locations as possible alternative outcomes if the null hypothesis were true. Before proceeding, the observations from different groups must therefore be centered onto a common group centroid. In other words, it is the residuals obtained after removing any location differences that are actually exchangeable under the null hypothesis of homogeneity of dispersions. Permutation of residuals (Freedman & Lane 1983) is an approach that has been demonstrated to have good asymptotic and empirical properties (Anderson & Legendre 1999, Anderson & Robinson 2001). PERMDISP uses permutation of residuals (i.e., permutation of samples among groups after centering all groups onto a common location) in order to generate P-values for the test.

Levene’s test (for univariate data) can be made more robust (i.e. less affected by outliers) by using deviations from medians rather than deviations from means (Brown & Forsythe 1974, Conover et al. 1981). However, for multivariate data, there is more than one way to define a median (Haldane 1948, Gower 1974, Brown 1983). One definition of a median for multivariate data is the vector of medians for each individual variable (e.g., O’Brien 1992, Manly 1994). Another possibility is to invoke a spatial concept for the median as the point in the multivariate cloud which minimises the sum of the distances from each observation to that point, called the ‘mediancentre’ by Gower (1974). This spatial median is invariant to rotational changes in the axes, a quality which is not shared by the vector of medians of individual variables. This invariance to rotation is important for our purposes here, where PCO axes (which can involve rotations of the original data cloud) might be used, so PERMDISP (optionally) uses spatial medians.

An analysis that calculates the z values as deviations from spatial medians (rather than centroids) will clearly be less affected by outliers, so will tend to be more robust if the distribution of points in the data cloud is highly skewed for some reason. PERMDISP provides an option to perform the test based on medians rather than centroids: simply choose (Distances are to ● Medians) in the PERMDISP dialog box (Fig. 2.4). Note that the residuals which are permuted for the test based on medians are not the least-squares (LS) residuals (as are used for the test on the basis of centroids). Instead, the multivariate analogue of least absolute deviation (LAD) residuals (e.g., Cade & Richards 1996) are permuted instead. That is, the groups are first centered on to a common spatial median before proceeding with the permutations.

**Relevant aside:** If you have a single variable and the analysis is based on a Euclidean distance matrix, then PERMDISP using centroids and tables will give Levene’s (1960) original univariate test, while PERMDISP using medians and tables will give the modification of Levene’s test proposed by Brown & Forsythe (1974)\textsuperscript{54} for univariate data.

**Recommendation:** Simulations with multivariate ecological datasets indicated that the test using PERMDISP on the basis of distances to centroids, with P-values obtained using permutations, gave the best overall results (in terms of type I error and power, Anderson 2006). These are thus the default options for the routine.

\textsuperscript{53} The use of PCO axes to calculate centroids and the use of permutation methods to obtain P-values are two important ways that the method implemented by PERMDISP differs from the method proposed by Underwood & Chapman (1998).

\textsuperscript{54} The version of Levene’s test used most commonly and available in many statistical computer packages is this version based on medians.
An ecological example of the test for homogeneity is provided by considering a study by Warwick et al. (1990b) on coral assemblages from South Tikus Island, Indonesia. Percentage cover was measured along 10 transects for 75 species of coral in each of several years (1981-1988) which spanned an El Niño weather event occurring in 1982-83. For simplicity, we shall consider here only the years 1981, 1983 and 1985. The data are located in the file tick.pri in the ‘Corals’ folder of the ‘Examples v6’ directory for PRIMER. Open up the file in PRIMER and select only the years 81, 83 and 85 from the dialog box obtained by choosing Select > Samples > Factor levels > Factor name: Year and clicking on the ‘Levels’ button and picking out these years. An MDS plot of these samples on the basis of the Bray-Curtis measure (Fig. 2.5) shows a clear pattern of apparently much greater dispersion (variability) among the samples obtained in 1983, right after the El Niño event, with relatively less dispersion among samples obtained either before (1981) or two years after (1985).

To perform the test for the homogeneity of dispersions for these data (Fig. 2.6), click on the resemblance matrix and choose PERMANOVA+ > PERMDISP > (Group factor: Year) & (Distances are to Centroids) & (P-values are from Permutation) & (Num. permutations: 9999) & ( Do pairwise tests). Not surprisingly, the overall test comparing all three groups is highly statistically significant (Fig. 2.6). Individual pairwise tests show that there was no significant difference in dispersion between 1981 and 1985, but that the dispersion of assemblages for 1983 was significantly larger than either of these, having an average Bray-Curtis distance-to-centroid of over 62% (Fig. 2.6).

One of the important reasons for the large dissimilarities among samples in 1983 was because the El Niño caused a massive coral bleaching event, and many of the corals died. Those that remained were very patchy, were generally not the same species in different transects and had small cover values. A quick scan of the data file reveals the large number of zeros recorded in 1983; none of the coral species achieved percentage cover values above 4%. The sparse data in 1983 therefore yielded many Bray-Curtis similarities of zero (dissimilarities of 100%) among samples.

An extremely important point is that the test of dispersion is going to be critically affected by the transformation, standardisation and resemblance measure used as the basis of the analysis. It is pretty well appreciated by most practitioners that transforming the data has important effects on relative dispersions. Consider the fact that one of the common remedies to heterogeneity of

---

55 Clarke et al. (2006c) have discussed the potentially erratic behaviour and lack of discrimination of the Bray-Curtis measure for sparse data such as these, proposing an adjustment which consists of adding a “dummy” species (present everywhere) into the dataset. They show that, for these coral data, the adjustment does have the effect of stabilising the dispersion across years and optimising the year-to-year separation.
variances in the analysis of univariate data is indeed to perform an appropriate transformation (e.g., Box & Cox 1964). Such a transformation is (usually) designed explicitly to make the data more normally distributed (less skewed), to remove intrinsic mean-variance relationships (if any), and to render the variances essentially homogeneous among groups. Although the consequences of the use of transformations on resulting inferences is rarely articulated explicitly (McArdle & Anderson 2004), such an approach has merit and is widely used in univariate analyses. Thus, it is not at all difficult to understand that transformations will also affect relative dispersions in multivariate space.

Fig. 2.6. PERMDISP for Tikus Island coral data using Bray-Curtis, and also using Manhattan distances on log(x+1)-transformed data.

Much less well appreciated is the extent to which the choice of resemblance measure can affect perceived patterns and results regarding relative dispersions among groups. For example, consider the analysis of the percentage cover data on corals from South Tikus Island. The Bray-Curtis measure is known to display erratic behaviour for sparse data such as these (Clarke et al. 2006c). We might consider an analysis of the same dataset using some other measure, such as a Euclidean or Manhattan distance on log(x+1)-transformed cover values. Such an approach could be considered reasonable on the grounds that the transformation will appropriately reduce the effects of large cover values, and a measure such as the Manhattan distance does not exclude joint absence information. In the present context, the joint absences of coral species (having been killed by bleaching) might indeed be considered to indicate similarity between samples. An MDS of the samples on the basis of the Manhattan measure on log(x+1)-transformed values shows a dramatically different pattern to what was shown using the Bray-Curtis measure (cf. Fig. 2.7 with Fig. 2.5). The points corresponding to samples from 1983 form a very tight cluster, samples from 1985 are a bit more dispersed, and the dispersion among samples from 1981 are much more dispersed (Fig. 2.7). The results of the PERMDISP analysis indicate that these differences in relative within-group dispersions among the groups (1981 > 1985 > 1983) are all highly statistically significant (see Fig. 2.6 above).

What we mean when we say “heterogeneity in multivariate dispersions” therefore must be qualified by reference to the particular resemblance measure we have chosen to use. Clearly, in the present case, the inclusion or exclusion of joint absence information can dramatically alter the results. In
this particular example, the effects of transformations were actually quite modest by comparison\textsuperscript{56}. Of course, the nature of the patterns obtained in any particular case will depend, however, on the nature of the data. In some cases, the transformation chosen will have dramatic effects.

\textbf{Fig. 2.7.} MDS of coral assemblages from South Tikus Island in each of three years on the basis of Manhattan distances of log(x+1)-transformed percentage cover.

There are many new areas to explore concerning the effects of different resemblance measures and transformations on relative within-group dispersions in different situations. The point is: be careful to define what is meant by “variability in assemblage structure” and realise that this is specific to the measure you have chosen to use and the nature of your particular data. Also, if you are intending to analyse the data using PERMANOVA as well, then of course it makes sense to choose the same transformation/standardisation/resemblance measure for the PERMDISP routine as were used for the PERMANOVA in order to obtain reasonable joint interpretations and inferences.

In addition, as PERMANOVA uses common measures of variability in the construction of \(F\)-ratios for tests, homogeneity is definitely implicit in the analysis of a resemblance matrix using PERMANOVA. Thus, after calculating a resemblance matrix which captures the desired ecological/community properties of the data, analysis by PERMDISP and visual assessment in ordination plots will highlight potential heterogeneities that could lead to a suitable degree of caution in the interpretation of results from a PERMANOVA model of the variation. This is not to say that a non-significant result using PERMDISP is an absolute requirement before using PERMANOVA; it is expected that PERMDISP will be powerful enough to detect small differences in dispersion that may not affect PERMANOVA adversely. However, the closer one can get to stabilising the relative dispersions among groups (or among cells in higher-way designs, etc.), the more valid and clear will be the interpretations from a PERMANOVA analysis.

When used on species composition (presence/absence) data in conjunction with certain resemblance measures, the test for homogeneity of multivariate dispersions is directly interpretable as a test for similarity in \textit{beta diversity} among groups (Anderson \textit{et al.} 2006). Whittaker (1960, 1972) defined beta diversity as the degree to which a set of observations in a given geographical area vary in the identities of species they contain. More specifically, he proposed a measure of beta diversity as the proportion by which a given area is richer than the average of the samples within it. Although there may be many ways to define beta diversity (e.g., Vellend 2001, Magurran 2004), Anderson \textit{et al.} (2006) considered that beta diversity can be broadly defined as the \textit{variability in species composition} among sampling units for a given area at a given spatial scale. Whittaker’s measure (as a proportion) only provides a single value per area (or group), so cannot be used to test

\textsuperscript{56} Results obtained using Bray-Curtis on log(x+1)-transformed data are similar to what is obtained in Fig. 2.5 and results obtained using Manhattan on untransformed data are similar to what is obtained in Fig. 2.7 (try it)!
for differences among groups in beta diversity\textsuperscript{57}. However, PERMDISP on the basis of ecological measures of compositional dissimilarity (e.g., Jaccard or Sørensen, which is just Bray-Curtis on presence/absence data) can be used for such a test. Note that the definition of beta diversity is focused on variability in composition. Thus, multivariate dispersion on the basis of any resemblance measure that includes relative abundance information as well will not necessarily provide a measure of beta diversity, per se.\textsuperscript{58}

Ellingsen & Gray (2002) studied beta diversity and its relationship with environmental heterogeneity in benthic marine systems over large spatial scales in the North Sea. Samples of soft-sediment macrobenthic organisms were obtained from \( N = 101 \) sites occurring in five large areas along a transect of 15 degrees of latitude (Fig. 2.8). A total of \( p = 809 \) taxa were recorded overall, and samples consisted of abundances pooled across five benthic grabs obtained at each site. The upper 5 cm of one additional grab was also sampled to measure environmental variables at each site. Of interest was to measure beta diversity (the degree of compositional heterogeneity) for each of these five areas and to compare this with variation in the environmental variables. The biological data are provided in the file norbio.pri in the ‘NorMac’ folder in the ‘Examples add-on’ directory.

An MDS plot on the basis of the Jaccard measure shows patterns of differences in assemblage composition among the five areas (Fig. 2.9). The Jaccard measure is directly interpretable as the percentage of unshared species. It uses only presence/absence information and measures of multivariate dispersion based on this measure are indeed interpretable as measures of beta diversity. Perhaps the most striking thing emerging from the plot is the quite large spread of sample points corresponding to area 3 and the quite tight cluster of points corresponding to area 1 compared to the other areas. The test for homogeneity reveals very strong differences among the five groups and, more particularly, identifies group 1 and group 3 as being significantly different from one another and from the other three groups (2, 4 and 5) in terms of their variability in species composition\textsuperscript{59} (Fig. 2.9). The average Jaccard distance-to-centroid is about 36\% for group 1, but is much larger (more than 56\%) for group 3. This pattern of heterogeneity was mirrored by similar patterns of variability in the environmental variables for the three areas (see Ellingsen & Gray 2002).

\textsuperscript{57}See Kiflawi & Spencer (2004), however, who have shown how expressing beta diversities in terms of an odds ratio does allow confidence intervals to be constructed.

\textsuperscript{58}If we allow dispersion based on any resemblance measure to be considered “beta diversity”, then beta diversity simply becomes a non-concept (\textit{sensu} Hurlbert 1971). For example, compare Figures 2.5 and 2.7; surely we cannot claim that both of these describe relative patterns in beta diversity for the same dataset!

\textsuperscript{59}Note that in PERMDISP, as in PERMANOVA, the pairwise tests are not corrected for multiple comparisons. See the section on Pair-wise comparisons in chapter 1 on PERMANOVA for more details.
and Anderson et al. 2006 for more details). One possible explanation for the relatively large biological and environmental variation in area 3 is that this may be an area of rapid transition from the southern to the northern climes.

Fig. 2.9. MDS of Norwegian macrofauna from 5 areas (labelled 1-5 as on the map in Fig. 2.8) based on the Jaccard measure and the results of PERMDISP, alongside, comparing beta diversity among the 5 areas.

There is one necessary restriction on the use of PERMDISP, which is that the number of replicate samples per group must exceed \( n = 2 \). The reason is that, if there are only two replicates, then, by definition, the distance to the centroid for those two samples must be equal to one another. Consider a single variable and a group with two samples having values of 4 and 6. The centroid (average) in Euclidean space for this group is therefore 5. The distance from sample 1 to the centroid is 1 and the distance from sample 2 to the centroid is also 1. These two values of \( z \) are necessarily equal to one another. This will also be the case for other groups having only 2 replicate samples, so the within-group variance of the \( z \)'s when \( n = 2 \) for all groups will be equal to zero. If the within-group variance is equal to zero, then the \( F \) statistic will be infinite, so the test loses all meaning. Clearly, the test is also meaningless for a group with \( n = 1 \), which will have only a single \( z \) value of zero. Thus, if the sample size for any of the groups is \( n \leq 2 \), then the PERMDISP routine will issue a warning accordingly. Although test results are meaningless in such cases, the individual deviations (the \( z \)'s) can nevertheless still be examined and compared in their value across the different groups, if desired. More generally, the issue here is the degree of correlation among values of \( z \), which increases the smaller the sample size. Levene (1960) showed the degree of correlation is of order \( n^{-2} \), which, he suggested, will probably not have a serious effect on the distribution of the \( F \) statistic. We suggest that formal tests using PERMDISP having within-group sample sizes less than \( n = 10 \) should be viewed with some caution and those having sample sizes less than \( n = 5 \) should probably be avoided, though (as elsewhere) further simulation studies for realistic multivariate cases would be helpful in refining such rules-of-thumb.

In many situations, the experimental design is not as simple as a one-way analysis among groups. For more complex designs, several tests of dispersion may be possible and relevant at a number of different levels. The sort of tests that will be logical to do in any given situation will depend on the design and the nature of existing location effects (if any) that were detected by PERMANOVA.
This is particularly the case when factors can interact with one another (see the section on crossed designs below). First, however, we shall consider relevant tests of dispersion that might be of interest in the case of a nested experimental design. For simplicity, we shall limit our discussion here to the two-factor nested design, although the essential principles discussed will of course apply to situations where there are greater numbers of factors in a nested hierarchy.

Consider the experimental design described by Anderson et al. (2004) investigating the potential effects of different depositional environments on benthic intertidal fauna of the Okura estuary in New Zealand. The hydrology and sediment dynamics of the estuary had been previously modeled (Cooper et al. 1999, Green & Oldman 1999), and areas corresponding to high, medium or low probability of sediment deposition were identified. There were \( n = 6 \) sediment cores (13 cm in diameter \( \times 15 \) cm deep) obtained from random positions within each of 15 sites along the estuary, with 5 sites from each of the high, medium and low depositional types of environments (Fig. 2.10). Sampling was repeated 6 times (twice in each of three seasons in 2001-2002), but we shall focus for simplicity on the following spatially nested design for the first time of sampling only:

Factor A: Deposition (fixed with \( a = 3 \) levels, high (H), medium (M) or low (L)).
Factor B: Site (random with \( b = 5 \) levels, nested in Deposition).

The design was balanced, with \( n = 6 \) replicate cores per site for a total of \( N = a \times b \times n = 90 \) samples obtained at each time. The data (counts of \( p = 73 \) taxa) are located in the file \okura.pri in the ‘Okura’ folder of the ‘Examples add-on’ directory. We shall follow Anderson et al. (2004) and perform the analysis on the basis of the Bray-Curtis measure on log(x+1)-transformed abundances. An MDS plot of the data from time 1 only (Fig. 2.11) revealed a fairly clear pattern to suggest that assemblages from different depositional environments were distinguishable from one another, especially those from sites having relatively high probabilities of deposition.

![Fig. 2.10. Okura estuary, with the 15 sites for sampling of benthic fauna labeled: high depositional sites are underlined, medium depositional sites are in italics and low depositional sites are in normal font.](image)

For a design such as this, there are two levels at which we may wish to think about relative dispersions. First, are the multivariate dispersions among the 6 cores within a site different among sites? Second, are the multivariate dispersions among the 5 site centroids different among the three different depositional environments? Another (third) possibility might be to compare the dispersions of the \( 5 \times 6 = 30 \) cores across the three depositional environments. However, such a test would only really be logical if there were no differences in location among sites (i.e. no significant ‘Site’ effects in the PERMANOVA).
First, consider the variability among cores within each site. For these particular data, the 15 sites are labeled 1-15 according to their position in the estuary (Fig. 2.10). Thus, to compare dispersions of assemblages in cores across sites, we simply run PERMDISP on the factor ‘Site’ for the full resemblance matrix. Note that these $a \times b = 15$ cells correspond to the lowest-level cells in the design. If the sites had been labeled 1-5, that is, if the sites had been given the same labels within each of the depositional environments (even though they are different actual sites, being nested), then we would have to first create a new factor corresponding to the fifteen cells (all combinations of factors A and B) by choosing Edit > Factors > Combine.
This first PERMDISP analysis reveals that the dispersion among cores varies significantly from site to site (Fig. 2.12, $F = 8.86, P < 0.001$). As ‘Site’ is a random factor, we are not especially interested in performing pairwise comparisons here, so these have not been done. There is heterogeneity in dispersions among cells and examining the output reveals that the average distance-to-centroid in Bray-Curtis space within a site generally varies from about 18 to 27%. Three of the sites, however, have an average distance-to-centroid of nearly 40% (Fig. 2.12).

Next, we shall consider the second question posed above: are there differences in the dispersions of the 5 site centroids for different depositional environments? Before proceeding, we first need to obtain a distance matrix among the site centroids. Recall, however, that the centroids in Bray-Curtis (or some other non-Euclidean space) are not the same as the arithmetic centroids calculated on the original variables. Thus, unfortunately, we cannot calculate the site centroids by going back to the raw data and just calculating site averages for the original variables. Instead, we shall use a new tool available as part of the PERMANOVA+ add-on, which calculates a resemblance matrix among centroids for groups identified by a factor in the space of the chosen resemblance measure. To do this for the present example, click on the resemblance matrix and then choose PERMANOVA+ > Distances among centroids… > Grouping factor: Site, then click ‘OK’. The resulting resemblance matrix contains the correct Bray-Curtis dissimilarities among the site centroids, which have been calculated using PCO axes (see the section Generalisation to dissimilarities above). Just to re-iterate, these centroids are not calculated on the original data, they are calculated on PCO axes obtained from the resemblance matrix, in order to preserve the resemblance measure chosen as the basis of the analysis.

To visualise the relative positions of centroids in multivariate space, choose Analyse > MDS from this resemblance matrix among centroids. Conveniently, PRIMER has retained all of the factors and labels associated with these points, so we can easily place appropriate labels and symbols onto the centroids in the plot. The dispersions of the site centroids appear to be roughly similar for the three depositional environments (H, M and L, Fig. 2.13), and indeed the test for homogeneity of dispersions revealed no significant differences among these three groups (Fig. 2.12, $F = 1.47, P = 0.43$). Note that ‘Deposition’ is a fixed factor and so we would indeed have been interested in the pair-wise comparisons among the three groups (had there been a significant $F$-ratio). This is why the option Do pairwise tests was chosen; these results are also shown in the output (Fig. 2.12).

Finally, we may consider the third question above: are there differences in dispersions among cores (ignoring sites) for the three depositional environments? Such a question only makes sense if sites have no effects. However, the results of the two-factor PERMANOVA for this experimental design
reveals that ‘Site’ effects are highly statistically significant (pseudo-\(F = 5.49, P < 0.001\), Fig. 2.14). Therefore, there is no logical reason to consider the multivariate dispersion of cores in a manner that ignores sites. As an added note of interest, the PERMANOVA test of the factor ‘Deposition’ in the two-factor nested design yields the same results as a one-way PERMANOVA for ‘Deposition’ using the resemblance matrix among site centroids (pseudo-\(F = 4.56\) with 2 and 12 df, \(P < 0.001\), Fig. 2.14). This clarifies how the nested model effectively treats the levels of the nested term (in this case, ‘sites’) as replicates for the analysis of the upper-level factor. Note that this equivalence would not hold if the centroids had been calculated as averages from the raw (or even the transformed) original data, which further emphasises that the centroids obtained using PCOs are indeed the correct ones for the analysis.

![Figure 2.14](image)

Fig. 2.14. PERMANOVA of the Okura data from time 1 according to the two-factor nested design and the test for depositional effects alone in a one-factor design on the basis of resemblances among site centroids.

When two factors are crossed with one another, there may be several possible hypotheses concerning relative dispersions among groups. An important aspect of analysing multi-factorial designs is to keep straight when you are focusing on dispersion effects and when you are focusing on location effects. Running PERMANOVA and PERMDISP in tandem, driven by carefully articulated hypotheses, provides a useful general approach.

Here we shall consider a study to investigate effects of marine reserve status and habitat on cryptic fish assemblages in northeast New Zealand, as described by Willis & Anderson (2003). Although marine reserves in this region afford protection for larger exploited fish species (such as snapper, Pagrus auratus, which occur in greater relative abundances and sizes inside reserves, see Willis et al. 2003), relatively little is known regarding the potential flow-on effects (either positive or negative) on many of the other components of these ecosystems (but do see Shears & Babcock 2002, 2003 regarding habitat-scale effects). For example, do assemblages of smaller and/or more cryptic fish (such as triplefins and gobies) change with increased densities of predators inside versus outside marine reserves? Is this effect (if any) modified in kelp forest habitat as opposed to urchin-grazed ‘barrens’ habitat? The two factors in the design were:

Factor A: Status (fixed with \(a = 2\) levels, reserve (R) or non-reserve (N) sites).

Factor B: Habitat (fixed with \(b = 2\) levels, kelp forest (k) or urchin-grazed rock flats (r), crossed with Status).

There were \(n = 6\) replicate \(9\) m\(^2\) plots surveyed within each of the \(a \times b = 4\) combinations of levels of the above two factors, yielding counts of abundances for \(p = 33\) species. These data are located in the file cryptic.pri in the ‘Cryptic’ folder of the ‘Examples add-on’ directory. An MDS on the basis of the Bray-Curtis measure on square-root transformed abundances does not show very clear effects of either of these factors (Fig. 2.15). Formal tests are necessary to clarify whether any
significant effects of either factor (or their interaction) are detectable in the higher-dimensional multivariate space.

First, we shall run a PERMANOVA on the two-factor design, bearing in mind that the test is specifically designed to detect location effects, but dispersion effects (if any) might also be detected by this test. Results indicate that the two factors do not interact ($P > 0.70$), but there are significant independent effects of ‘Status’ and ‘Habitat’ ($P < 0.01$ in each case, Fig. 2.16).

**Fig. 2.15.** MDS of cryptic fish assemblages at reserve (R) or non-reserve (N) sites in either kelp forest or urchin-grazed rock flat habitats.

**Fig. 2.16.** PERMANOVA for the two-way analysis of cryptic fish, along with pair-wise comparisons for each of the main effects.

Although there are only two levels of each of the factors (so the tests in the main PERMANOVA analysis are enough to indicate the significant difference between the two levels, in each case), pair-wise comparisons done on each factor will give us a little more information regarding the average similarity between and within the different groups (Fig. 2.16). The average within-group
similarities are comparable for ‘N’ and ‘R’, but the average within-group similarity in the kelp forest habitat (‘k’, 40.4%) is much smaller than that in the rock flat habitat (‘r’, 55.8%). The former is even smaller than the average similarity between kelp and rock-flat habitats (44.6%)! The pattern in the MDS plot (Fig. 2.15) also indicates greater dispersion of points for ‘k’ (squares) than for ‘r’ (downward triangles). These observations provide signals to suggest (but do not prove) that the significant effect of habitat might be driven somewhat by differences in dispersions. Of course, we can investigate this idea formally using PERMDISP.

First, although there is no significant interaction detected by PERMANOVA, it is still possible that the individual cells may differ in their relative dispersions. This can be tested by combining the two factors to make a single factor with \(a \times b = 4\) levels and then running PERMDISP to compare these within-cell dispersions. Choose Edit > Factors > Combine and then ‘Include’ both factors in the ‘Ordered selection’ dialog box (Fig. 2.17). Next choose PERMANOVA+ > PERMDISP > (Group factor: StatusHabitat) & (Distances are to •Centroids) & (P-values are from •Permutation) & (Num. permutations: 9999). For these data, the null hypothesis of equal dispersions among the four cells is retained (\(F = 2.0, P > 0.25\), Fig. 2.17).

![Fig. 2.17. Menus used to generate a combined factor consisting of the 4 combinations of ‘Status’ and ‘Habitat’, and the results window from PERMDISP which formally compares dispersions among these four cells.](image)

Next, to consider analyses of dispersions for main effects in a crossed design (such as this), it is important to consider what might be meaningful in light of the PERMANOVA results. When two (or more) factors are crossed with one another, it is easily possible to confuse a dispersion effect with an interaction effect. To illustrate this point, suppose I have a two-factor design, where there is a significant interaction such as that shown in Fig. 2.18. More specifically, suppose there is no effect of factor A in level 1 of factor B (grey/filled symbols), but there is a significant effect of factor A in level 2 of factor B (white/open symbols). A test for equality of dispersions among the levels of factor B that ignored A would clearly find a significant difference between the two levels of factor B (with white symbols obviously being more dispersed than grey symbols). However, this is not really caused by dispersion differences per se, but rather is due to the fact that factor A interacts with factor B. Therefore, before performing tests of dispersion for main effects in crossed designs, the potential for interactions to confound interpretations of the results should be examined.
This is generally achieved by first testing and investigating the nature of any significant interaction terms using PERMANOVA, and also by examining patterns in ordination plots.

We have seen, for the cryptic fish data, that no interaction was detected between ‘Status’ and ‘Habitat’ (Fig. 2.16), so we can perform the test of homogeneity of dispersions for each of the main effects separately (ignoring the other factor) using PERMDISP. These reveal no significant differences in dispersion for the ‘Status’ factor \((F = 0.06, P > 0.82)\), but significant differences in dispersion for the ‘Habitat’ factor, with greater variability in the fish assemblages observed in kelp forests compared to the urchin-barrens rock flat habitat \((F = 7.8, P < 0.05, \text{Fig. 2.19})\)\(^{60}\).

![Fig. 2.18. Schematic diagram in two dimensions of an interaction between two factors that can result in misinterpretation of tests of dispersion for main effects.](image)

![Fig. 2.19. Results of PERMDISP on each of the main effects from the two-factor study of cryptic fish.](image)

Note that the significant effect of the marine reserve ‘Status’ detected by the PERMANOVA is therefore now interpretable to be purely (in essence) a location effect, as the test for homogeneity of dispersions for this factor was not significant. However, the effect of ‘Habitat’ appears to be primarily (although perhaps not entirely) driven by differences in dispersion, with greater variability in the structure of cryptic fish assemblages observed in kelp forest habitats compared to rock flat habitats. If there is a location effect (i.e. a shift in multivariate space) due to habitat in addition to the dispersion effect detected (which of course is possible), then this should be discernible in MDS plots (preferably carried out separately for each reserve status in 3-d, when the stress will drop to lower values than in Fig. 2.15).

---

\(^{60}\) A word of warning: PERMDISP analyses and measures of dispersion for main effects like this will include the location shifts (if any) of other main effects. Although this won’t necessarily confound the analysis if the factors do not interact, it can nevertheless compromise power. Ideally, we would like to remove location effects of all potential factors before proceeding with the test. This is not implemented in the current version of PERMDISP, but there is clearly scope for further development here.
PERMDISP is designed to test the null hypothesis of no differences in dispersions among *a priori* groups. Although this may be the only goal in some cases, PERMDISP can also provide a useful companion test to PERMANOVA in order to clarify the nature of multivariate effects on the basis of a chosen resemblance measure.

Table 2.1 provides a synopsis of the logical inferences arising from the outcomes of PERMDISP and PERMANOVA for the one-way (single-factor) case. Three of the four scenarios yield a clear inferential outcome; the fourth leaves some uncertainty for interpretation. Namely, if both PERMDISP and PERMANOVA tests are significant, we will know that dispersion effects occur, but we may not necessarily know if location effects are also present. Examining ordination plots and the relative sizes of within and between-group resemblances will be important here, but it is undoubtedly true that there is scope for further work on the implications of a significant outcome for the PERMDISP test on the structure and interpretation of the associated PERMANOVA tests. Finally, in more complex designs with crossed or nested factors, several different PERMDISP analyses may be needed (e.g., at different levels in the design) in order to clarify dispersion effects, if any.

**Table 2.1.** Outcomes and potential inferences to be drawn from joint analyses done using PERMANOVA and PERMDISP for the one-way case.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>PERMANOVA</th>
<th>PERMDISP</th>
<th>Inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Not significant</td>
<td>Not significant</td>
<td>No location or dispersion effects</td>
</tr>
<tr>
<td>2</td>
<td>Significant</td>
<td>Not significant</td>
<td>Location effect only</td>
</tr>
<tr>
<td>3</td>
<td>Not significant</td>
<td>Significant</td>
<td>Dispersion effect only</td>
</tr>
<tr>
<td>4</td>
<td>Significant</td>
<td>Significant</td>
<td>Dispersion effect and perhaps (although not necessarily) a location effect as well.</td>
</tr>
</tbody>
</table>
3. Principal coordinates analysis (PCO)

Method: Torgerson (1958), Gower (1966)

PCO is a routine for performing principal coordinates analysis (Gower 1966) on the basis of a (symmetric) resemblance matrix. PCO is also sometimes referred to as classical scaling, with origins in the psychometric literature (Torgerson 1958). PCO places the samples onto Euclidean axes (i.e., so they can be drawn) using only a matrix of inter-point dissimilarities (e.g., Legendre & Legendre 1998). Ordination using non-metric multi-dimensional scaling (MDS) focuses on preserving only the rank order of dissimilarities for an a priori chosen number of dimensions, whereas ordination using PCO (like PCA) is a projection onto axes, but (unlike PCA) in the space of the dissimilarity measure chosen. The user chooses the number of PCO axes to include in the output, but generally only the first two or three axes are drawn in ordination plots. A further utility provided in the PERMANOVA+ add-on that uses PCO axes is the option to calculate distances among centroids for levels of a chosen factor. This allows visualisation of centroids from an experimental design in the space of the resemblance measure chosen.

Rationale

It is difficult to visualise patterns in the responses of whole sets of variables simultaneously. Each variable can be considered a dimension, with its own story to tell in terms of its mean, variance, skewness, etc. For most sets of multivariate data, there are also correlations among the variables. Ordination is simply the ordering of samples in Euclidean space (e.g., on a page) in some way, using the information provided by the variables. The primary goal of ordination methods is usually to reduce the dimensionality of the data cloud in order to allow the most salient patterns and structures to be observed.

Different ordination methods have different criteria by which the picture in reduced space is drawn (see chapter 9 of Legendre & Legendre 1998 for a more complete discussion). For example:

- Non-metric multi-dimensional scaling (MDS) preserves the rank order of the inter-point dissimilarities (for whatever resemblance measure has been chosen) as well as possible within the constraints of a small number of dimensions (usually just two or three). The adequacy of the plot is ascertained by virtue of how well the inter-point distances in the reduced-dimension, Euclidean ordination plot reflect the rank orders of the underlying dissimilarities (see, for example, chapter 5 in Clarke & Warwick 2001).
- Principal components analysis (PCA) is a projection of the points (perpendicularly) onto axes that minimise residual variation in Euclidean space. The first principal component axis is defined as the straight line drawn through the cloud of points such that the variance of sample points, when projected perpendicularly onto the axis, is maximised (see, for example, chapter 4 of Clarke & Warwick 2001).
- Correspondence analysis (CA) is also a projection of the points onto axes that minimise residual variation, but this is done in the space defined by the chi-squared distances among points (ter Braak 1987, Minchin 1987, Legendre & Legendre 1998); PCA performed on a resemblance matrix of Euclidean distances reproduces the pattern and results that would be obtained by a PCA on the original variables. Similarly, PCO performed on a resemblance matrix of chi-squared distances reproduces the pattern that would be obtained by a CA on the original variables. Thus, PCO is a more general procedure than either PCA or CA, yielding a projection in the space indicated by the resemblance measure chosen.

PCO performed on a resemblance matrix of Euclidean distances reproduces the pattern and results that would be obtained by a PCA on the original variables. Similarly, PCO performed on a resemblance matrix of chi-squared distances reproduces the pattern that would be obtained by a CA on the original variables. Thus, PCO is a more general procedure than either PCA or CA, yielding a projection in the space indicated by the resemblance measure chosen.

Two other features of PCO serve to highlight its affinity with PCA (as a projection). First, the scales of the resulting PCO axes are interpretable in the units of the original resemblance measure. Although the distances between samples in a plot of few dimensions will underestimate the distances in the full-dimensional space, they are, nevertheless, estimated in the same units as the
original resemblance measure, but as projected along the PCO axes. Thus, unlike MDS axes, the PCO axes refer to a non-arbitrary quantitative scale defined by the chosen resemblance measure. Second, PCO axes (like PCA axes) are centered at zero and are only defined up to their sign. So, any PCO axis can be reflected (by changing the signs of the sample scores), if convenient.61

For many multivariate applications (especially for species abundance data), MDS is usually the most appropriate ordination method to use for visualising patterns in a small number of dimensions, because it is genuinely the most flexible and robust approach available (Kenkel & Orlóci 1986, Minchin 1987). There is a clear relationship in the philosophy underlying the ANOSIM testing procedure and non-metric MDS in PRIMER. Both are non-parametric approaches that work on the ranks of resemblance measures alone. However, when using PERMANOVA to perform a partitioning for more complex designs, it is the actual dissimilarities (and not just their ranks) that are of interest and which are being modeled directly (e.g., see the section PERMANOVA versus ANOSIM in chapter 1). Therefore, we may wish to use an ordination procedure that is a little more consistent with this philosophy, and PCO may do this by providing a direct projection of the points in the space defined by the actual dissimilarities themselves. Although MDS will provide a more optimal solution for visualising in a few dimensions what is happening in the multi-dimensional cloud, the PCO can in some cases provide additional insights regarding original dissimilarities that might be lost in the non-metric MDS, due to ranking. In addition (as stated in chapter 2), the Euclidean distance between two points in the space defined by the full set of PCO axes (all together) is equivalent to the original dissimilarity between those two points using the chosen resemblance measure on the original variables.62 So another main use of PCO axes is to obtain distances among centroids, which can then form the basis of further analyses when dealing with more complex and/or large multi-factorial datasets.

To construct axes that maximise fitted variation (or minimise residual variation) in the cloud of points defined by the resemblance measure chosen, the calculation of eigenvalues (sometimes called “latent roots”) and their associated eigenvectors is required. It is best to hold on to the conceptual description of what the PCO is doing and what it produces, rather than to get too bogged down in the matrix algebra required for its computation. More complete descriptions are available elsewhere (e.g., Gower 1966, Legendre & Legendre 1998), but in essence, the PCO is produced by doing the following steps (Fig. 3.1):

1. From the matrix of dissimilarities, D, calculate matrix A, defined (element-wise) as minus one-half times each dissimilarity (or distance)63;  
2. Centre matrix A on its row and column averages, and on the overall average, to obtain Gower’s centred matrix G;  
3. Eigenvalue decomposition of matrix G yields eigenvalues (\(\lambda_i\), \(i = 1, \ldots, N\)) and their associated eigenvectors.  
4. The PCO axes \(Q\) (also called “scores”) are obtained by multiplying (scaling) each of the eigenvectors by the square root of their corresponding eigenvalue.64

The eigenvalues associated with each of the PCO axes provide information on how much of the variability inherent in the resemblance matrix is explained by each successive axis (usually expressed as a percentage of the total). The eigenvalues (and their associated axes) are ordered from largest to smallest, and their associated axes are also orthogonal to (i.e., perpendicular to or independent of) one another. Thus, \(\lambda_1\) is the largest and the first axis is drawn through the cloud of points in a direction that maximises the total variation along it. The second eigenvalue, \(\lambda_2\), is second-largest and its corresponding axis is drawn in a direction that maximises the total variation

---

61 This is done within PRIMER in the same manner as for either an MDS or PCA plot, by choosing Graph > Flip X or Graph > Flip Y in the resulting configuration.

62 With appropriate separate treatment of the axes corresponding to the positive and negative eigenvalues, if any, see McArdle & Anderson (2001) and Anderson (2006) and the section on Negative eigenvalues for details.

63 If a matrix of similarities is available instead, then the PCO routine in PERMANOVA+ will automatically translate these into dissimilarities as an initial step in the analysis.

64 If the eigenvalue (\(\lambda_j\)) is negative, then the square root of its absolute value is used instead, but the resulting vector is an imaginary axis (recall that any real number multiplied by \(i = \sqrt{-1}\) is an imaginary number).
along it, with the further caveat that it must be perpendicular to (independent of) the first axis, and so on. Although the decomposition will produce \( N \) axes if there are \( N \) points, there will generally be a maximum of \( N - 1 \) non-zero axes. This is because only \( N - 1 \) axes are required to place \( N \) points into a Euclidean space. (Consider: only 1 axis is needed to describe the distance between two points, only 2 axes are needed to describe the distances among three points, and so on…). If matrix \( \mathbf{D} \) has Euclidean distances to begin with and the number of variables (\( p \)) is less than \( N \), then the maximum number of eigenvalues will be \( p \) and the PCO axes will correspond exactly to principal component axes that would be produced using PCA.

\[
\mathbf{D}_{N \times N} = (d_{ij})
\]

Calculate matrix \( \mathbf{A}_{N \times N} = (a_{ij}) \)

\[
a_{ij} = (-\frac{1}{2} d_{ij}^2)
\]

Centering

\[
\mathbf{G}_{N \times N} = (g_{ij})
\]

\[
(g_{ij}) = (a_{ij} - \bar{a}_i - \bar{a}_j + \bar{a})
\]

Eigenvalue decomposition

\[
\mathbf{Q}_{N \times N}
\]

**Fig. 3.1.** Schematic diagram of the mechanics of a principal coordinates analysis (PCO).

The full set of PCO axes *when taken all together* preserve the original dissimilarities among the points given in matrix \( \mathbf{D} \). However, the *adequacy* of the representation of the points as projected onto a smaller number of dimensions is determined for a PCO by considering how much of the total variation in the system is explained by the first two (or three) axes that are drawn. The two- (or three-) dimensional distances in the ordination will underestimate the true dissimilarities\(^{65}\). The percentage of the variation explained by the \( i \)th PCO axis is calculated as \((100 \times \lambda_i / \Sigma \lambda_i)\). If the percentage of the variation explained by the first two axes is low, then distances in the two-dimensional ordination will not necessarily reflect the structures occurring in the full multivariate space terribly well. How much is “enough” for the percentage of the variation explained by the first two (or three) axes in order to obtain meaningful interpretations from a PCO plot is difficult to establish, as it will depend on the goals of the study, the original number of variables and the number of points in the system. We suggest that taking an approach akin to that taken for a PCA is appropriate also for a PCO. For example, a two-dimensional PCO ordination that explains ~70% or more of the multivariate variability inherent in the full resemblance matrix would be expected to provide a reasonable representation of the general overall structure. Keep in mind, however, that it is possible for the percentage to be lower, but for the most important features of the data cloud still to be well represented. Conversely, it is also possible for the percentage to be relatively high, but for considerable distortions of some distances still to occur, due to the projection.

As an example, consider the data on Victorian avifauna at the level of individual surveys, in the file vicsurv.pri of the ‘VielAvi’ folder in the ‘Examples add-on’ directory. For simplicity, we shall focus only on a subset of the samples: *Select > Samples > Factor levels > Factor name: Treatment > Levels…* and select only those samples taken from either ‘good’ or ‘poor’ sites. Duplicate this sheet containing only the subset of data and rename it vics.good.poor. Next, choose *Analyse > Resemblance > (Analyse between•Samples) & (Measure•Bray-Curtis similarity) & (✓Add

---

\(^{65}\) Except in certain rare cases, where the first two or three axes might explain greater than 100% of the variability! See the section on Negative eigenvalues.
3. PCO

dummy variable > Value: \[1\]). From this resemblance matrix (calculated using the adjusted Bray-Curtis measure), choose **PERMANOVA+ > PCO >** (Maximum no of PCOs: \([15]\)) & (✓Plot results). The default maximum number of PCOs corresponds to \(N - 1\).

The output provided from the analysis includes two parts. First, a text file (Fig. 3.2) with information concerning the eigenvector decomposition of the \(G\) matrix, including the percentage of the variation explained by each successive PCO axis. Values taken by individual samples along each PCO axis (called “scores”) are also provided. Like any other text file in PRIMER, one can cut and paste this information into a spreadsheet outside of PRIMER, if desired. Note there is also the option (✓Scores to worksheet) in the PCO dialog box (Fig. 3.2), which allows further analyses of one or more of these axes from within PRIMER.

The other part of the output is a graphical object consisting of the ordination itself (Fig. 3.3). By default, this is produced for the first two PCO axes in two dimensions, but by choosing **Graph > Special**, the full range of options already in PRIMER for graphical representation is made available for PCO plots, including 3-d scatterplots, bubble plots, the ability to overlay trajectories or cluster solutions, and a choice regarding which axes to plot. One might choose, for example, to also view the third and fourth PCO axes or the second, third and fourth in 3-d, etc., if the percentage of the variation explained by the first two axes is not large. In addition, a new feature provided as part of the PERMANOVA+ add-on is the ability to superimpose vectors onto the plot that correspond to the raw correlations of individual variables with the ordination axes (see the section Vector overlays below).

![Fig. 3.2. PCO on a subset of the Victorian avifauna survey data.](image)
In the case of the Victorian avifauna, we see the percentage of the total variation inherent in the resemblance matrix that is explained by these first two axes is 75.4% (Fig. 3.2), so the two-dimensional projection is likely to capture the salient patterns in the full data cloud. Also, the two-dimensional plot shows a clear separation of samples from ‘poor’ vs ‘good’ sites on the basis of the (adjusted) Bray-Curtis measure (Fig. 3.3). Although not labeled explicitly, the two ‘poor’ sites are also clearly distinguishable from one another on the plot — the labels 1-4 in the lower-right cluster of points correspond to the four surveys at one of these sites, while the 4 surveys for the other ‘poor’ site are all located in the upper right of the diagram. This contrasts with the avifauna measured from the ‘good’ sites, which are well mixed and not easily distinguishable from one another through time (towards the left of the diagram).

The sharp-sighted will have noticed a conundrum in the output given for the Victorian avifauna shown in Fig. 3.2. The values for the percentage variation explained for PCO axes 10 through 15 are negative! How can this be? Variance, in the usual sense, is always positive, so this seems very strange indeed! It turns out that if the resemblance measure used is not embeddable in Euclidean space, then some of the eigenvalues will be negative, resulting in (apparently) negative explained variation. How does this happen and how can it be interpreted?

Negative eigenvalues can occur when the resemblance measure used does not fulfill the four mathematical properties required for it to be classified as a metric distance measure. These are:

1. The minimum distance is zero: if point A and point B are identical, then $d_{AB} = 0$.
2. All distances are non-negative: if point A and point B are not identical, then $d_{AB} > 0$.
3. Symmetry: the distance from A to B is equal to the distance from B to A: $d_{AB} = d_{BA}$.
4. The triangle inequality: $d_{AB} \leq (d_{AC} + d_{BC})$.

Almost all measures worth using will fulfill at least the first three of these properties. A dissimilarity measure which fulfills the first three of the above properties, but not the fourth, is called semi-metric. Most of the resemblance measures of greatest interest to ecologists are either semi-metric (such as Bray-Curtis) or, although metric, are not necessarily embeddable in Euclidean space (such as Jaccard or Manhattan) and so negative eigenvalues can still be produced by the PCO.

---

66 For a review of the mathematical properties of many commonly used measures, see Gower & Legendre (1986) and chapter 7 of Legendre & Legendre (1998).
The fourth property, known as the triangle inequality states that the distance between two points (A and B, say) is equal to or smaller than the distance from A to B via some other point (C, say). How does violation of the triangle inequality produce negative eigenvalues? Perhaps the best way to tackle this question is to step through an example.

![Fig. 3.4. Demonstration of violation of the triangle inequality and how PCO generates a Euclidean solution in two dimensions; the second dimension is imaginary.](image)

Consider a resemblance measure calculated among three points: A, B and C, which produces the dissimilarity matrix shown in Fig. 3.4a. Here, the triangle inequality is clearly violated. Consequently, there is no way to place these three points into Euclidean space in two real dimensions in such a manner which preserves all three original dissimilarities. Point C should lie at a distance of 0.1 from point A (i.e., it may lie at any position along the circumference of a circle of radius 0.1 whose centre is at point A, such as C1). However, point C should also lie at a distance of 0.2 from point B (such as at point C2). Clearly, point C cannot fulfil both of these criteria simultaneously – it cannot be in two places at once! Well, we have to put point C somewhere in order to draw this configuration in Euclidean space (i.e., on the page) at all. Suppose point C is given a position along a single dimension between points A and B, as shown in Fig. 3.4b. This could be anywhere, perhaps, along the straight-line continuum from A to B, but it would make sense for it to lie somewhere in the “gap” between the two circles. If we had done a PCO of the original dissimilarity matrix, the position for point C shown in Fig. 3.4b is precisely that which is given for it along the first PCO axis.

Clearly, in order to draw this ordination (in one dimension), we have had to “stretch” the distances from A to C and from B to C, just so that C could be represented as a single point in the diagram. A consequence of this is that the total variation (the sum of squared inter-point distances divided by the number of points, see Figs. 1.3 and 1.4 in chapter 1) is inflated. That is, the original total sum of squares was $SS_T = 0.07$ (Fig. 3.4a), and now, even using only one dimension, the total sum of squares is actually larger than this ($SS_{PCO1} = 0.08$, Fig. 3.4b). We can remove this “extra” variance, introduced by placing the points from a semi-metric system into Euclidean space, by introducing one or more imaginary axes. Distances along an imaginary axis can be treated effectively in the same manner as those along real axes, except they need to be multiplied by the constant $i = \sqrt{-1}$. Consequently, squared values of distances along an imaginary axis are multiplied by $i^2 = -1$, and, therefore, are simply negative. In our example, the distances among the three points obtained along the second PCO axis (which is imaginary) are shown in Fig. 3.4c. In practice, the configuration output treats the second axis like any other axis (so that it can be drawn), but we need to keep in mind that this second axis (theoretically and algebraically) is actually defined only in imaginary
The total sum of squares along this second axis alone is $SS_{PCO2} = -0.01$, and so the variance occurring along this imaginary axis is negative. Despite this, we can see in this example that $SS_T = SS_{PCO1} + SS_{PCO2}$. It is true in this example, and it is true in general, that the sum of the individual SS for each of the PCO’s will add up to the total SS from the original dissimilarities provided we take due care when combining real and imaginary axes – the SS of the latter contribute negatively. In this example, there is one positive and one negative eigenvalue, corresponding to PCO1 and PCO2, respectively. The sum of these two eigenvalues, retaining their sign, is equal to the total variance.

Returning now to the Victorian avifauna, we can plot the percentage of the variation explained by each of the PCO axes, as given in the output in Fig. 3.2. This is known as a scree plot (Fig. 3.5), and it provides a visual diagnostic regarding the number of axes that capture the majority of variation for this system. For this example, there are a series of positive eigenvalues which are progressively smaller in size, followed by a series of negative eigenvalues (Fig. 3.5). We can also plot these percentages cumulatively (Fig. 3.5) and see that the percentage of explained variation goes up past 100% when we get to PCO axis 5. The first 9 PCO axes together actually explain ~106% of the original variation (Fig. 3.2). Once again, the reason for this is that the analysis has had to “make do” and “stretch” some of the original distances (for which the triangle inequality does not hold) in order to represent these points on real Euclidean axes. Then, with the addition of subsequent PCO axes (10 through 15) corresponding to negative eigenvalues (imaginary Euclidean axes), this is brought back down to precisely 100% of the original variation based on adjusted Bray-Curtis resemblances.

For purposes of ordination, our interest will be in plotting and visualising patterns using just the first two or three axes. Generally, these first few axes will correspond to large positive eigenvalues, the axes corresponding to negative eigenvalues will be negligible (i.e., in the “tail” of the scree plot) and need not be of any concern (Sibson 1979, Gower 1987). However, if:

- the first two or three PCO axes together explain > 100% of the variation,
- any of the plotted PCO axes are associated with a negative eigenvalue (i.e. corresponding to an imaginary axis), or
- the largest negative eigenvalue in the system as a whole is larger (in absolute value) than the smallest positive eigenvalue associated with the PCO axes in the plot (Caillez & Pagès 1976, as cited by Legendre & Legendre 1998),

then interpreting the plot will be problematic. In any of these cases, one would be dealing with a resemblance matrix that rather seriously violates the triangle inequality. This can occur if the resemblance measure being used is inappropriate for the type of data, if there are a lot of repeated values in the resemblance matrix or if the resemblances are ranked prior to analysis. With the PCO routine, as with PERMANOVA or PERMDISP, it makes sense to carefully consider the meaning

---

67 Indeed, for any other purposes apart from attempts to physically draw a configuration, PRIMER and the PERMANOVA+ add-on will indeed treat such axes, algebraically, as imaginary, and so maintain the correct sign of the associated eigenvalue.
of the resemblance measure used in the context of the data to be analysed and the questions of interest.

Perhaps not surprisingly, previous workers have been troubled by the presence of negative eigenvalues from PCO analyses and how they should be interpreted (e.g., see pp. 432-438 in Legendre & Legendre 1998). One possibility is to “correct” for negative eigenvalues by adding a constant to all of the dissimilarities (Cailliez 1983) or to all of their squares (Lingoes 1971). Clearly, these correction methods inflate the total variation (e.g., Legendre & Anderson 1999) and are not actually necessary (McArdle & Anderson 2001). Cailliez & Pagès (1976) also suggested correcting the percentage of the variation among $N$ points explained by (say) the first $m$ PCO axes from what is used above, i.e.:

$$100 \times \frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{N} \lambda_i}$$

(3.1)

to:

$$100 \times \frac{\sum_{i=1}^{m} \lambda_i + m|\lambda_N|}{\sum_{i=1}^{N} \lambda_i + (N-1)|\lambda_N|}$$

(3.2)

where $|\lambda_N|$ is the absolute value of the largest negative eigenvalue (see p. 438 in Legendre & Legendre 1998). Although it was suggested that (3.2) would provide a better measure of the quality of the ordination in the presence of negative eigenvalues, this is not entirely clear. The approach in (3.2), like the proposed correction methods, inflates the total variation. Alternatively, Gower (1987) has indicated that the adequacy of the ordination can be assessed by calculating percentages using the squares of the eigenvalues, when negative eigenvalues are present. In the PCO routine of the PERMANOVA+ package, we chose simply to retain the former calculation (3.1) in the output, so that the user may examine the complete original diagnostic information regarding the percentage of variation explained by each of the real and imaginary axes, individually and cumulatively (e.g., Figs. 3.2 and 3.5).

An important take-home message is that the PCO axes together do explain 100% of the variation in the original dissimilarity matrix (Fig. 3.4), provided we are careful to retain the signs of the eigenvalues and to treat those axes associated with negative eigenvalues as imaginary. The PCO axes therefore (when taken together) provide a full Euclidean representation of the dissimilarity matrix, albeit requiring both real and complex (imaginary) axes for some (semi-metric) dissimilarity measures. As a consequence of this property, note also that the original dissimilarity between any two samples can be obtained from the PCO axes. More particularly, the square root of the sum of the squared distances between two points along the PCO axes\(^{68}\) is equal to the original dissimilarity for those two points. For example, consider the Euclidean distance between points A and C in Fig. 3.4 that would be obtained using the PCO axes alone:

$$d_{AC} = \sqrt{(d_{AC}^{PCO})^2 + (d_{AC}^{PCO2})^2} = \sqrt{(0.167)^2 + (0.134)^2}.$$  

As $i^2 = -1$, this gives

$$d_{AC} = \sqrt{(0.167)^2 - (0.134)^2} = 0.100$$

which is the original dissimilarity between points A and C in Fig. 3.4a. In other words, we can recreate the original dissimilarities by calculating Euclidean distances on the PCO scores (Gower 1966). In order to get this exactly right, however, we have to use all of the PCO axes and we have to treat the imaginary axes correctly and separately under the square root sign (e.g., McArdle & Anderson 2001).

A new feature of the PERMANOVA+ add-on package is the ability to add vector overlays onto graphical outputs. This is offered as purely an exploratory tool to visualise potential linear or monotonic relationships between a given set of variables and ordination axes. For example, returning to the Victorian avifauna data, we may wish to know which of the original species

---

\(^{68}\) (i.e., the Euclidean distance between two points calculated from the PCO scores)
variables are either increasing or decreasing in value from left to right across the PCO diagram. These would be bird species whose abundances correlate with the differences seen between the ‘good’ and the ‘poor’ sites, which are clearly split along PCO axis 1 (Fig. 3.3). From the PCO plot, choose **Graph > Special** to obtain the ‘Configuration Plot’ dialog box (Fig. 3.6), then, under ‘Vectors’ choose /*Worksheet variables: vic.good.poor */ & **(Correlation type: Spearman)**, click on the ‘Select…’ button and choose (Select Vectors •Correlation > 0.5), OK.

This produces a vector overlay onto the PCO plot as shown in Fig. 3.7. In this case, we have restricted the overlay to include only those variables from the worksheet that have a vector length that is greater than 0.5. Alternatively, Pearson correlations may be used instead. These will specifically highlight linear relationships, whereas Spearman correlations are a bit more flexible, being based on ranks, and so will highlight more simply the overall increasing or decreasing relationships of individual variables across the plot. The primary features of the vector overlay are:

- The circle is a unit circle (radius = 1.0), whose relative size and position of origin (centre) is arbitrary with respect to the underlying plot.
- Each vector begins at the centre of the circle (the origin) and ends at the coordinates \((x, y)\) consisting of the correlations between that variable and each of PCO axis 1 and 2, respectively.
- The length and direction of each vector indicates the strength and sign, respectively, of the relationship between that variable and the PCO axes.

For the Victorian avifauna, we can see that the abundance of Red Wattlebird has a strong negative relationship with PCO 1 (indicative of ‘good’ sites), while the abundance of Golden Whistler has a fairly strong positive relationship with this axis (indicative of ‘poor’ sites). These two species have very weak relationships with PCO 2. There are other species that are correlated with PCO 2, (either positively or negatively), which largely separates the two ‘poor’ sites from one another (Fig. 3.7).

By clicking on the ‘Correlations to worksheet’ button in the ‘Configuration Plot’ dialog box (Fig. 3.6), individual correlations between each variable in the selected worksheet and the PCO axes are output to a new worksheet where they can be considered individually, exported to other programs,
or analysed in other ways within PRIMER\(^69\). Examining the Spearman correlations given in a worksheet for the Victorian avifauna data helps to clarify how the axes were drawn (Fig. 3.8). For example, the Yellow-plumed Honeyeater has a correlation of \(\rho_1 = 0.376\) with PCO axis 1 and \(\rho_2 = 0.564\) with PCO axis 2. The length of the vector for that species is therefore \(\ell = \sqrt{\rho_1^2 + \rho_2^2} = 0.678\) and it occurs in the upper-right quadrant of the circle, as both correlations are positive (Fig. 3.8). The correlations and associated vectors are also shown for the Yellow-tufted Honeyeater and the Buff-rumped Thornbill (Fig. 3.8).

**Fig. 3.7.** Vector overlay on the PCO of the Victorian avifauna, showing birds with vectors longer than 0.5.

**Fig. 3.8.** Spearman correlations in datasheet and schematic diagram of calculations used to produce vector overlays on the PCO of the Victorian avifauna data.

There are several important caveats on the use and interpretation of these vector overlays. First, just because a variable has a long vector when drawn on an ordination in this way does not confirm that

\[^69\] Beware of the fact that if you choose to reflect positions of points along axes by changing their sign (i.e., if you choose Graph > Flip X or Flip Y), the signs of the correlations given in the worksheet will no longer correspond to those shown in the diagram!
this variable is necessarily responsible for differences among samples or groups in that direction. These are correlations only and therefore they cannot be construed to indicate causation of either the effects of factors or of dissimilarities between individual sample points. Second, just because a given variable has a short vector when drawn on an ordination in this way does not mean that this variable is unimportant with respect to patterns that might be apparent in the diagram. Pearson correlations will show linear relationships with axes, Spearman rank correlations will show monotonic increasing or decreasing relationships with axes, but neither will show Gaussian, unimodal or multi-modal relationships well at all. Yet these kinds of relationships are very common indeed for ecological species abundance data, especially for a series of sites along one or more environmental gradients (ter Braak 1985, Zhu et al. 2005, Yee 2006). Bubble plots, which are also available within PRIMER, can be used to explore these more complex relationships (see chapter 7 in Clarke & Gorley 2006).

It is best to view these vector overlays as simply an exploratory tool. They do not mean that the variables do or do not have linear relationships with the axes (except in special cases, see the section PCO vs PCA below). For the above example, the split in the data between ‘good’ and ‘poor’ indicates a clear role for PCO axis 1, so seeking variables with increasing or decreasing relationships with this axis (via Spearman raw correlations) is fairly reasonable here. For ordinations that have more complex patterns and gradients, however, the vector overlays may do a poor job of uncovering the variables that are relevant in structuring multivariate variation.

Principal components analysis (PCA) is described in detail in chapter 4 of Clarke & Warwick (2001). As stated earlier, PCO produces an equivalent ordination to a PCA when a Euclidean distance matrix is used as the basis of the analysis. Consider the environmental data from the Firth of Clyde (Pearson & Blackstock 1984), as analysed using the PCA routine in PRIMER in chapter 10 of Clarke & Gorley (2006). These data consist of 11 environmental variables, including: depth, the concentrations of several heavy metals, percent carbon and percent nitrogen found in soft sediments from benthic grabs at each of 12 sites along an east-west transect that spans the Garroch Head sewage-sludge dumpground, with site 6 (in the middle) being closest to the actual dump site. The data are located in the file clev.pri in the ‘Clydemac’ folder of the ‘Examples v6’ directory.

As indicated by Clarke & Gorley (2006), it is appropriate to first check the distributions of variables (e.g., for skewness and outliers) before proceeding with a PCA. This can be done, for example, by choosing Analyse>Draftsman Plot in PRIMER. As recommended for these data by Clarke & Gorley (2006), log-transform all of the variables except depth. This is done by...
3. PCO

highlighting all of the variables except depth and choosing **Tools > Transform(individual)** > **Expression**: \(\log(0.1+V)\) > OK. Rename the transformed variables, as appropriate (e.g. Cu can be renamed \(\ln \text{Cu}\), and so on), and then rename the data sheet of transformed data elevt. Clearly, these variables are on quite different measurement scales and units, so a PCA on the correlation matrix is appropriate here. Normalise the transformed data by choosing **Analyse > Pre-treatment > Normalise variables**. Rename the data sheet of normalised variables elevtn. Finally, do a PCA of the normalised data by choosing **Analyse > PCA**.

Recall that PCA is simply a centred rotation of the original axes and that the resulting PC axes are therefore linear combinations of the original (in this case, transformed and normalised) variables. The ‘Eigenvectors’ in the output file from a PCA in PRIMER provide explicitly the coefficients associated with each variable in the linear combination that gives rise to each PC axis. Importantly, the vectors shown on the PCA plot are these eigenvectors, giving specific information about these linear combinations. This means, necessarily, that the lengths and positions of the vectors depend on the other variables included in the analysis.

![PCA](image1)

**Fig. 3.10.** Results of PCA and PCO on Euclidean distances for the transformed and normalised environmental data from the Firth of Clyde.

Next we may replicate the pattern of points seen in the PCA ordination precisely by doing a PCO on the basis of a Euclidean distance matrix. From the elevtn data file, choose **Analyse > Resemblance** > **(Analyse between •Samples) & (Measure •Euclidean distance)** > OK. From the resulting resemblance matrix, choose **PERMANOVA+ > PCO** > OK. For PCO, as for PCA, the sign of axes is arbitrary. It may be necessary to choose **Graph > Flip X** and/or **Graph > Flip Y** before the orientation of the PCO plot can be seen to match that of the PCA. Once this is done, however, it should be readily apparent that the patterns of points and even the scaling of axes are

\[70\] This is directly analogous to the fact that the regression coefficient for a variable \(X_1\) in the simple regression of \(Y\) vs \(X_1\), will be different from the partial regression coefficient in the multiple regression of \(Y\) vs \(X_1\) and \(X_2\) together. That is, the relationship between \(Y\) and \(X_1\) changes once \(X_2\) is taken into account. Correlation (non-independence) between \(X_1\) and \(X_2\) is what causes simple and partial regression coefficients to differ from one another.
identical for these two analyses (Fig. 3.9). Further proof of the equivalence is seen by examining the % variation explained by each of the axes, as provided in the text output files (Fig. 3.10). The first two axes alone explain over 89% of the variation in these 11 environmental variables, indicating that the 2-d ordination is highly likely to have captured the majority of the salient patterns of variation in the multi-dimensional data cloud.

By choosing Graph > Special > (Vectors ●Worksheet variables: elevtn > Correlation type: Pearson), vectors are drawn onto the PCO plot that correspond to the Pearson correlations of individual variables with the PCO axes. These are clearly different from the eigenvectors that are drawn by default on the PCA plot. For the vector overlay on the PCO plot, each variable was treated separately and the vectors arise from individual correlations that do not take into account any of the other variables. In contrast, the eigenvectors of the PCA do take into account the other variables in the analysis, and would obviously change if one or more of the variables were omitted. It is not surprising, in the present example, that log-concentrations of many of the heavy metals are strongly correlated with the first PCO axis, showing a gradient of decreasing contamination with increasing distance in either direction away from the dumpsite (site 6).

The PCA eigenvectors are not given in the output from a Euclidean PCO analysis. This is because the PCO uses the Euclidean distance matrix alone as its starting point, so has no “memory” of the variables which gave rise to these distances. We can, however, obtain these eigenvectors retrospectively, as the resulting PCO axes in this case coincide with what would be obtained by running a PCA – they are therefore linear combinations of the original (transformed and normalised) variables. This is done by choosing Graph > Special > (Vectors ●Worksheet variables: elevtn > Correlation type: Multiple). What this option does is to overlay vectors that show the multiple partial correlations of the variables in the chosen worksheet with the configuration axes. In this case, these correspond to the eigenvectors. If you click on the option ‘Correlations to worksheet’ in the ‘Configuration plot’ dialog, you will see in this worksheet that the values for each variable correspond indeed to their eigenvector values as provided in the PCA output file. This equivalence will hold provided variables have been normalised prior to analysis.

More generally, the relevant point here is that the choice of ‘Correlation type: Multiple’ will produce a vector overlay of multiple partial correlations, where the relationships between variables and configuration axes take into account other variables in the worksheet. This contrasts with the choice of ‘Correlation type: Pearson’ (or Spearman), which plot raw correlations for each variable, ignoring all other variables. For the PCA, note that it is also possible to replace the eigenvectors of original variables that are displayed by default on the ordination with an overlay of some other vectors of interest. This is done by simply choosing ‘●Worksheet variables’ and indicating the worksheet that contains the variables of interest, instead of ‘●Base variables’ in the ‘Configuration Plot’ dialog.

A further point to note is the fact that there are no negative eigenvalues in this example. Indeed, the cumulative scree plot for either a PCA or a PCO based on Euclidean distances will simply be a smooth increasing function from zero to 100%. Any measure that is Euclidean-embeddable and fulfils the triangle inequality will have strictly non-negative eigenvalues in the PCO and thus will have all real and no imaginary axes (Torgerson 1958, Gower 1982, Gower & Legendre 1986).

If PCO is done on a resemblance matrix obtained using the chi-squared distance measure (measure D16 under the ‘Distance’ option under the ‘More’ tab of the ‘Resemblance’ dialog), then the resulting ordination will be identical to a correspondence analysis (CA) among samples that has been drawn using scaling method 1 (see p. 456 and p. 467 of Legendre & Legendre 1998 for details). Although the relationship between the original variables and the resulting ordination axes is linear for PCO on Euclidean distances (a.k.a. PCA) and unimodal for PCO on chi-squared distances (a.k.a. CA), when PCO is based on some other measure, such as Bray-Curtis, these relationships are likely to be highly non-linear and are generally unknown. Clearly, PCO is more general and flexible than either PCA or CA. This added flexibility comes at a price, however. Like MDS, PCO necessarily must lose any direct link with the original variables by being based purely

71 The sign associated with the values given in this file and given for the eigenvectors will depend, of course, on the sign of the PCO and PCA axes, respectively, that happen to be provided in the output. As these are arbitrary, the signs may need to be “flipped” to see the concordance.
on the resemblance matrix. Relationships between the PCO axes and the original variables (or any other variables for that matter) can only be investigated retrospectively (e.g., using vector overlays or bubbles), as seen in the above examples.

In chapter 1, the difficulty in calculating centroids for non-Euclidean resemblance measures was discussed (see the section entitled Huygens’ theorem). In chapter 2, the section entitled Generalisation to dissimilarities indicated how PCO axes could be used in order to calculate distances from individual points to their group centroids in the space of a chosen resemblance measure as part of the PERMDISP routine. Furthermore (see the section entitled Dispersion in nested designs in chapter 2), it was shown how a new tool available in PERMANOVA+ can be used to calculate distances among centroids in the space of a chosen resemblance measure. One important use of this tool is to provide insights into the relative sizes and directions of effects in complex experimental designs. Specifically, once distances among centroids from the cells or combinations of levels of an experimental design have been calculated, these can be visualised using PCO (or MDS). The results will provide a suitable visual complement to the output given in PERMANOVA regarding the relative sizes of components of variation in the experimental design and will also clarify the relative distances among centroids for individual levels of factors.

For example, we shall consider again the data on macrofauna from the Okura estuary (Anderson et al. 2004), found in the file okura.pri of the ‘Okura’ folder in the ‘Examples add-on’ directory. Previously, we considered only the first time of sampling. Now, however, we shall consider the full data set, which included 6 times of sampling. The full experimental design included four factors:

Factor A: Season (fixed with $a = 3$ levels: winter/spring (W.S), spring/summer (S.S) or late summer (L.S)).

Factor B: Rain.Dry (fixed with $b = 2$ levels: after rain (R) or after a dry period (D)).

Factor C: Deposition (fixed with $c = 3$ levels: high (H), medium (M) or low (L) probability of sediment deposition).

Factor D: Site (random with $d = 5$ levels, nested in the factor Deposition).

There were $n = 6$ sediment cores collected at each site at each time of sampling.

A PERMANOVA analysis based on the Bray-Curtis resemblance measure calculated from log(X+1)-transformed variables yielded quantitative measures for the components of variation...
associated with each term in the model (Fig. 3.11). Focusing on the main effects for the non-nested factors only, we can see that the greatest effect size was attributable to Deposition, followed by Season, followed by Rain.Dry. To visualise the variation among the relevant cell centroids in this design, we begin by identifying the cells that correspond to what we wish to plot as levels of a single factor. In this case, we wish to examine the cells corresponding to all combinations of Season × Rain.Dry × Deposition. From the full resemblance matrix, choose Edit > Factors > Combine, then click on each of these main effect terms in turn, followed by the right arrow in order to move them over into the ‘Include’ box on the right, then click ‘OK’. Next, to calculate distances among these centroids, choose PERMANOVA+ > Distances among centroids… > Grouping factor: SeasonRain.DryDeposition > OK (Fig. 3.12). The resulting resemblance matrix will contain distances (in Bray-Curtis space, as that is what formed the basis of the analysis here) among the $a \times b \times c = 3 \times 2 \times 3 = 18$ cells. Note that the names of the samples (centroids) in this newly created resemblance matrix will be the same as the levels of the factor that was used to identify the cells. Whereas the original resemblance matrix had 540 samples, we have now obtained Bray-Curtis resemblances among 18 centroids, each calculated as an average (using PCO axes, not the raw data mind!) from $d \times n = 5 \times 6 = 30$ cores.

From the distance matrix among centroids (‘Resem2’), choose PERMANOVA+ > PCO and examine the resulting ordination of centroids (Fig. 3.13). The first PCO axis explains the majority of the variation among these cells (nearly 75%) and is strongly associated with the separation of assemblages in high depositional environments (on the right) from those in low or medium depositional environments (on the left). Of far less importance (as was also shown in the PERMANOVA output) is the seasonal factor. Differences among seasons are discernible along PCO axis 2, explaining less than 8% of the variation among cells and with an apparent progression of change in assemblage structure from winter/spring, through spring/summer and then to late summer/autumn.

---

72 See chapter 1 for a more complete description of how to use the PERMANOVA routine to analyse complex experimental designs.
summer from the top to the bottom of the PCO plot (Fig. 3.13). Of even less importance is the effect of rainfall – the centroids corresponding to the two sampling times within each season and depositional environment occur quite close together on the plot, further indicating that, for these data (and based on the Bray-Curtis measure), seasonal and depositional effects were of greater importance than rainfall effects, which were negligible.

Interactions among these factors did not contribute large amounts of variation, but they were present and some were statistically significant (see the PERMANOVA output in Fig. 3.11). The seasonal effects appeared to be stronger for the high depositional environments than for either of the medium or low depositional environments, according to the plot, so it is not surprising that the Season×Deposition term is sizable in the PERMANOVA output as well. Similarly, the Season×Rainfall interaction term contributes a reasonable amount to the overall variation, and the plot of centroids suggests that rainfall effects (i.e. the distances between R and D) were more substantial in winter/spring than in late/summer. Of course, ordination plots of appropriate subsets of the data and relevant pair-wise comparisons can be used to further elucidate and interpret significant interactions.

This example demonstrates that ordination plots of distances among centroids can be very useful in unraveling patterns among levels of factors in complex designs. The new tool Distances among centroids in PERMANOVA+ uses PCO axes to calculate these centroids, retaining the necessary distinctions between sets of axes that correspond to positive and negative eigenvalues, respectively, and so maintaining the multivariate structure identified by the choice of resemblance measure as the basis for the analysis as a whole.

**Fig. 3.13.** PCO of distances among centroids on the basis of the Bray-Curtis measure of log(X+1)-transformed abundances for Okura macrofauna, first with ‘Deposition’ symbols and ‘Season’ labels (top panel) and then with ‘Season’ symbols and ‘Rain.Dry’ labels (bottom panel).

Importantly, an analysis that proceeds instead by first calculating centroids (averages) from the raw data first, followed by the rest (i.e. calculating the transformation and Bray-Curtis resemblances on these averages) would not provide the same results. Patterns from the latter would also not necessarily accord with the relative sizes of components of variation from a PERMANOVA
partitioning that had been performed on the full set of data. We recommend that the decision to either sum or average the raw data before analysis should be driven by an \textit{a priori} judgment regarding the appropriate scale of observation of the communities of interest. For example, in some cases, the individual replicates are too small and too highly variable in composition to be considered representative samples of the communities of interest. In such cases, pooling together (summing or averaging) small-scale replicates to obtain an appropriate sample unit for a given spatial scale \textit{before} performing a PERMANOVA (or other) analysis may indeed be appropriate.\textsuperscript{73} The tool provided here to calculate distances among centroids, instead, assists in understanding the partitioning of the variation in the multivariate space identified by the resemblance measure, well after a decision regarding what should comprise an appropriate lowest-level representative sampling unit has been made.

We recommend that, for routine ordination to visualise multivariate data on the basis of a chosen resemblance measure, non-metric MDS is the method of choice (Kruskal 1964; Minchin 1987). MDS is robust to outliers, and it explicitly uses a criterion of preserving rank orders of dissimilarities in a reduced number of dimensions, so has excellent distance-preserving properties.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{Okura_estuary_2001-2002_MDS.png}
\caption{MDS of distances among centroids on the basis of the Bray-Curtis measure of log(X+1)-transformed abundances for Okura macrofauna (top panel) and associated Shepard diagram (bottom panel).}
\end{figure}

In practice, PCO and MDS will tend to give similar results for a given resemblance matrix. Generally, far more important to the resulting patterns seen in the ordination will be the decisions made regarding the choice of transformation, standardisation and resemblance measure. Trials with a few examples will demonstrate this and are left to the reader to explore. There are, however, a few notable exceptions. Differences between MDS and PCO will be sharpest when there is a large

\textsuperscript{73} An example where this was done was the Norwegian macrofauna (norbio.pri in the ‘NorMac’ folder), where 5 benthic grabs at a site were pooled together and considered as a single sampling unit for analysis.
split between groups of one or more samples in the multivariate space. In such cases, MDS can yield what is called a “degenerate” solution (see chapter 5 in Clarke & Warwick 2001), where all of the points within a group are tightly clustered or collapsed onto a single point in the MDS configuration. This occurs when all of the “within-group” dissimilarities are smaller than all of the “between-group” dissimilarities. As pointed out by Clarke & Warwick (2001), in such cases “there is clearly no yardstick within our non-parametric approach for determining how far apart the groups should be placed in the MDS plot”. However, by choosing to use PCO in such cases, we are provided with just such a yardstick.

A case in point is the resemblance matrix among centroids analysed using PCO in Fig. 3.13 for the macrofauna from Okura estuary. An MDS plot of this resemblance matrix, and the associated Shepard plot, highlighting the two disjunctive sets of “within-group” and “between-group” dissimilarities, is shown in Fig. 3.14. Clearly, little or no information about the actual relative positions of these centroids can be gained from the MDS plot. The usual solution suggested for dealing with such cases is to carry out separate ordinations on each of the two groups. However, if our interest lies in visualising (as well as we can in a reduced number of dimensions, that is) the relative positions of the full set of centroids in the higher-dimensional space, especially to help us understand the relative quantitative differences among levels and associated effect sizes for factors, the MDS approach may let us down. In this case, it is not possible to relate the information or patterns in the MDS plot to the PERMANOVA output for this experimental design, and splitting the data into pieces will not particularly help here. The PCO routine does a much better job (Fig. 3.13).

As an aside, PCO is not only different from non-metric MDS, it also differs from what might generally be referred to as metric MDS. Both metric and nonmetric MDS encompass many methods (see Gower & Hand 1996), but their main focus is to minimise the criterion known as stress, a measure of how well the distances in the Euclidean configuration represent the original dissimilarities. Whereas the non-metric algorithms minimise stress as a monotonic function of the dissimilarities, their metric counterparts minimise stress using a linear or least-squares type of approach. Metric methods are also sometimes called least-squares scaling. Minimising a nonmetric stress criterion with a linear constraint is the same as minimising metric stress, though neither is equivalent to PCO. The point here is that MDS (metric or non-metric) is focused purely on preserving dissimilarities or distances in the configuration for a given number of dimensions, whereas PCO is a projection from the space of the resemblance measure onto Euclidean axes. The success of that projection, with respect to preserving dissimilarities, will therefore depend somewhat on just how high-dimensional the underlying data are and how ‘non-Euclidean’ the original resemblance measure is.

The strength of non-metric MDS lies in its flexible ‘stretching and squeezing’ of the resemblance scale, for example as dissimilarities push up against their upper limit of 100% (communities with no species in common). This focus on preserving rank-order relationships will generally give more sensible descriptions, e.g. of long-baseline gradients, in low dimensions than can be obtained by PCO. (The reader is encouraged to try out the comparison for some of the well-known data sets in the PRIMER ‘Examples v6’ directory, such as the macrofauna data in the Clydemac directory, the study met in Fig. 3.9). Paradoxically, however, the strength of PCO in the PERMANOVA context is precisely that it does not ‘stretch and squeeze’ the dissimilarity scale, so that where a low-dimensional PCO plot is able to capture the high-dimensional structure adequately (as reflected in the % variation explained), it is likely to give a closer reflection of the resemblance values actually used in the partitioning methods such as PERMANOVA and PERMDISP.
4. Distance-based linear models (DISTLM) and distance-based redundancy analysis (dbRDA)

Key references

General description
DISTLM is a routine for analysing and modeling the relationship between a multivariate data cloud, as described by a resemblance matrix, and one or more predictor variables. For example, in ecology, the resemblance matrix commonly describes dissimilarities (or similarities) among a set of samples on the basis of multivariate species abundance data, and interest may lie in determining the relationship between this data cloud and one or more environmental variables that were measured for the same set of samples. The routine allows predictor variables to be fit individually or together in specified sets. P-values for testing the null hypothesis of no relationship (either for individual variables alone or conditional on other variables) are obtained using appropriate permutation methods. Not only does DISTLM provide quantitative measures and tests of the variation explained by one or more predictor variables, the new routine in PERMANOVA+ has a suite of new tools for building models and generating hypotheses. Parsimonious models can be built using a choice of model selection criteria and procedures. Coupled with preliminary diagnostics to assess multi-collinearity among predictor variables, several potentially relevant models can be explored. Finally, for a given model, the user may also visualise the fitted model in multi-dimensional space, using the dbRDA routine.

Rationale
Just as PERMANOVA does a partitioning of variation in a data cloud described by a resemblance matrix according to an ANOVA model, DISTLM does just such a partitioning, but according to a regression (or multiple regression) model. For an ANOVA design, the predictor variables are categorical (i.e., levels of factors), whereas in a regression model, the predictor variables are (generally) quantitative and continuous. ANOVA is simply a specific kind of linear model, so DISTLM can therefore also be used to analyse models that contain a mixture of categorical and continuous predictor variables.

The approach implemented by DISTLM is called distance-based redundancy analysis (dbRDA), which was first coined by Legendre & Anderson (1999) and later refined by McArdle & Anderson (2001). Legendre & Anderson (1999) described dbRDA as a multivariate multiple regression of PCO axes on predictor variables. They included a correction for negative eigenvalues for situations when these occurred (see chapter 3 regarding negative eigenvalues and how they can arise in PCO). McArdle & Anderson (2001) refined this idea to provide a more direct approach which is the method now implemented by DISTLM and described here. It does not require PCO axes to be calculated, nor does it require any corrections for negative eigenvalues. These two approaches are equivalent for situations where no negative eigenvalues would arise in a PCO of the resemblance matrix being analysed.

Both of the PERMANOVA+ routines discussed in this chapter (DISTLM and dbRDA) actually do distance-based redundancy analysis. The DISTLM routine is used to perform partitioning, test hypotheses and build models, while the dbRDA routine is used to perform an ordination of fitted values from a given model. In the dbRDA routine, the structure of the data cloud is viewed through the eyes of the model (so-to-speak) by doing an eigen-analysis of the fitted data cloud. While a PCO on the original resemblance matrix is an unconstrained ordination (because the resemblance matrix alone is examined, free of any specific model or hypothesis), dbRDA is constrained to find linear combinations of the predictor variables which explain the greatest variation in the data cloud.

Partitioning
Consider an \((N \times p)\) matrix of response variables \(Y\), where \(N\) = the number of samples and \(p\) = the number of variables. Consider also an \((N \times q)\) matrix, \(X\), which contains \(q\) explanatory (predictor) variables in DISTLM models are always treated as fixed, so use the PERMANOVA routine instead when dealing with random factors. Quantitative continuous variables can be included in a PERMANOVA model by treating them as covariables (see chapter 1).
variables of interest (e.g., environmental variables). The purpose of DISTLM is to perform a permutational test for the multivariate null hypothesis of no relationship between matrices Y and X on the basis of a chosen resemblance measure, using permutations of the samples to obtain a P-value. In essence, the purpose here is to ask the question: does X explain a significant proportion of the multivariate variation in the data cloud described by the resemblance matrix obtained from Y (Fig. 4.1)? Note that the analysis done by DISTLM is directional and that these sets of variables have particular roles. The variables in X are being used to explain, model or predict variability in the data cloud described by the resemblance matrix arising from Y.

![Fig. 4.1. Conceptual diagram of regression as a partitioning of the total variation into a portion that is explained by the predictor variables in matrix X, and a portion that is left unexplained (the residual).](image)

Details of how dbRDA is done are provided by Legendre & Anderson (1999) and McArdle & Anderson (2001). Maintaining a clear view of the conceptual framework is what matters most (Fig. 4.1), but a thumbnail sketch of the matrix algebra involved in the mechanics of dbRDA is also provided here. Suppose that D is an N × N matrix of dissimilarities (or distances) among samples. The analysis proceeds through the following steps (Fig. 4.2):

1. From the matrix of dissimilarities, D, calculate matrix A, then Gower’s centred matrix G, as outlined in steps 1 and 2 for doing a PCO (see Fig. 3.1);
2. The total sum of squares (SS_total) of the full multivariate data cloud is the sum of the diagonal elements (called the “trace” and symbolised here by “tr”) of matrix G, that is, tr(G);
3. From matrix X, calculate the “hat” matrix H = X'X'. This is the matrix derived from the solutions to the normal equations ordinarily used in multiple regression (e.g., Johnson & Wichern 1992, Neter et al. 1996);
4. The explained sum of squares for the regression (SS_regression) is then calculated directly as tr(HGH);
5. The unexplained (or residual) sum of squares is then SS_residual = SS_total – SS_regression. This is also able to be calculated directly as tr([I – H]G[I – H]), where I is an N × N identity matrix.

Once the partitioning has been done, the proportion of the variation in the multivariate data cloud that is explained by the variables in X is calculated as:

$$R^2 = \frac{SS_{\text{Regression}}}{SS_{\text{Total}}}$$

(4.1)

Furthermore, an appropriate statistic for testing the general null hypothesis of no relationship is:

$$F = \frac{SS_{\text{Regression}}/q}{SS_{\text{Residual}}/(N – q – 1)}$$

(4.2)

This is the pseudo-F statistic, as already seen for the ANOVA case in chapter 1 (equation 1.3). It is a direct multivariate analogue of Fisher’s F ratio used in traditional regression. However, when using DISTLM, we do not presume to know the distribution of pseudo-F in equation 4.2, especially for p > 1 and where a non-Euclidean dissimilarity measure has been used as the basis of the

---

75 If a resemblance matrix of similarities is available instead, then DISTLM in PERMANOVA+ will automatically transform these into dissimilarities; the user need not do this as a separate step.

76 Regression models usually include an intercept term. This is obtained by including a single column of 1’s as the first column in matrix X. DISTLM automatically includes an intercept in all of its models. This amounts to centreing the data before analysis and has no effect on the calculations of explained variation.

77 An identity matrix (I) behaves algebraically like a “1” in matrix algebra. It is a square diagonal matrix with 1’s all along its diagonal and zeros elsewhere.
analysis. Typical values we might expect for pseudo-$F$ in equation (4.2) if the null hypothesis of no relationship were true can be obtained by randomly re-ordering the sample units in matrix $Y$ (or equivalently, by simultaneously re-ordering the rows and columns of matrix $G$), while leaving the ordering of samples in matrix $X$ (and $H$) fixed. For each permutation, a new value of pseudo-$F$ is calculated ($F^\pi$). There is generally going to be a very large number of possible permutations for a regression problem such as this; a total of $N!$ (i.e., $N$ factorial) re-orderings are possible. A large random subset of these will do in order to obtain an exact $P$-value for the test (Dwass 1957). The value of pseudo-$F$ obtained with the original ordering of samples is compared with the permutation distribution of $F^\pi$ to calculate a $P$-value for the test (see equation 1.4 in chapter 1).

Fig. 4.2. Schematic diagram of distance-based redundancy analysis as performed by DISTLM.

If Euclidean distances are used as the basis of the analysis, then DISTLM will fit a traditional linear model of $Y$ versus $X$, and the resulting $F$ ratio and $R^2$ values will be equivalent to those obtained from:

- a simple regression (when $p = 1$ and $q = 1$);
- a multiple regression (when $p = 1$ and $q > 1$); or
- a multivariate multiple regression (when $p > 1$ and $q > 1$). Traditional multivariate multiple regression is also called redundancy analysis or RDA (Gittins 1985, ter Braak 1987).

An important remaining difference, however, between the results obtained by DISTLM compared to other software for such cases is that all $P$-values in DISTLM are obtained by permutation, thus avoiding the usual traditional assumption that errors be normally distributed.

The added flexibility of DISTLM, just as for PERMANOVA, is that any resemblance measure can be used as the basis of the analysis. Note that the “hat” matrix ($H$) is so-called because, in traditional regression analysis, it provides a projection of the response variables $Y$ onto $X$ (Johnson & Wichern 1992) transforming them to their fitted values, which are commonly denoted by $\hat{Y}$. In other words, we obtain fitted values (“y-hat”) by pre-multiplying $Y$ by the hat matrix: $\hat{Y} = HY$. Whereas in traditional multivariate multiple regression the explained sum of squares (provided each of the variables in $Y$ are centred on their means) can be written as $tr(\hat{Y}\hat{Y}^\prime) = tr(HYY^\prime H^\prime)$, we achieve the ability to partition variability more generally on the basis of any resemblance matrix simply by replacing $YY^\prime$ in this equation with $G$ (McArdle & Anderson 2001). If Euclidean
In our first example of DISTLM, we will examine the relationship between the Shannon diversity \((H)\) of macrofauna and log copper concentration from benthic sampling at 12 sites along the Garroch Head dumpground in the Firth of Clyde, using simple linear regression. How much of the variation in macrofaunal diversity \((Y)\) is explained by variation in the concentration of copper in the sediment \((X)\)? In this case, \(p = q = 1\), so both \(Y\) and \(X\) are actually just single vectors containing one variable (rather than being whole matrices). We shall do a DISTLM analysis on the basis of Euclidean distances, which is equivalent to univariate simple linear regression, but where the \(P\)-value for the test is obtained using permutations, rather than by using traditional tables.

In PRIMER, open up the environmental data file \texttt{clev.pri}\footnote{This equivalence is noted on p. 426 of Legendre & Legendre (1998) and is also easy to verify by numerical examples.} and also the macrofaunal data file \texttt{clma.pri}, which are both located in the ‘Clydemac’ folder of the ‘Examples v6’ directory. Here, we will focus first only on a single predictor variable: copper. Highlight the variable ‘Cu’ and choose \textbf{Select} \textgreater{} \textbf{Highlighted}. Next, transform the variable (for consistency, we shall use the log-transformation suggested by Clarke & Gorley 2006), by choosing \textbf{Tools} \textgreater{} \textbf{Transform(individual)} \textgreater{} \textbf{Expression}: log(0.1+V). Rename the transformed variable ‘ln Cu’, and also rename the data sheet containing this variable ln Cu. Next, go to the sheet containing the macrofaunal data clma and choose \textbf{Analyse} \textgreater{} \textbf{DIVERSE}. In the resulting dialog, remove the tick marks from all of the measures (shown under the ‘Other’ tab or the ‘Simpson’ tab), so that the only diversity measure.

\footnote{An interesting point worth noting here is to consider how many \(df\) are actually available for a multivariate analysis. If \(p = 1\), then we have a total of \(N\) independent units. If \(p > 1\), then if all variables were independent of one another, we would have \(N \times p\) independent units. We do not expect, however, to have independent variables, so, depending on their degree of inter-correlation, we expect the true \(df\) for the multivariate system to lie somewhere between \(N\) and \(N \times p\). The Mantel-type procedure, which treats with \(N(N-1)/2\) units is generally going to overestimate the true \(df\), or information content of the multivariate system, while PERMANOVA and dbRDA’s use of \(N\) is, if anything, an underestimate. This is clearly an area that warrants further research.}
that will be calculated is Shannon’s index $H'$ (shown under the ‘Shannon’ tab) using Log base e. Place a tick mark next to Results to worksheet, then click OK. Rename the resulting data sheet $H'$. Calculate Euclidean distances among samples on the basis of $H'$ alone. That is, from the $H'$ data sheet, choose Analyse > Resemblance > (Analyse between •Samples) & (Measure •Euclidean distance).

Fig. 4.3. DISTLM procedure for simple linear regression of Clyde macrofaunal diversity ($H'$) versus log copper concentration (ln Cu).

With the predictor ($X$) variable(s) in one sheet and the resemblance matrix (arising from the response variable(s) $Y$) in another, we are now ready to proceed with the analysis. DISTLM, like all of the methods in the PERMANOVA+ add-on, begins from the resemblance matrix. Note that the number and names of the samples in the resemblance matrix have to match precisely the ones listed in the worksheet of predictor variables. They do not necessarily have to be in the same order, but they do have to have the same strict names, so that they can be matched and related to one another directly in the analysis. Choose PERMANOVA+ > DISTLM > (Predictor variables worksheet: ln Cu) & (Selection procedure •All specified) & (Selection criterion •R^2) & (Num. permutations: 9999) (Fig. 4.3).

The results file (Fig. 4.4) shows that the proportion of the variation in Shannon diversity explained by log copper concentrations is quite large ($R^2 = 0.815$, as seen in the column headed ‘Prop.’) and, not surprisingly, statistically significant by permutation ($P = 0.0001$). Thus, 81.5% of the variation in macrofaunal diversity among these 12 sites (as measured by $H'$ alone) is explained by variation in log copper concentration. Also shown in the output is the explicit quantitative partitioning: $SS_{Total} = 7.63$ and $SS_{Regression} = 6.22$. As there is only one predictor variable here, the information given under the heading ‘Marginal tests’ (Fig. 4.4) is all that is really needed from the output file. To see a scatterplot for this simple regression case, go to the ln Cu worksheet and choose Tools > Merge > Second worksheet: $H'$ > OK. From this merged data worksheet (called ‘Data1’ by default), choose Analyse > Draftsman Plot > •Correlations to worksheet > OK. The plot shows that $H'$ decreases strongly with increasing ln Cu and the Pearson linear correlation ($r$) is $-0.903$ (Fig. 4.4). Indeed, this checks out with what was given in DISTLM for $R^2$, as $(-0.903)^2 = 0.815$. 127
4. DISTLM and dbRDA

**Fig. 4.4.** DISTLM results for the regression of diversity of Clyde macrofauna ($H'$) versus log copper concentration (ln Cu).

More generally, when $X$ contains more than one variable, we may also be interested in *conditional* or *partial* tests. For example, if $X$ contains two variables $X_1$ and $X_2$, or (more generally) two sets of variables $X_1$ and $X_2$, one may ask: how much of the variation among samples in the resemblance matrix is explained by $X_2$, given that $X_1$ has already been fitted in the model? Usually, the two variables (or sets of variables) are not completely independent of one another, and the degree of correlation between them will result in there being some overlap in the variability that they explain (Fig. 4.5). This means that the amount explained by an individual variable will be different from the amount that it explains after one or more other variables have been fitted. A test of the relationship between a response variable (or multivariate data cloud) and an individual variable alone is called a *marginal test*, while a test of such a relationship after fitting one or more other variables is called a *conditional* or *partial* test. When fitting more than one quantitative variable in a regression-type of context like this, the *order* of the fit matters.

**Fig. 4.5.** Schematic diagram of partitioning of variation according to two predictor variables (or sets of variables), $X_1$ and $X_2$. 
Suppose we wish to test the relationship between the response data cloud and $X_2$, given $X_1$. The variable(s) in $X_1$ in such a case are also sometimes called covariates. If we consider that the model with both $X_1$ and $X_2$ is called the “full model”, and the model with only $X_1$ is called the “reduced model” (i.e. the model with all terms fitted except the one we are interested in for the conditional test), then the test statistic to use for a conditional test is:

$$F = \frac{(SS_{\text{full}} - SS_{\text{reduced}})/(q_{\text{full}} - q_{\text{reduced}})}{(SS_{\text{total}} - SS_{\text{full}})/(N - q_{\text{full}} - 1)}$$

(4.3)

where $SS_{\text{full}}$ and $SS_{\text{reduced}}$ are the explained sums of squares from the full and reduced model regressions, respectively. Also, $q_{\text{full}}$ and $q_{\text{reduced}}$ are the number of variables in $X$ for the full and reduced models, respectively. From Fig. 4.6, it is clear how this statistic isolates only that portion of the variation attributable to the new variable(s) $X_2$ after taking into account what is explained by the other(s) in $X_1$.

![Fig. 4.6. Schematic diagram of partitioning of variation according to two sets of predictor variables, as in Fig. 4.5, but here showing the sums of squares used explicitly by the $F$ ratio in equation (4.3) for the conditional test.](image)

The other consideration for the test is how the permutations should be done under a true null hypothesis. This is a little tricky, because we want any potential relationship between the response data cloud and $X_1$ along with any potential relationship between $X_1$ and $X_2$ to remain intact while we “destroy” any relationship between the response data cloud and $X_2$ under a true null hypothesis. Permuting the samples randomly, as we would for the usual regression case, will not really do here, as this approach will destroy all of these relationships. What we can do instead is to calculate the residuals of the reduced model (i.e. what is “left over” after fitting $X_1$). By definition, these residuals are independent of $X_1$, and they are therefore exchangeable under a true null hypothesis of no relationship (of the response) with $X_2$. The only other trick is to continue to condition on $X_1$, even under permutation. Thus, with each permutation, we treat the residuals as the response data cloud and calculate pseudo-$F$ as in equation (4.3), in order to ensure that $X_1$ is always taken into account in the analysis. This is required because, interestingly, once we permute the residuals, they will no longer be independent of $X_1$ (Anderson & Legendre 1999; Anderson & Robinson 2001)! This method of permutation is called permutation of residuals under a reduced model (Freedman & Lane 1989). Anderson & Legendre (1999) showed that this method performed well compared to alternatives in empirical simulations.

The definition of an exact test was given in chapter 1; for a test to be exact, the rate of type I error (probability of rejecting $H_0$ when it is true) must be equal to the significance level (such as $\alpha = 0.05$) that is chosen \textit{a priori}. Permutation of residuals under a reduced model does not provide an exact test, but it is asymptotically exact — that is, its type I error approaches $\alpha$ with increases in the sample size, $N$. The reason it is not exact is because we must estimate the relationship between the

---

80 See also the section entitled Designs with covariates in chapter 1.
81 But see the discussion in Manly (2006) and also Anderson & Robinson (2001), who show that this approach can still give reasonable results in some cases, provided there are no outliers in the covariates.
82 See the section entitled Methods of permutation in chapter 1 on PERMANOVA for more about permutation of residuals under a reduced model.
data cloud and $\mathbf{X}_1$ to get the residuals, and this is (necessarily) imperfect – the residuals only approximate the true errors. If we knew the true errors, given $\mathbf{X}_1$, then we would have an exact test. Clearly, the estimation gets better the greater the sample size and, as shown by Anderson & Robinson (2001), permutation of residuals under a reduced model is as close as we can get to this conceptually exact test.

There are (at least) two ways that the outcome of a conditional test can surprise the user, given results seen in marginal tests for individual variables:

- Suppose the marginal test of $Y$ versus $\mathbf{X}_2$ is statistically significant. The conditional test of $Y$ versus $\mathbf{X}_2$ can be found to be non-significant, after taking into account the relationship between $Y$ and $\mathbf{X}_1$.
- Suppose the marginal test of $Y$ versus $\mathbf{X}_2$ is not statistically significant. The conditional test of $Y$ versus $\mathbf{X}_2$ can be found, nevertheless, to be highly significant, after taking into account the relationship between $Y$ and $\mathbf{X}_1$.

The first of these scenarios is somewhat straightforward to visualise, by reference to the graphical representation in Fig. 4.7 (left-hand side). Clearly, if there is a great deal of overlap in the variability of $Y$ explained by $\mathbf{X}_1$ and $\mathbf{X}_2$, then the fitting of $\mathbf{X}_1$ can effectively “wipe-out” a large portion of the variation in $Y$ explained by $\mathbf{X}_2$, thus rendering it non-significant. On the other hand, the second of these scenarios is perhaps more perplexing: how can $\mathbf{X}_2$ suddenly become important in explaining the variation in $Y$ after fitting $\mathbf{X}_1$ if it was originally deemed unimportant when considered alone in the marginal test? The graphical representation in Fig. 4.7 (right-hand side) should help to clarify this scenario: here, although $\mathbf{X}_2$ explains only a small portion of the variation in $Y$ on its own, its component nevertheless becomes a significant and substantial proportion of what is left (compared to the residual) once $\mathbf{X}_1$ is taken into account. These relationships become more complex with greater numbers of predictor variables (whether considered individually or in sets) and are in fact multi-dimensional, which cannot be easily drawn schematically. The primary take-home message is not to expect the order of the variables being fitted in a sequential model, nor the associated sequential conditional tests, necessarily to reflect the relative importance of the variables shown by the marginal tests.

Fig. 4.7. Schematic diagrams showing two scenarios (left and right) where the conditional test of $Y$ versus $\mathbf{X}_2$ given $\mathbf{X}_1$ can differ substantially from the outcome of the marginal test of $Y$ versus $\mathbf{X}_2$ alone.

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 *hold.pri* in the ‘HoldNZ’ folder of the ‘Examples add-on’ directory. These multivariate data were analysed in the section entitled Designs with covariates 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 *et al.* 1996, Anderson *et al.* 2005a, b). 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 *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 *holdenv.pri*. The associated draftsman plot (Fig. 1.45) shows that these variables are reasonably evenly distributed across their range (see Assumptions & diagnostics, 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 *hold.pri* in PRIMER, select the molluscan species only and duplicate these into their own sheet, named *molluscs*, as shown in Fig. 1.27 in the section entitled Nested designs of chapter 1. Choose Analyse > DIVERSE and retain only the tick mark next to *Total species: S* on the ‘Other’ tab, but remove all others. Also choose to get *Total species* results to worksheet and name the resulting sheet *No. species*. 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 *holdenv.pri* 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: *holdenv*) 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: 9999).

The results (Fig. 4.8) 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.
4. DISTLM and dbRDA

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 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’).

Thus far, we have only done examples for a univariate response variable in Euclidean space, using DISTLM to fit linear models, but with tests being done by permutation. However, the fact that any resemblance measure can be used as the basis of the analysis in dbRDA yields considerable flexibility in terms of modeling. In traditional regression and RDA, the fitted values are a linear
combination of the variables in $X$. So, the relationship between the multivariate data $Y$ and predictor variables $X$ is assumed to be linear for the purposes of modeling. In dbRDA, however, the relationship being modeled between $Y$ and $X$ variables is more complex and depends on the chosen resemblance measure. Another way to describe dbRDA is that it is a traditional RDA done on the PCO axes from a resemblance matrix, rather than on the original $Y$ variables. Thus, in dbRDA, we effectively assume a linear relationship between the PCO axes derived from the resemblance matrix and the variables in $X$ for purposes of modeling (Legendre & Anderson 1999). In many cases, this is quite appropriate provided the resemblance measure used as the basis of the analysis is a sensible one for the data. By “sensible”, we mean that the resemblance measure describes multivariate variation in a way that emphasises the features of the data that are of interest (e.g., changes in composition, relative abundance, etc.) for specific hypotheses postulated by the user. Note, for example, that if one were to perform dbRDA on a chi-squared distance matrix, then this will assume unimodal relationships between $Y$ and $X$, as is done in canonical correspondence analysis (CCA, ter Braak 1986a, b)\(^{83}\). Clearly dbRDA goes beyond what either RDA or CCA can provide, by allowing any resemblance measure (e.g., Bray-Curtis, Manhattan, etc.) to define multivariate variation. In addition, once a given resemblance measure has been chosen, many other kinds of non-linear relationships between the PCO axes and $X$ can be modeled by introducing polynomials of the variables in $X$, if desired (e.g., Makarenkov & Legendre 2002).

Although dbRDA does provide quite impressive flexibility with respect to the response variables ($Y$), it pays to spend some time with the $X$ variables to examine their distributions and the relationships among them, as these are being treated in the same way in dbRDA as they would for any linear multiple regression model. Although DISTLM does not make any explicit assumptions about the distributions of the $X$ variables, they should nevertheless be reasonable for purposes of linear modeling – they should not be heavily skewed or contain extreme outliers. It is a very good idea, therefore, to examine the $X$ variables using a draftsman plot in PRIMER and to transform them (individually if necessary) to avoid skewness before proceeding.

The issue of multi-collinearity – strong inter-correlations among the $X$ variables – is also something to watch out for in dbRDA, as it is for RDA or multiple regression (e.g., Neter et al. 1996). If two variables in $X$ are very highly co-linear (with correlation $|r| \geq 0.95$, for example), then they contain effectively the same information and are redundant for purposes of the analysis. A redundant variable should be dropped before proceeding (keeping in mind that the variable which is retained for modeling may of course simply be acting as a proxy for the one that was dropped). Once again, PRIMER’s Draftsman Plot tool will provide useful direct information about multi-collinearity among the variables in $X$. See pp. 122-123 of chapter 11 in Clarke & Gorley (2006), for example, which demonstrate how to use the Draftsman Plot tool to identify skewness and multi-collinearity for a set of environmental variables.

In traditional multiple regression, the errors are assumed to be independent and identically distributed (i.i.d.) as normal random variables. DISTLM uses permutations to test hypotheses, however, so normality is not assumed. For a permutation test in regression, if we consider that the null hypothesis is true and $Y$ and $X$ are not related to one another, then the matching of a particular row of $Y$ (where rows identify the samples, as in Fig. 4.2) with a particular row of $X$ does not matter, and we can order the 1 to $N$ rows of $Y$ (or, equivalently, the rows and columns of matrix $D$) in any way we wish (e.g., Manly 2006). Thus, all that is assumed is that the sample rows are exchangeable under a true null hypothesis. For conditional tests, we assume that the residuals obtained after fitting covariates are exchangeable under a true null hypothesis. This means, more particularly, that we assume that the linear model being used to fit the covariate(s) to multivariate

---

\(^{83}\) Being careful, that is, to do all computations on the axes associated with positive and negative eigenvalues separately, and combining them only when they are squared (e.g., as sums of squares). Axes corresponding to the negative eigenvalues contribute negatively in the squared terms. See chapter 3 regarding negative eigenvalues and McArdle & Anderson (2001) for more details.

\(^{84}\) If the user performs dbRDA on the basis of the chi-squared distance measure, the results will produce patterns highly similar to those obtained using CCA. Any differences will be due to the intrinsic weights used in CCA. See ter Braak (1987), Legendre & Legendre (1998) and Legendre & Gallagher (2001) for details regarding the differences between CCA and RDA and the algebraic formulations for these two approaches.
data in the space of the resemblance measure is appropriate and that the errors (estimated by the residuals from this model) have homogeneous dispersions in that space, so they are exchangeable.

In many situations, a scientist may have measured a large number of predictor variables that could be potentially important, and interest lies in determining which ones are best at explaining variation in the response data cloud and also whether particular combinations of variables, working together, do a better job than other combinations in this respect. More specifically, one may wish to build a model for the response data cloud, using the best possible combination of predictor variables available. There are two primary issues one is faced with when trying to build models in this way: first, what criterion should be used to identify a “good” model and second, what procedure should one use to select the variables on the basis of said criterion?

In addition to providing tests of hypotheses for specific regression-style problems, DISTLM also provides the user with a flexible model-building tool. A suite of selection procedures and selection criteria are available, as seen in the DISTLM dialog box and described below.

**Selection procedures**

- **All specified** will simply fit all of the variables in the predictor variables worksheet, either in the order in which they appear in the worksheet (by default) or in the order given explicitly under the ‘Available’ column in the ‘Selection’ dialog. The ‘Selection’ dialog can also be used to force the exclusion or inclusion of certain variables from this or any of the other selection procedures as well.

- **Forward selection** begins with a null model, containing no predictor variables. The predictor variable with the best value for the selection criterion is chosen first, followed by the variable that, together with the first, improves the selection criterion the most, and so on. Forward selection therefore adds one variable at a time to the model, choosing the variable at each step which results in the greatest improvement in the value of the selection criterion. At each step, the conditional test associated with adding that variable to the model is also done. The procedure stops when there is no further possible improvement in the selection criterion.

- **Backward elimination** begins with a full model, containing all of the predictor variables. The variable which, when removed, results in the greatest improvement in the selection criterion is eliminated first. The conditional test associated with removing each variable is also done at each step. Variables are eliminated from the model sequentially, one at a time, until no further improvement in the criterion can be achieved.

- **Step-wise** begins with a null model, like forward selection. First, it seeks to add a variable that will improve the selection criterion. It continues in this fashion, but what distinguishes it from forward selection is that, after every step, it attempts to improve the criterion by removing a term. This approach is therefore like doing a forward selection, followed by a possible backward elimination at every step. The conditional test associated with either the addition or removal of a given variable is done at each step. The procedure stops when no improvements in the selection criterion can be made by either adding or deleting a term. Forward selection is often criticised because it does not allow removal of a term, once it is in the model. The rationale of the step-wise approach responds directly to this criticism.

- **Best** is a procedure which examines the value of the selection criterion for all possible combinations of predictor variables. One can choose the level of detail provided in the output file as ‘Normal’, ‘Brief’ or ‘Detailed’, which mirrors similar choices to be made when running the BEST procedure in PRIMER (see chapter 11 in Clarke & Gorley 2006). The default output from the Best selection procedure in DISTLM is to provide the best 1-variable model, the best 2-variable model, and so on, on the basis of the chosen selection criterion. The overall 10 best models are also provided (by default) in the output, but this number can be increased if desired. Be aware that for large numbers of predictor variables, the time required to fit all possible models can be prohibitive.

**Selection criteria**

- \( R^2 \) is simply the proportion of explained variation for the model, shown in equation (4.1). Clearly, we should wish for models that have good explanatory power and so, arguably, the larger the value of \( R^2 \), the better the model. The main drawback to using this as a selection
criterion is that, as already noted, its value simply increases with increases in the number of predictor variables. Thus, the model containing all \( q \) variables will always be chosen as the best one. This ignores the concept of parsimony, where we wish to obtain a model having good explanatory power that is, nevertheless, as simple as possible (i.e. having as few predictor variables as are really useful).

- **Adjusted \( R^2 \)** provides a more useful criterion than \( R^2 \) for model selection. We may not wish to include predictor variables in the model if they add no more to the explained sum of squares than would be expected by adding some random variable. Adjusted \( R^2 \) takes into account the number of parameters (variables) in the model and is defined as:

\[
R^2_{\text{adjusted}} = 1 - \frac{SS_{\text{Residual}}/(N-v)}{SS_{\text{Total}}/(N-1)}
\]

(4.4)

where \( v \) is the number of parameters in the model (e.g., for the full model with all \( q \) variables, we would have \( v = q+1 \), as we are also fitting an intercept as a separate parameter). Adjusted \( R^2 \) will only increase with decreases in the residual mean square, as the total sum of squares is constant. If adding a variable increases the value of \( v \) without sufficiently reducing the value of \( SS_{\text{Residual}} \), then adjusted \( R^2 \) will go down and the variable is not worth including in the model.

- **AIC** is an acronym for “An Information Criterion”, and was first described by Akaike (1973). The criterion comes from likelihood theory and is defined as:

\[
AIC = -2\ell + 2v
\]

(4.5)

where \( \ell \) is the log-likelihood associated with a model having \( v \) parameters. Unlike \( R^2 \) and adjusted \( R^2 \), smaller values of AIC indicate a better model. The formulation of AIC from normal theory in the univariate case (e.g., see Seber & Lee 2003) can also be written as:

\[
AIC = N \log(SS_{\text{Residual}}/N) + 2v
\]

(4.6)

DISTLM uses a distance-based multivariate analogue to this univariate criterion, by simply inserting the \( SS_{\text{Residual}} \) from the partitioning (as is used in the construction of pseudo-\( F \)) directly into equation (4.6). Although no explicit qualities of statistical likelihood, *per se*, are necessarily associated with the use of AIC in this form, we see no reason why this directly analogous function should not provide a reasonable approach. Unlike \( R^2 \), the value of AIC will not continue to get better with increases in the number of predictor variables in the model. The “+2v” term effectively adds a “penalty” for increases in the number of predictor variables.

- **AICc**, a modification of the AIC criterion that was developed to handle situations where the number of samples \( (N) \) is small relative to the number of predictor variables \( (q) \). AICc was found to perform rather poorly in these situations (Sugiura 1978, Sakamoto *et al.* 1986, Hurvich & Tsai 1989). AICc is calculated as:

\[
AICc = N \log(SS_{\text{Residual}}/N) + 2v(N/(N-v-1))
\]

(4.7)

In essence, the usual AIC penalty term (+2v) has been adjusted by multiplying it by the following correction factor: \((N/(N-v-1))\). Burnham & Anderson (2002) recommend, in the analysis of a univariate response variable, that AICc should be used instead of AIC whenever the ratio \( N/v \) is small. They further suggest that a ratio of (say) \( N/v < 40 \) should be considered as “small”! As the use of information criteria such as this in multivariate analysis (including based on resemblance matrices) is still very much in its infancy, we shall make no specific recommendations about this at present; further research and simulation studies are clearly needed.

- **BIC**, an acronym for “Bayesian Information Criterion” (Schwarz 1978), is much like AIC in flavour (it is not actually Bayesian in a strict sense). Smaller values of BIC also indicate a better model. The difference is that it includes a more severe penalty for the inclusion of extraneous predictor variables. Namely, it replaces the “+2v” in equation (4.6) with “+\( \log(N)v \)” instead. In the DISTLM context, it is calculated as:

\[
BIC = N \log(SS_{\text{Residual}}/N) + \log(N)v
\]

(4.8)

For any data set having a sample size of \( N \geq 8 \), then \( \log(N) > 2 \), and the BIC penalty for including variables in the model will be larger (so more strict) than the AIC penalty.
Depending on the resemblance measure used (and, to perhaps a lesser extent, the scales of the original response variables), it is possible for $AIC$ (or $AIC_c$ or $BIC$) to be negative for a given model. This is caused, not by the model having a negative residual sum of squares, but rather by $SS_{Residual}$ being less than 1.0 in value. When the log is taken of a value less than 1.0, the result is a negative number. However, in these cases (as in all others), smaller values of $AIC$ (or $AIC_c$ or $BIC$) still correspond to a better model.

Although there are other model selection criteria, we included in DISTLM the ones which presently seem to have the greatest general following in the literature (e.g., Burnham & Anderson 2002). For example, Godínez-Domínguez & Freire (2003) used a multivariate analogue to $AIC$ in order to choose among competing models in a multivariate canonical correspondence analysis (CCA). However, the properties and behaviour of these proposed criteria are still largely unknown in the context of dbRDA, especially with multiple response variables ($p > 1$) and for non-Euclidean resemblance measures. More research in this area is certainly required. In the context of univariate model selection, $AIC$ is known to be a rather generous criterion, and will tend to err on the side of including rather too many predictor variables; that is, to “overfit” (e.g., Nishii 1984, Zhang 1992, Seber & Lee 2003). On the other hand, trials using $BIC$ suggest it may be a bit too severe, requiring the removal of rather too many potentially useful variables. Thus, we suggest that if the use of $AIC$ and $BIC$ yield similar results for a given dataset, then you are probably on the right track! One possibility is to plot a scatter-plot of the $AIC$ and $BIC$ values for the top 20 or so models obtained for a given dataset and see which models fall in the lower left-hand corner (that is, those which have relatively low values using either of these criteria). These are the ones that should be considered as the best current contenders for a parsimonious model. An example of this type of approach is given in an analysis of the Ekofisk macrofauna below.

Cautionary notes

Before proceeding, a few cautionary notes are appropriate with respect to building models. First, the procedures of forward selection, backward elimination and step-wise selection are in no way guaranteed to find the best overall model. Second, even if the search for the “best” overall model is done, the result will depend on which selection criterion is used (adjusted $R^2$, $AIC$, $AIC_c$, or $BIC$). Third, DISTLM fits a linear combination of the $X$ variables, which may or may not be appropriate in a given situation (e.g., see the section on Linkage trees in chapter 11 of Clarke & Gorley 2006). As a consequence, it is certainly always appropriate to spend some time with the $X$ variables doing some diagnostic plots and checking out their distributions and relationships with one another as a preliminary step. Fourth, the particular predictor variables that are chosen in a model should not be interpreted as being necessarily causative\textsuperscript{85}. The variables chosen may be acting as proxies for some other important variables that either were not measured or were omitted from the model for reasons of parsimony. Finally, it is not appropriate to use a model selection procedure and then to take the resulting model and test for its significance in explaining variability. This approach uses circular and therefore invalid logic, because it is already the purposeful job of the model selection procedure to select useful explanatory variables. To create a valid test, the inherent bias associated with model selection would need to be taken into account by performing the selection procedure anew with each permutation. Such a test would require a great deal of computational time and is not currently available in DISTLM\textsuperscript{86}. In sum, model-building using the DISTLM tool should generally be viewed as an exploratory hypothesis-generating activity, rather than a definitive method for finding the one “true” model.

We shall now use the DISTLM tool to identify potential parsimonious models for benthic macrofauna near the Ekofisk oil platform in response to several measured environmental variables. The response data (in file ekma.pri in the ‘Ekofisk’ folder of the ‘Examples v6’ directory) consist of abundances of $p = 173$ species from 3 grab samples at each of $N = 39$ sites in a 5-spoke radial design (see Fig. 10.6a in Clarke & Warwick 2001). Also measured were $q = 10$ environmental variables (ekev.xls) at each site: Distance (from the oil platform), THC (total hydrocarbons),

\textsuperscript{85} Unless a predictor variable has been expressly manipulated experimentally in a structured and controlled way to allow causative inferences, that is.

\textsuperscript{86} The approach of including the selection procedure as part of the test is, however, available for examining non-parametric relationships between resemblance matrices as part of PRIMER’s BEST routine. See pp. 124-125 in chapter 11 of Clarke & Gorley (2006) for details.
Redox, % Mud, Phi mean (another grain-size characteristic), and concentrations of several heavy metals: Ba, Sr, Cu, Pb and Ni.

We begin by considering some diagnostics for the environmental variables. A draftsman plot indicates that several of the variables show a great deal of right-skewness (Fig. 4.10). The following transformations seemed to do a reasonable job of evening things out: a log transformation for THC, Cu, Pb and Ni, a fourth-root transformation for Distance and % Mud, and a square-root transformation for Ba and Sr. No transformation seems to be necessary for either Redox or Phi mean (Fig. 4.10), the latter already being on a log scale. Transformations were done by highlighting (not selecting) the columns for THC, Cu, Pb and Ni and using Tools > Transform(individual) > (Expression: log(V)) & (Rename variables). The procedure was then repeated on the resulting data file, first using (Expression: (V)^0.25) on the highlighted columns of Distance and % Mud, and then using (Expression: sqr(V)) on the highlighted columns of Ba and Sr. For future reference, rename the data file containing all of the transformed variables ekevt and save it as ekevt.pri. By choosing Analyse > Draftsman Plot > (Correlations to worksheet) we are able to examine not just the bivariate distributions of these transformed variables but also their correlation structure (Fig. 4.11).

For these data, several variables showed strong correlations, such as log(Cu), sqr(Sr) and log(Pb) (|r| > 0.8). The greatest correlation was between sqr(Sr) and log(THC) (|r| = 0.92). These strong inter-correlations provide our first indication that not all of the variables may be needed in a parsimonious model. We may choose to remove one or other of sqr(Sr) or log(THC), as these are effectively redundant variables in the present context. However, their correlation does not quite reach the usual cut-off of 0.95 and it might be interesting to see how the model selection procedures deal with this multi-collinearity. It is worth bearing in mind in what follows that wherever one of these two variables is chosen, then the other could be used instead and would effectively serve the same purpose for modeling.

Note that although the environmental variables may well be on different measurement scales or units, it is not necessary to normalise them prior to running DISTLM, because normalisation is done automatically as part of the matrix algebra of regression (i.e., through the formation of the hat matrix, see Fig. 4.2)\(^87\). If we do choose to normalise the predictor variables, this will make no difference whatsoever to the results of DISTLM.

---

\(^{87}\) This contrasts with the use of the RELATE or BEST procedures in PRIMER, which would require normalisation prior to analysis for situations such as this.
We are now ready to proceed with the analysis of the macrofauna. We shall base the analysis on the Bray-Curtis resemblance measure after square-root transforming the raw abundance values. An MDS plot of these data shows a clear gradient of change in assemblage structure with increasing distance from the oil platform (Fig. 1.7). From the ekma worksheet, choose Analyse > Pre-treatment > Transform (overall) > (Transformation: Square root), then choose Analyse > Resemblance > (Measure •Bray-Curtis). For exploratory purposes, we shall begin by doing a forward selection of the transformed environmental variables, using the $R^2$ criterion. This will allow us to take a look at the marginal tests and also to see how much of the variation in the macrofaunal data cloud (based on Bray-Curtis) all of the environmental variables, taken together, can explain. From the resemblance matrix, choose PERMANOVA+ > DISTLM > (Predictor variables worksheet: ekevt) & (Selection Procedure •Forward) & (Selection Criterion •$R^2$) & (●Do marginal tests) & (Num. permutations: 9999).

![Graph2](Ekofisk sediments - transformed)

![Resem7](Ekofisk sediments - transformed Correlation (-1 to 1))

**Fig. 4.11.** Draftsman plot and correlation matrix for the transformed Ekofisk sediment variables.

In the marginal tests, we can see that every individual variable, except for Redox and log(Ni), has a significant relationship with the species-derived multivariate data cloud, when considered alone and ignoring all other variables ($P < 0.001$, Fig. 4.12). What is also clear is that the variable (Distance)$^{(0.25)}$ alone explains nearly 25% of the variability in the data cloud, and other variables (sqr(Sr), log(Cu) and log(Pb)) also individually explain substantial portions (close to 20% or more) of the variation in community structure (Fig. 4.12). For the forward selection based on $R^2$, it follows that (Distance)$^{(0.25)}$ must be chosen first. Once this term is in the model, the variable that increases the $R^2$ criterion the most when added is log(Cu). Together, these first two variables explain 34.4% of the variability in the data cloud (shown in the column headed ‘Cumul.’). Next,
given these two variables in the model, the next-best variable to add in order to increase $R^2$ is $\text{sqr}(\text{Ba})$, which adds a further 4.17% to the explained variation (shown in the column headed ‘Prop.’). The forward selection procedure continues, also doing conditional tests at each step along the way, until no further increases in $R^2$ are possible. As we have chosen to use raw $R^2$ as our criterion, this eventually simply leads to all of the variables being included. The total variation explained by all 10 environmental variables is 52.2%, a figure which is given at the bottom of the ‘Cumul.’ column for the sequential tests, and which is also given directly under the heading ‘Best solution’ in the output file (Fig. 4.12).

Fig. 4.12. Results of DISTLM for Ekofisk macrofauna using forward selection of transformed environmental sediment variables.

A notable aspect of the sequential tests is that after fitting the first four variables: $(\text{Distance})^{0.25}$, log(Cu), $\text{sqr}(\text{Ba})$ and $\text{sqr}(\text{Sr})$, the $P$-value associated with the conditional test to add log(THC) to the model is not statistically significant and is quite large ($P = 0.336$). There is probably no real further mileage to be gained, therefore, by including log(THC) or any of the other subsequently fitted terms in the model. The first four variables together explain 42.1% of the variation in community structure, and subsequent variables add very little to this (only about 1.5-2% each). Given that we had noticed earlier a very strong correlation between log(THC) and $\text{sqr}(\text{Sr})$, it is not surprising that the addition of log(THC) is not really worthwhile after $\text{sqr}(\text{Sr})$ is already in the model. Although all of the $P$-values for this and the subsequent conditional tests are reasonably large ($P > 0.17$ in all cases), the conditional $P$-values in forward selection do not necessarily continue to increase and, indeed, a “significant” $P$-value can crop up even after a large one has been encountered in the list. It turns out that little meaning can be drawn from $P$-values for individual terms (whether large or small) after the first large $P$-value has been encountered in a series of sequential tests, as the inclusion of a non-significant term in the model such as this will affect subsequent results in various unpredictable ways, depending on the degree of inter-correlations among the variables.

Based on the forward selection results, we might consider constructing and using a model with these first four chosen variables only. This is clearly a more parsimonious model than using all 10 variables. However, the forward selection procedure is not necessarily guaranteed to find the best possible model for a given number of variables. We shall therefore explore some alternative
possible parsimonious models using the AIC and BIC criteria, in turn. From the resemblance matrix, choose PERMANOVA+ > DISTLM > (Predictor variables worksheet: ekevt) & (Selection Procedure •Best) & (Selection Criterion •AIC) & (Best > (Max num of best results: 10) 
& (Results detail: Normal)). There is no need to do the marginal tests again, so remove the ✓ from this option.

When the ‘Best’ selection procedure is used, there are two primary sections of interest in the output. The first is entitled ‘Best result for each number of variables’ (Fig. 4.13). For example, in the present case, the best single variable for modelling the species data cloud is identified as variable 1, which is (Distance)^0.25. The best 2-variable model, interestingly, does not include variable 1, but instead includes variables 6 and 7, corresponding to sqr(Ba) and sqr(Sr), respectively. The best 4-variable model has the variables numbered 1 and 6-8, which correspond to (Distance)^0.25, sqr(Ba), sqr(Sr) and log(Cu). Note that the best 4-variable model is not the same as the 4-variable model that was found using forward selection, although these two models have values for $R^2$ that are very close. Actually, it turns out that it doesn’t matter which selection criterion you choose to use ($R^2$, adjusted $R^2$, AIC, AIC, or BIC), this first section of results in the output from a ‘Best’ selection procedure will be identical. This is because, for a given number of predictor variables, the ‘penalty’ term being used within any of these criteria will be identical, so all that will distinguish models having the same number of predictor variables will be the value of $SS_{Residual}$ or, equivalently, $R^2$.

![Fig. 4.13. Results of DISTLM for Ekofisk macrofauna using the ‘Best’ selection procedure on the basis of the AIC selection criterion and then (below this) on the basis of the BIC selection criterion.](image)

The next section of results is entitled ‘Overall best solutions’, and this provides the 10 best overall models that were found using the AIC criterion (Fig. 4.13). The number of ‘best’ overall models and the amount of detail provided in the output can be changed in the DISTLM dialog. For the Ekofisk data, the model that achieved the lowest value of AIC (and therefore was the best model on
4. DISTLM and dbRDA

the basis of this criterion) had three variables: 1, 6 and 7 (i.e., (Distance)^(0.25), sqr(Ba) and sqr(Sr)). Another model that achieved an equally low value for \textit{AIC} had 4 variables: 1, 6 and 7. Indeed, a rather large number of models having 3 or 4 variables achieved an \textit{AIC} value that was within 1 unit of the best overall model, and even one of the 2-variable models (6, 7) was within this range. Burnham & Anderson (2002, p. 131) suggested that models having \textit{AIC} values within 2 units of the best model should be examined more closely to see if they differ from the best model by 1 parameter while still having essentially the same value for the first (non-penalty) term in equation (4.6). In such cases, the larger model is not really competitive, but only appears to be “close” because it adds a parameter and is therefore within 2 units of the best model (2\(v\) = 2 where \(v\) = 1), even though the fit (as measured by the first term in equation 4.6) is not genuinely improved. Generally, when a number of models produce quite similar \textit{AIC} values (within 1 to 2 units of each other, as seen here), this certainly suggests that there is a reasonable amount of redundancy among the variables in \(X\), so whichever model is eventually settled on, it is likely that a number of different combinations of predictor variables could be used inter-changeably in order to explain the observed relationship, due to these inter-correlations.

For comparison, we can also perform the ‘Best’ selection procedure using the \textit{BIC} criterion (Fig. 4.13). The more “severe” nature of the \textit{BIC} criterion is apparent straight away, as many of the best overall solutions contain only 2 or 3 variables, rather than 3 or 4, as were obtained using \textit{AIC}. However, a few of the models listed in the top 10 using \textit{BIC} coincide with those listed using \textit{AIC}, including \{6, 7\}, \{1, 6, 7\}, \{1, 6, 8\}, \{1, 7, 8\} and \{1, 2, 8\}. To achieve a balance between the more severe \textit{BIC} criterion and the more generous \textit{AIC} criterion, we could output the top, say, 20 models for each and then examine a scatter-plot of the two criteria for these models. This has been done for the Ekofisk data (Fig. 4.14). An astute choice of symbols (one being hollow and larger than the other) makes it easy to identify the models that appeared in both lists. Note that from the \textit{AIC} list, it is easy to calculate the \textit{BIC} criterion (and \textit{vice versa}), as SS\textsubscript{Residual} (denoted ‘RSS’ in the output file) and the number of variables in the model (which is \(v - 1\)) are provided for each model in the list.

![Fig. 4.14. Scatterplot of the \textit{AIC} and \textit{BIC} values for each of the top twenty models on the basis of either the \textit{AIC} criterion or on the basis of the \textit{BIC} criterion. Some of these models overlap (i.e. were chosen by both criteria).](image)

The exact correlation between \textit{AIC} and \textit{BIC} for models having a given number of variables is evident in the scatter plot, as individual models occur along a series of parallel lines (Fig. 4.14). These lines correspond to models having 1 variable, 2 variables, 3 variables, and so on. In this example, most of the best models have 2, 3 or 4 variables. The greatest overlap in models that were listed in the top 20 using either the \textit{AIC} or \textit{BIC} criterion occurred for 3-variable models. Moreover,
the plot suggests that any of the following models would probably be reasonable parsimonious choices here: \{1, 6, 7\}, \{6, 7\} or \{1, 6, 8\}. Balancing “severity” with “generosity”, we might choose for the time being to use the 3-variable model (bowing to AIC) which had the lowest BIC criterion, i.e., model \{1, 6, 7\}: (Distance)^a(0.25), sqr(Ba) and sqr(Sr), which together explained nearly 40\% of the variability in macrofaunal community structure. In making a choice such as this, it is important to bear in mind the caveats articulated earlier with respect to multi-collinearity and to refrain from making any direct causative inferences for these particular variables.

We may wish to visualise a given model in the multivariate space of our chosen resemblance matrix. The ordination method of PCO was introduced in chapter 3 as a way of obtaining orthogonal (independent) axes in Euclidean space that would represent our data cloud, as defined by the resemblance measure chosen, in order to visualise overall patterns. PCO is achieved by doing an eigenvalue decomposition of Gower’s centred matrix \(G\) (Fig. 3.1). Such an ordination is considered to be unconstrained, because we have not used any other variables or hypotheses to draw it. Suppose now that we have a model of the relationship between our data cloud and one or more predictor variables that are contained in matrix \(X\). We have constructed the projection matrix \(H\), and produced the sums of squares and cross-products matrix of fitted values \(HGH\) in order to do the partitioning (Fig. 4.2). Distance-based redundancy analysis (dbRDA) is simply an ordination of the fitted values from a multivariate regression model; it is achieved by doing an eigenvalue decomposition of matrix \(HGH\) (Fig. 4.15). That is:

1. Eigenvalue decomposition of matrix \(HGH\) yields eigenvalues \((\gamma_i^2, \ell = 1, \ldots, s)\) and their associated eigenvectors.
2. The dbRDA axes \(Z\) (also called dbRDA “scores”) are obtained by scaling (multiplying) each of the eigenvectors by the square root of their corresponding eigenvalue.

Fig. 4.15. Schematic diagram of distance-based redundancy analysis as a constrained ordination: an eigenvalue decomposition of the sums of squares and cross products of the fitted values from a model.

The result is a constrained ordination, because it will produce axes that must necessarily be directly and linearly related to the fitted values and, therefore, the predictor variables. If Euclidean distances are used at the outset to produce matrix \(D\), then \(G = YY’\) (where \(Y\) has been centered) and dbRDA is equivalent to doing a traditional RDA directly on a matrix of \(p\) response variables \(Y\). The number of dbRDA axes that will be produced (denoted by \(s\) above and in Fig. 4.15) will be the minimum of \((q, (N – 1))\). If the resemblance measure used is Euclidean, then \(s\) will be the minimum of \((p, q, (N – 1))\). As with most ordination methods, we can draw the first two or three dbRDA axes and examine patterns among the samples seen on the plot.
A dbRDA plot can be obtained within PRIMER in one of two ways using the new PERMANOVA+ add-on: either directly for a given resemblance matrix by choosing PERMANOVA+ > dbRDA and then providing the name of a worksheet containing the predictor variables to be used (all of the variables in this worksheet that are selected will be used in this case), or by ticking the option (✓ Do dbRDA plot) in the DISTLM dialog. For the latter, the dbRDA plot will be done on the predictor variables selected by the DISTLM routine, as obtained using the selection procedure and selection criterion chosen by the user.

Initial interest lies in determining the adequacy of the plot. More specifically, how much of the fitted model variation is captured by the first two (or three) dbRDA axes? As for other eigenvalue decomposition techniques (such as PCA or PCO), the eigenvalues are ordered from largest to smallest and yield information regarding the variation explained by each orthogonal (independent) dbRDA axis. The sum of the eigenvalues from dbRDA is equal to the total explained variation in the fitted model, that is: \( \sum \gamma_i = tr(GH) = SS_{Regression} \). So, the percentage of the fitted model’s variation that is explained by the \( \ell \) th dbRDA axis is \( (100 \times \gamma_i / \sum \gamma_i) \). As with PCO or PCA, we consider that if the percentage of the fitted variation that is explained by the diagram exceeds ~70%, then the plot is likely to capture most of the salient patterns in the fitted model.

Another relevant conceptual point regarding dbRDA is that there is an important difference between the percentage of the fitted variation explained by a given axis, as opposed to the percentage of the total variation explained by that axis. The latter is calculated as \((100 \times \gamma_i / tr(G))\) or \((100 \times \gamma_i / \sum \lambda_i)\), where \(\sum \lambda_i\) is the sum of PCO eigenvalues. The dbRDA axis might be great at showing variation according to the fitted model, but if the model itself only explains a paltry amount of the total variation in the first place, then the dbRDA axis may be of little overall relevance in the multivariate system as a whole. It is therefore quite important to consider the percentage of explained variation for each dbRDA axis out of the fitted and out of the total variation. Another way of thinking about this is to compare the patterns seen in the dbRDA plot with the patterns seen in the unconstrained PCO plot. If the patterns are similar, then this indicates that the model is a pretty good one and captures much of what can be seen overall in the multivariate data cloud. Such cases should also generally correspond to situations where the \(R^2\) for the model \((SS_{Regression} / SS_{Total} = \gamma_i / \sum \lambda_i)\) is fairly large. In contrast, if the pattern in the dbRDA and the PCO are very different, then there are probably other factors that are not included in the model which are important in driving overall patterns of variation.

Something which certainly should come as no surprise is to see the X variables playing an important role in driving the variation along dbRDA axes. Of course, the X variables must feature strongly here, because it is from these that the fitted variation is derived! To characterise the dbRDA axes, it is useful to determine the strength and direction of the relationship between individual X variables and these axes. A common method for visualising these relationships is to examine vector overlays on the ordination diagram, with one vector for each predictor variable. The positions and directions of appropriate vectors can be obtained by calculating the multiple partial correlations between each of the X variables and the dbRDA axis scores. These vectors can be interpreted as the effect of a given predictor variable on the construction of the constrained ordination picture – the longer the vector, the bigger the effect that variable has had in the construction of the dbRDA axes being viewed. If the ordination being viewed explains a large proportion of the variation in the fitted model (and we would hope that this is indeed generally true), then these vectors are also representative of the strength and direction of influence of individual X variables in the model itself. Note that these vectors show the strength of the relationships between each predictor variable and the dbRDA axes, given that it is fitted simultaneously with the other X variables in the model. In other words, the calculation of the correlation between each X variable and each dbRDA axis is conditional upon (i.e. takes into account) all of the other X variables in the worksheet.

Although the above vector overlay is shown by default in the dbRDA ordination graphic, there are other kinds of potentially informative vector overlays. For example, one might choose to plot the simple Pearson (or Spearman) correlations of the X variables with the dbRDA axes, as was shown

---

88 The eigenvalues from a redundancy analysis are sometimes called “canonical eigenvalues”, as in chapter 11 in Legendre & Legendre (1998).
for PCO plots in the section Vector overlays in chapter 3. Note, however, that each of these vectors is drawn individually, ignoring the other variables in the worksheet. In PRIMER, there are therefore three different kinds of vector overlays offered for dbRDA plots as part of the PERMANOVA+ add-on. These are obtained as follows from within the ‘Configuration Plot’ dialog (presuming for the moment that the predictor variables that gave rise to the model are located in a worksheet named X):

- Default dbRDA vector overlay, corresponding to multiple partial correlations of the (centred) X variables with the dbRDA axes: choose (Vectors: • Base variables). These vectors are identical to those that are obtained using the choice (Vectors: • Worksheet variables: X > Correlation type: Multiple).
- Pearson simple linear correlations of individual X variables with the dbRDA axes: choose (Vectors: • Worksheet variables: X > Correlation type: Pearson)
- Spearman simple rank correlations of the X variables with the dbRDA axes: choose (Vectors: • Worksheet variables: X > Correlation type: Spearman)

Of course, some other variables (and not just the predictor variables that gave rise to the dbRDA axes) can also be superimposed on the dbRDA ordination diagram using this general vector overlay tool. Bubble plots (see p. 83 in chapter 7 of Clarke & Gorley 2006) can also be useful to visualise the relative values of an individual variable among sample points in a dbRDA ordination.

---

Fig. 4.16. dbRDA of the Ekofisk data from the parsimonious model with three environmental variables.

Clearly, the type of vectors to plot is up to the user and depends on the type of information being sought. When publishing a dbRDA plot as part of a study, it is important to give information about the variation explained by the axes (out of the fitted and out of the total variation) and also to indicate which type of vector overlays are being used, if any are displayed. Of course, all of the other graphical options and tools available from within PRIMER are also able to be used in dbRDA.

---

89 Although ter Braak (1990) suggested that individual Pearson correlations (ignoring other X variables) are most informative, Rencher (1988, 1995) disagrees, stating that the multivariate nature of these inter-relationships is ignored by using raw correlations, which show only univariate relationships. Both types of vector overlay are available within the PERMANOVA+ add-on.

90 The ‘Configuration Plot’ dialog is obtained for any plot by choosing Graph > Special.
Let us examine the constrained dbRDA ordination for the parsimonious model obtained earlier using DISTLM on the Ekofisk data. The parsimonious model we had settled on included three predictor variables: (Distance)^0.25, sqr(Ba) and sqr(Sr). Go to the ekfvt worksheet, highlight and then select these three variables by choosing Select > Highlighted. Now return to the resemblance matrix based on the macrofaunal data and choose PERMANOVA+ > dbRDA > (Predictor variable worksheet: ekfvt) & (Max. no. of dbRDA axes: 3) & (Plot results). The resulting pattern among the samples (Fig. 4.16) suggests that there are effectively two gradients (forming sort of an upside-down “V” shape) in the community structure of the macrofauna that can be modeled by these environmental variables. The first largely distinguishes among groups A, B, C and D and is (not surprisingly) driven largely by distance from the oil platform, as well as the concentrations of Ba in the sediments. The second gradient identifies variability among the sites within group D that are close to the oil platform (< 250 m). These differences are apparently mostly driven by differences in the concentrations of Sr in the sediments. We shouldn’t forget, however, that sqr(Sr) had a very strong relationship (r = 0.92) with log(THC), so the modeled variation in community structure among these samples near the platform could just as easily be due to variation in total hydrocarbons as to strontium concentrations (or to the combined effects of both, or to some other unmeasured variables)\(^91\).

\(^{91}\) Quite similar roles for Ba and THC as those described here were outlined for this data set by Clarke & Gorley (2006, p. 84, chapter 7) from examining bubble plots of these environmental variables on the species MDS plot, although no formal modeling of these relationships was done in that purely non-parametric setting.
individual axes’ relate to the percentage explained out of the fitted model (i.e., out of $tr(HGH)$ or $SS_{Regression}$). The next two columns in this section relate to the percentage explained out of the total variation in the resemblance matrix (i.e., out of $tr(G)$ or $SS_{Total}$). In the present case, the first two dbRDA axes explain 94.0% of the fitted variation, and this is about 36.7% of the total variation in the resemblance matrix. So, we can rest assured that these two dbRDA axes are capturing pretty much everything we should wish to know about the fitted model, although there is still quite a lot of residual variation in the original data matrix which is not captured in this diagram. Note that there are three dbRDA axes as $s = \min(q, (N-1)) = \min(3, 38) = 3$. All of the dbRDA axes together explain 100% of the fitted variation. Taken together, they also explain 39.1% of the total variation. This is equal to $100 \times R^2$ from the DISTLM model for these three variables (Fig. 4.13).

Another way to check the overall utility and adequacy of the model as it is shown in the constrained dbRDA plot is to compare it with an unconstrained PCO plot of the same data. If the patterns of sample points in the two plots are similar, then this indicates that the dbRDA (and by implication, the model) is doing a good job of finding and explaining the most salient patterns of variation across the data cloud as a whole. If, on the other hand, these two plots are quite different, then, although it does not mean the model is useless, it does indicate that there are likely to be some other structuring forces out there which were not measured or included in the model. The PCO plot of the Ekofisk data shows a remarkably similar pattern to the dbRDA plot (Fig. 4.18), indicating that this three-variable model is indeed capturing the most salient overall patterns of variability. Although there is still quite a lot of unexplained variation (even the first two PCO axes explain only 43.1% of the total), these are the axes of greatest variation through the data cloud defined by the resemblance measure chosen and thus the environmental drivers used in the model (and variables correlated with them) are very likely to be the most important drivers of differences in macrofaunal community structure among these samples.

The next portion of the output file provides the dbRDA coordinate scores (Fig. 4.17). These are the positions of the samples along the dbRDA axes and they can also be output into a separate worksheet by ticking the (Scores to worksheet) option in the dbRDA dialog. Next, the output file provides information regarding the vector overlay positions of each variable along each axis (Fig. 4.17). Specifically, under the heading ‘Relationships between dbRDA coordinate axes and orthonormal X variables’, are given the values of the multiple partial correlations; these are the values that are plotted in the default dbRDA vector overlay.

The last portion of the output from dbRDA shows the relative importance of each X variable in the formation of the dbRDA axes (Fig. 4.17). These are entitled ‘Weights’ and subtitled ‘(Coefficients for linear combinations of X’s in the formation of dbRDA coordinates)’. Recall that the dbRDA axes are linear combinations of the fitted values and that these, in turn, are linear combinations of
the original \( X \) variables. Thus, we can obtain the direct conditional relationships of individual \( X \) variables in the formation of each dbRDA axis by multiplying these two sets of linear combinations together. The resulting weights given here are identical to the standardised partial regression coefficients that would be obtained by regressing each of the dbRDA axes scores \((Z)\) directly onto the matrix of normalised explanatory variables \( X \).\(^{92}\)

In some situations, it is useful to be able to partition variability in the data cloud according to \emph{sets} of predictor variables, rather than treating each variable individually. For example, Borcard \emph{et al.} (1992) discussed the partitioning of variation in multivariate species data among two sets of variables: a set of environmental variables and a set of spatial variables (such as latitude and longitude, arbitrary spatial coordinates, or polynomial functions of these). They considered that analyses of the relationship between species data and environmental variables should include a consideration of the intrinsic spatial structuring caused simply by the relative geographic distances among samples at a given scale. One might expect, for example, that samples close together would be more similar than those further apart. By analysing the data in sets, one can explicitly examine the proportion of variation in the species data that is explained by the environmental variables over and above the amount explained by the spatial variables alone.

DISTLM can treat variables either individually or in sets. To identify sets of variables, one needs to first define an \emph{indicator} which will identify the set that a particular predictor variable belongs to. Recall that an \emph{indicator} in PRIMER identifies groups of \emph{variables} the same way that a \emph{factor} identifies groups of \emph{samples} (see chapter 2 in Clarke & Gorley 2006).

As an example, we shall analyse a data set on the responses of heterotrophic bacteria grown in different media (labeled ‘Ma’ and ‘Bna’) to sets of environmental and spatial variables obtained from each of 20 sites in the Thau lagoon (Amanieu \emph{et al.} 1989) provided by Legendre & Legendre 92. If a Euclidean distance matrix was used as the basis of the dbRDA analysis, these weights correspond precisely to the matrix called \( C \) in equation 11.14 on p. 585 of Legendre & Legendre (1998).
Data on the bacteria are located in the file \texttt{thbac.pri}, and the associated environmental and spatial variables are located in the file \texttt{thevsp.pri}, both in the ‘Thau’ folder of the ‘Examples add-on’ directory. For the bacteria, the variable \textit{Bna} is the concentration of colony-forming units of aerobic heterotrophs growing on bioMérieux nutrient agar (low salinity) and \textit{Ma} is the concentration growing on marine agar. The environmental variables are \textit{NH4} (ammonium in the water column, in \textmu mol per litre), \textit{Phaeo.a} (phaeopigments from degraded chlorophyll \textit{a}, in \textmu g per litre) and \textit{Prod} (bacterial production, determined by incorporation of titrated thymidine in bacterial DNA, in nmol per litre per day). Each of these environmental variables and also the bacteria concentrations have already been transformed using \textit{ln}(X+1). The spatial variables (named \(X\) and \(Y\)) are the positions of the samples in terms of geographic coordinates according to an arbitrary grid, and have been centred on their means. The spatial variable \(X^2\) is the square of \(X\) and is included as another spatial predictor variable of potential importance.

Open up both of these data files in PRIMER. Focusing first on the sheet containing environmental and spatial variables, choose \textit{Edit > Indicators} and an indicator called (rather uncreatively) ‘Indicators’ will be shown that identifies which set (environmental or spatial, denoted by ‘Env’ or ‘Geo’, respectively) each of these predictor variables belongs to (Fig. 4.19). Indicators, like factors, can either be created from within PRIMER, or they can be brought into PRIMER along with the data file. Next, it is wise to examine a draftsman plot to see the distributions of these variables and the relationships among them, prior to fitting the model. Whether we fit the predictor variables alone or in sets, we do need to satisfy ourselves that their distributions are reasonable and check for high collinearity, removing any clearly redundant variables if necessary. For these data (Fig. 4.19), it is interesting to note that the scatterplot of the spatial variables \(X\) and \(Y\) effectively shows a map of the spatial positions of the samples. Furthermore, it is no surprise to see that \(X^2\) has a perfect quadratic relationship with \(X\). For the rest, all seems well and we are ready to proceed.

The response variables, in this case, are concentrations of bacteria that contain no zeros and (after the log-transformation, which has already been done for these data) are quite well-behaved in their distributions (i.e., show fairly equal scatter throughout their range and are not heavily skewed). Thus, we can quite reasonably consider doing an analysis on the basis of the Euclidean distance measure. The variables are also in the same units of measurement and occur on similar scales, so no preliminary normalisation is needed here. Although normalisation of \textit{predictor} variables is never required in DISTLM (and would make no difference to the results in any event), it is important to consider, when calculating Euclidean distances on the basis of a set of \textit{response}
variables, whether or not a prior normalisation is appropriate. For this, the usual considerations (such as the units of the variables and their relative scales) apply.

Calculate a Euclidean distance matrix from the thbac data sheet and then choose **PERMANOVA+** > **DISTLM** > (Predictor variables worksheet: thevsp) & (✓ Group variables (indicator): Indicators) & (Selection Procedure • All specified) & (Selection Criterion • \(R^2\)) & (Num. permutations: 9999) & (✓ Do marginal tests), as shown in Fig. 4.20. Before clicking the ‘OK’ button, check out the order in which these sets of variables will be fitted by clicking on the ‘Select variables/groups…’ button. The order shown will be the order in which these groups were provided in the data file. In our case, however, it makes sense specifically to fit the ‘Geo’ set first, followed by the ‘Env’ set, as we are interested in testing the hypothesis of there being no relationship between the bacteria concentrations and the environmental variables, **given** the spatial variables. This can be done by highlighting Geo appearing inside the ‘Available:’ column and then clicking on the upward arrow to move it up to first in the list (Fig. 4.20).

The file of results from this analysis is shown in Fig. 4.21. We can see that the three spatial variables alone accounted for 36.4% of the variation in the bacteria concentrations, while the environmental variables accounted for 45.0%. However, after fitting the spatial variables, the environmental variables explained an additional 18.5%, resulting in a total explained variation of 55%. This additional amount was not statistically significant, however, according to the sequential test (\(P = 0.16\)).

![Fig. 4.21. Results of DISTLM fitting sets of predictor variables to bacteria concentrations from Thau lagoon.](image-url)

Partitioning of multivariate data described by a resemblance matrix of choice in response to multiple sets of variables (as in Anderson & Gribble 1998, for example, who partitioned temporal, spatial and environmental components) can be done in this manner, and the degree of overlap between individual sets can be readily determined by changing the order of fit of the variables, as desired. Keep in mind, however, that sets having different numbers of predictor variables will naturally have a different capability when it comes to explaining variation – a set with one variable would be expected naturally to explain less than a set with 5 variables, simply because it has fewer degrees of freedom. We have chosen in this example to use the simple \(R^2\) criterion (both sets had 3 variables, so the percentage of variation explained could be directly compared), but an adjusted \(R^2\), \(AIC\), \(AIC_c\) or \(BIC\) criterion could also be used, and these would (each in their own way) take into...
account different numbers of variables in the different sets\textsuperscript{93}. Such an approach is directly analogous to examining components of variation derived from mean squares, rather than just looking at the raw partitioning of the sums of squares when comparing the relative importance of different sources of variation in a PERMANOVA model. In this example, we also had a specific hypothesis which dictated the order of the fit (the ‘Geo’ set first, followed by the ‘Env’ set), but in other situations, a different selection procedure (i.e., forward, backward, step-wise, best) might be more appropriate for the hypotheses or goals of the study.

Sometimes the predictor variables of interest are not quantitative, continuous variables, but rather consist of categories or groups, called categorical or nominal variables. There are also situations where we have mixtures of variable types that we want to include in a single DISTLM analysis. For example, we might want to consider a single model that includes temperature and salinity (continuous and quantitative) but which also includes a categorical variable, habitat, that might be recorded for each sample as one out of a number of different types: sand, coral, rock, seagrass, etc. How can we analyse categorical predictor variables using DISTLM?

As an example, we will consider an analysis of environmental control and spatial structure in an oribatid mite community, described by Borcard \textit{et al.} (1992) and Borcard & Legendre (1994). These data come from a site adjacent to a small Laurentide lake, Lake Geai, at the Station de Biologie des Laurentides in Quebec, Canada. The species (response) data (in the file \texttt{ormites.pri} in the ‘OrbMit’ folder of the ‘Examples add-on’ directory) consist of counts of abundances for each of \(p = 35\) species of oribatid mites from each of \(N = 70\) sites. The file \texttt{orenvgeo.pri} (in the same folder) contains a number of predictor variables associated with each of the sites. Some of these are spatial variables and correspond to the geographic coordinates of each sample on a grid measured in metres (‘x’, ‘y’) and up to third-order polynomial functions of these (‘x2’, ‘xy’, ‘y2’, etc.). The others are environmental variables, which include two quantitative variables:

- Substratum density (‘Substr.dens’), measured in grams per litre of dry uncompressed matter, and
- Water content of the substratum (‘H2O.cont’), measured in grams per litre,

and three categorical (nominal) variables:

- Substrate (7 categories: Sphagnum groups 1-4, litter, bare peat, interfaces),
- Shrubs (3 categories: no shrubs, few shrubs, many shrubs), and
- Microtopography (2 categories: blanket (flat) and hummock (raised))

The nominal variable ‘Shrubs’ is actually ordinal or semi-quantitative, so the simple rank order values of ‘1’, ‘2’ and ‘3’ in a single column could also be used here to treat this as a single (1 df) continuous quantitative predictor variable. However, we shall treat it as a nominal variable here to be consistent with the way this variable was treated by Borcard \textit{et al.} (1992) and Borcard & Legendre (1994).

\textbf{Fig. 4.22.} Example of a categorical (nominal) variable and its expanded binary form.

\textsuperscript{93}Peres-Neto \textit{et al.} (2006) discussed the use of an adjusted \(R^2\) criterion to perform a partitioning of multivariate response data in RDA and CCA models and to allow comparison of the sizes of portions explained by different sets of predictor variables.
To start with, we need to expand each categorical variable of interest into a set of “on/off” or *binary* variables (e.g., Fig. 4.22). For each categorical variable, there will be as many binary variables as there are categories. The binary variables will take a value of ‘1’ for samples where that category occurs, and zero (‘0’) elsewhere. Once this has been done, the binary variables associated with a particular nominal variable need to be specified as a set, using an indicator in PRIMER (see the previous section on setting up indicators to analyse sets of variables). For example, the seven “on/off” variables that “code” for each of the seven substrate categories are identified (by having a common name for the indicator) as belonging to the ‘Substrate’ set. If there are also quantitative variables to be analysed as part of the same predictor variable worksheet, then each of these can simply be specified in the indicator as belonging to its own “set”. The environmental data, with the categorical variables already expressed as binaries and identified in sets using the indicator called ‘Type’, are located in the file *orenvgeo.pri* (Fig. 4.23).

The *df* for each categorical set is the number of categories minus 1. Each quantitative variable has one degree of freedom (= 1 *df*) associated with it for a regression. Another way of thinking about this is that we have to estimate one slope parameter (or coefficient) for each quantitative variable we wish to include in the model. When we have a categorical variable, however, the number of degrees of freedom associated with it is the number of categories minus 1. The reason for the “minus 1” becomes clear when we think about the way the binary “codes” for a given set are specified. Consider a categorical variable that has two categories, like ‘Microtopography’ in the oribatid mite example. Once you are given the first binary variable ‘Blanket’, you automatically know what the other binary variable must be: ‘Hummock’ will have 1’s wherever ‘Blanket’ has zeros, and *vice versa*. Therefore, the variable ‘Hummock’ is actually completely redundant here. This issue is also readily seen by calculating the correlation between ‘Blanket’ and ‘Hummock’, $r = -1.0$, a perfect negative correlation! So one (or the other) of these variables needs to be removed for the analysis. Fortunately, the DISTLM routine automatically determines the correct degrees of freedom for a given set of variables and will omit unnecessary variables, even when all of the binary variables have been included by the user.

As an aside, analysis of variance *is* simply a regression on categorical variables. Therefore, we can actually do PERMANOVA-style analyses using DISTLM by setting up ‘binary codes’ for each of

---

94 This phenomenon is sometimes referred to as the “overparameterisation” of an ANOVA model.
the factors (as in Fig. 4.22). With the PERMANOVA+ tool at our disposal, however, it is much more efficient to analyse the responses of multivariate variables to an ANOVA design using the PERMANOVA routine instead, as this already caters for balanced and unbalanced designs with or without quantitative (or categorical) covariables. PERMANOVA also allows factors to be nested within other factors and correctly analyses random factors and mixed models, whereas DISTLM will treat all of the predictor variables (whether they be quantitative or categorical) as fixed\(^95\). In some sense, it may seem odd to treat quantitative predictor variables as if they are “fixed”. In many ecological applications, for example, a scientist has measured the \(X\) predictor variables (say, a suite of environmental variables) for each sample in much the same way as the \(Y\) response variables (say, abundances or counts of species). So arguably the \(X\) variables should be modeled with error, just as the \(Y\) variables are (e.g., McArdle 1988, Warton & Weber 2002, McArdle 2003, Warton et al. 2006). Another way of viewing the regression approach, however, is not so much that the \(X\) variables are measured without error, but rather that the models we construct from them are conditional on the actual values we happened to observe (e.g., Neter et al. 1996). One natural consequence of this philosophical stance (among others) is that it is not appropriate to extrapolate the resulting models beyond the range of observed values of \(X\), a well-known caveat for regression.

 Returning to the example, for simplicity we shall restrict our attention to the environmental variables only and omit the spatial variables from the analysis in what follows. The distributions of the categorical variables need not be considered using the usual diagnostic tools, as these only consist of 1’s and 0’s. A scatter plot using the draftsman plot tool in PRIMER for the two quantitative predictor variables alone (‘H2O.cont’ and ‘Substr.dens’) shows fairly even scatter and no indication that either of these variables requires any transformation prior to analysis. So, we are ready to proceed. First, calculate a Bray-Curtis resemblance matrix after square-root transforming the oribatid mite species abundance data. From this resemblance matrix, choose PERMANOVA+ > DISTLM > (Predictor variables worksheet: orenvgeo) & (Group variables (indicator) Type) & (Selection Procedure Step-wise) & (Selection Criterion Adjusted R^2) & (Num. permutations: 9999) & (Do marginal tests) & (Do dbRDA plot). Before clicking the ‘OK’ button for the

---

\(^95\) The treatment of quantitative predictor variables as random (sometimes called “Model II” regression) does not exist within the current framework of either PERMANOVA or DISTLM and is a topic for further research. The CAP routine, however, is designed to examine the canonical correlations between two (sphericised) data clouds; it can be considered as one type of multivariate analogue to Model II regression. See chapter 5 for details.
DISTLM dialog, force the exclusion of the spatial variables by clicking on the ‘Select variables/groups’ button and moving the group ‘Geo’ over into the ‘Force exclusion:’ column of the ‘Selection’ dialog (Fig. 4.24). The choice of adjusted $R^2$ as the selection criterion for this case is especially appropriate, because this criterion will take into account the fact that the sets of predictor variables have different numbers of variables in them, whereas the use of the raw $R^2$ criterion will not. If model selection were the ultimate aim, then one might alternatively choose to use $AIC$ or $BIC$ instead, which also would take into account the different numbers of variables in different sets. In our case, interest lies in determining how much of the variability is explained by each set of variables (thus the choice to examine marginal tests), and also what a sequential step-wise model fit of these sets of variables would produce.

Marginal tests (Fig. 4.25) show that each of the sets explains a significant proportion of the variation in the mite data, when considered alone ($P < 0.05$ in all cases). The single variable of water content (‘H2O.cont’) alone explains the greatest amount of variation in the oribatid mite species data cloud (based on Bray-Curtis), at 29.2%, while substrate density explains the least (only 3.6%). The set of variables that increased the value of adjusted $R^2$ the most after fitting ‘H2O.cont’ was ‘Substrate’, followed by ‘Shrub’, ‘Microtop’ and then ‘Substr.dens’. Although a step-wise procedure was used, at no stage were there any eliminations of sets from the model once they had been added and the conditional tests associated with each of the sequential additions were statistically significant ($P < 0.001$ in all cases). All of these predictor variables together explained 57.7% of the variation in the species data cloud and the adjusted $R^2$ for the full model was 0.497 (Fig. 4.25).

The full model can be visualised by examining the dbRDA ordination, requested as part of the output. The first two dbRDA axes captured nearly 80% of the variability in the fitted model, 45.7% of the total variation in the data cloud (Fig. 4.26). The vector overlay shows how the first dbRDA
axis is particularly strongly related to water content, shrubs (none versus many) and microtopography (blanket versus hummock). When categorical variables are included in a DISTLM analysis, then the length of each categorical vector is a measure of the strength of the relationship between that category and the dbRDA axes. More particularly, if the separation of groups is clear in the plot, then we would expect the vectors for those categories to be relatively long. To help interpret ordination plots, it may be useful to provide the categorical variables also as factors; this will allow labels and symbols to be placed on the plot according to these different groupings. For example, by superimposing symbols corresponding to the three categories of ‘Shrubs’, a gradient from ‘none’ to ‘many’ (left to right) is apparent in the dbRDA diagram (Fig. 4.26). We leave it to the user to explore other categorical variables in this way, to examine unconstrained ordinations for the oribatid mite data and to consider analyses that might include the spatial variables as well.

**Fig. 4.26.** dbRDA ordination for the fitted model of oribatid mite data (based on Bray-Curtis after square-root transformation of abundances) versus environmental variables.

DISTLM vs BEST/ BIOENV

On the face of it, the DISTLM routine might be thought of as playing a similar role to PRIMER’s BEST routine in the analysis of multivariate species data. More particularly, the BEST (BIOENV or BVSTEP) procedure in PRIMER is designed to find a combination of environmental variables which, together, result in resemblances among samples whose rank order best matches the rank order of the inter-sample resemblances arising from biological (species) data. There are some

---

96 See chapter 11 in Clarke & Warwick (2001) and chapter 11 in Clarke & Gorley (2006) for more details regarding these routines.
important differences, however, between the BEST/BIOENV approach and models of a multivariate data cloud obtained using DISTLM.

First, DISTLM actually formally fits a linear model of the predictor (environmental) variables to the response (species) data cloud, in the space defined by the chosen resemblance measure. This means that the dissimilarities (or similarities) themselves are important here. They will define the shape and structure of the data cloud, so (as in PERMANOVA), the resemblance measure the user chooses is quite important. Indeed, although our ability to view the structure of this data cloud using unconstrained ordination (such as PCO) is necessarily imperfect, such an ordination should nevertheless be used to provide some information on whether there are any gross outliers, for example, which would make a linear modeling approach inappropriate. DISTLM does not assume anything specific about the shape of the data cloud, and any resemblance measure that is deemed appropriate for the nature of the data and hypotheses of interest can be used to construct it, but outliers or “high leverage” points\(^7\) in that space, if present, will tend to have a strong influence on the results.

The advantages of fitting a formal model using DISTLM are fairly clear. First, we achieve a direct quantitative partitioning of the multivariate variability that is explained by each of several environmental variables. Thus, we can determine how much of the variability is attributable to individual predictor variables (either acting alone or in pre-defined sets), and we can determine explicitly how much overlap there is in this explained variation. Of course, in order to do this, we have to be explicit about what we mean by “variation”, so that is where (and why) the choice of resemblance measure becomes so important.

Of course, DISTLM also has some clear limitations. First, despite the flexibility afforded by being able to choose any resemblance measure we wish as the basis of the analysis (so the models are usually not at all linear with respect to the original \(Y\) variables in the majority of cases), these models are strictly linear in the \(X\) variables. We can use polynomials of the \(X\) variables to get around this to some extent, but this is not the only potential issue. Another is that DISTLM’s reliance on the traditional partitioning approach means that we can run out of degrees of freedom if there are more predictor variables than there are samples. More particularly, in order to get sensible results, the largest possible full model is restricted to having \(q \leq N - 1\), at most. This is a simple consequence of it being possible to perfectly fit a linear model with \((N - 1)\) parameters (variables) to \(N\) points \((R^2 = 1.0)\). Although we can use criteria that are not strictly monotonic on \(R^2\) with increases in predictor variables (such as adjusted \(R^2\), \(AIC\), \(AIC\), or \(BIC\)), which will certainly help to find parsimonious models, all of the models fit by DISTLM partition the total variation using a linear function of the \(X\)'s and so will necessarily have this restriction of an upper bound on \(q\).

In contrast, the BEST/BIOENV procedure arises out of the purely non-parametric approach inherent in the majority of the routines already available in PRIMER, such as non-metric MDS and ANOSIM. The differences between DISTLM and BIOENV are therefore directly analogous to many of the differences between PERMANOVA and ANOSIM already discussed in chapter 1. In essence, the BEST/BIOENV procedure does not attempt to model the data cloud at all, but rather tries to find the best possible rank-order match between the inter-point dissimilarities and the inter-point distances derived from sets of environmental variables. The criterion used for this matching is either a Spearman or a Kendall rank correlation, so it is only the rank orders of the resemblances that are being considered. There are several advantages to this approach. First, we can have as many variables as we want in either of the original matrices. The “matching” is being done on the basis of rank resemblances only, so there is simply no limit to how many original variables may be used to calculate either the species resemblances or environmental distance matrices. Second, the rank correlation (whether we use Spearman, weighted Spearman or Kendall) yields a criterion for the success of the match which (unlike \(R^2\)) is not monotonically related to the number of variables in the environmental data matrix at all. In fact, the inclusion of variables that do nothing to enhance the match will clearly cause a decrease in rank correlation. This criterion has intuitive appeal for identifying parsimonious sets of environmental variables that produce patterns among samples that are similar to the patterns produced among those same samples using the biotic data. Furthermore,

---

\(^7\) For a discussion of outliers and high leverage points in multiple regression, see for example Neter et al. (1996).
the permutation test associated with the BEST/BIOENV routine includes the selection step with each permutation. This is really rather neat and allows the user validly to test the significance of the relationship between the two matrices given that some effort has gone into selecting environmental variables that will provide a good match.

The limitations of the BEST/BIOENV approach become apparent, however, when we realise that, once a purportedly “useful” set of environmental variables have been selected, we are not in a position to say how much of the variation inherent in the species resemblance matrix is “explained” by these variables, either individually or collectively. Such a variance is a function of the precise measurement scale of the resemblances, i.e. is a “metric” concept that cannot be captured by a non-(para)metric approach. The rank correlation between the two resemblance matrices does provide a valuable non-parametric index of how closely the collective set of environmental variables captures the multivariate pattern of the species variables (on a scale from $\rho \approx 0$ to 1), and this is an index with an absolute validity in comparisons across different transformations, resemblance measures, etc. (as with the similarly rank-based ANOSIM R statistic). However, it does not directly provide a quantitative measure of the relative importance of the individual environmental variables that have been selected; this can only be inferred by comparing the match ($\rho$) to the multivariate species cloud for different subsets of these environmental variables. Most tellingly, it cannot provide sequential (partial) tests, i.e. of the statistical significance of adding (or deleting) an explanatory variable from the current set. In other words, by going “non-parametric” (BEST/BIOENV), we relinquish our ability to explicitly measure and model multivariate variation. On the other hand, if we want to create such a model (DISTLM), then we must define what we mean by “multivariate variation” and decide how we are going to model it. This requires some decisions (e.g., which resemblance measure shall I use?) and some model assumptions (e.g., fitting linear combinations of predictor variables, and that the residual variability is additive and homogeneous across the different levels of the predictor variables). In short, we believe that DISTLM retains much of the flexibility of the non-parametric approach by allowing any (reasonable) resemblance measure to define what we mean by “multivariate variation”. However, in order to take the step towards formally modeling this variation, we are forced to let go of the fully non-parametric (and completely assumption-free) setting. Nevertheless, by using permutation procedures for the marginal and sequential tests, these additional assumptions can, however, entirely avoid being distributional.

---

98 By “reasonable” we generally mean a measure that fulfills at least the first 3 of the 4 criteria of a metric distance measure (see the section Negative eigenvalues in chapter 3) and also one which is meaningful to the researcher for interpretation.
CAP is a routine for performing canonical analysis of principal coordinates. The purpose of CAP is to find axes through the multivariate cloud of points that either: (i) are the best at discriminating among a priori groups (discriminant analysis) or (ii) have the strongest correlation with some other set of variables (canonical correlation). The analysis can be based on any resemblance matrix of choice. The routine begins by calculating principal coordinates from the resemblance matrix among $N$ samples and it then uses these to do one of the following: (i) predict group membership; (ii) predict positions of samples along some other single continuous variable ($q = 1$); or (iii) find axes having maximum correlations with some other set of variables ($q > 1$). There is a potential problem of overparameterisation here, because if we have ($N – 1$) PCO axes, they will clearly be able to do a perfect job of modeling $N$ points. To avoid this problem, diagnostics are required in order to choose an appropriate subset of PCO axes (i.e., $m < (N – 1)$) to use for the analysis. The value of $m$ is chosen either: (i) by maximising a leave-one-out allocation success to groups or (ii) by minimising a leave-one-out residual sum of squares in a canonical correlation. These diagnostics and an appropriate choice for $m$ are provided by the routine. The routine also can perform a permutation test for the significance of the canonical relationship, using either the trace (sum of canonical eigenvalues) or the first canonical eigenvalue as test statistics. Also provided is a plot of the canonical axis scores. A new feature is the capacity to place new observations into the canonical space, based only on their resemblances with prior observations. For a discriminant-type analysis, this includes the allocation of new observations to existing groups.

In some cases, we may know that there are differences among some pre-defined groups (for example, after performing a cluster analysis, or after obtaining a significant result in PERMANOVA), and our interest is not so much in testing for group differences as it is in characterising those differences. With CAP, the central question one asks is: can I find an axis through the multivariate cloud of points that is best at separating the groups? The motivation for the CAP routine arose from the realisation that sometimes there are real differences among a priori groups in multivariate space that cannot be easily seen in an unconstrained ordination (such as a PCA, PCO or MDS plot). This happens primarily when the direction through the data cloud that distinguishes the groups from one another is fundamentally different from the direction of greatest total variation across the data cloud.

For example, suppose we have two groups of samples in a two-dimensional space, as shown in Fig. 5.1. These data consist of two morphometric measurements (the distance of the transverse groove from the posterior border of the prothorax, in microns, and the length of the elytra, in 0.01 mm) on two species of flea-beetle: Haltica oleracea (19 individuals) and H. carduorum (20 individuals). For simplicity in what follows, both variables were normalised. The data are due to Lubischew (1962), appeared in Table 6.1 of Seber (1984) and (along with 2 other morphometric variables) can be found in the file flea.pri in the ‘FleaBeet’ folder of the ‘Examples add-on’ directory.

Clearly there is a difference in the location of these two data clouds (corresponding to the two species) when we view the scatter plot of the normalised variables (Fig. 5.1). Imagine for a moment, however, that we cannot see two dimensions, but can only see one. If we were to draw an axis through this cloud of points in such a way that maximises total variation (i.e., a principal component), then, by projecting the samples onto that axis, we would get a mixture of samples from the two groups along the single dimension, as shown on the left-hand side of Fig. 5.1. If patterns along this single dimension were all we could see, we would not suppose there was any difference between these two groups.

Next, suppose that, instead, we were to consider drawing an ordination that utilises our hypothesis of the existence of two groups of samples in some way. Suppose we were to draw an axis through this cloud of points in such a way that maximises the group differences. That is, we shall find an axis that is best at separating the groups. For this example, the direction of this axis is clearly quite different from the one that maximises total overall variation. Projecting the points onto this new
axis shows, in the reduced space of one dimension, that there is indeed a clear separation of these two groups in the higher-dimensional space, even though this was not clear in the one-dimensional (unconstrained) ordination that was done in the absence of any hypothesis (Fig. 5.1).

Note that the cloud of points has not changed at all, but our view of that data cloud has changed dramatically, because our criterion for drawing the ordination axis has changed. We might liken this to visiting a city for the first time and deciding which avenue we wish to stroll down in order to observe that city. Clearly, our choice here is going to have consequences regarding what kind of “view” we will get of that city. If we choose to travel in an east-west direction, we will see certain aspects of the buildings, parks, landmarks and faces of the city. If we choose a different direction (axis) to travel, however, we will likely see something very different, and maybe get a quite different impression of what that city is like. Tall buildings might obscure certain landmarks or features from us when viewed from certain directions, and so on. Once again, the city itself (like points in a multi-dimensional space) has not really changed, but our view of it can depend on our choice regarding which direction we will travel through it.

More generally, it is clear how this kind of phenomenon can also occur in a higher-dimensional context and on the basis of some other resemblance measure (rather than the 2-d Euclidean space shown in Fig. 5.1). That is, we may not be able to see group differences in two or three dimensions in an MDS or PCO plot (both of these are unconstrained ordination methods), but it may well be possible to discriminate among groups along some other dimension or direction through the multivariate data cloud. CAP is designed precisely to do this in the multivariate space of the resemblance measure chosen.

The other primary motivation for using CAP is as a tool for classification or prediction. That is, once a CAP model has been developed, it can be used to classify new points into existing groups. For example, suppose I have classified individual fish into one of three species on the basis of a suite of morphometric measures. Given a new fish that has values for each of these same measures, I can allocate or classify that fish into one of the groups, using the CAP routine. Similarly, suppose
I have done a cluster analysis of species abundance data on the basis of the Bray-Curtis resemblance measure and have identified that multivariate samples occur in four different distinct communities. If I have a new sample, I might wish to know – in which of these four communities does this new sample belong? The CAP routine can be used to make this prediction, given the resemblances between the new sample and the existing ones.

Another sense in which CAP can be used for prediction is in the context of canonical correlation. Here, the criterion being used for the ordination is to find an axis that has the strongest relationship with one (or more) continuous variables (as opposed to groups). One can consider this a bit like putting the DISTLM approach rather on its head. The DISTLM routine asks: how much variability in the multivariate data cloud is explained by variable $X$? In contrast, the CAP routine asks: how well can I predict positions along the axis of $X$ using the multivariate data cloud? So, the essential difference between these two approaches is in the role of these two sets of variables in the analysis. DISTLM treats the multivariate data as a response data cloud, whereas in CAP they are considered rather like predictors instead. Thus, for example, I may use CAP to relate multivariate species data to some environmental variable or gradient (altitude, depth, percentage mud, etc.). I may then place a new point into this space and predict its position along the gradient of interest, given only its resemblances with the existing samples.

### Details of CAP and how it is related to other methods

Details of CAP and how it is related to other methods are provided by Anderson & Robinson (2003) and Anderson & Willis (2003). In brief, a classical canonical analysis is simply done on a subset of the PCO axes. Here, we provide a thumbnail sketch to outline the main features of the analysis. The important issues to keep in mind are the conceptual ones, but a little matrix algebra is included here for completeness. Let $D$ be an $N \times N$ matrix of dissimilarities (or distances) among samples. Let $X$ be an $N \times q$ matrix that contains either codes for groups (for discriminant analysis) or one or more quantitative variables of interest (for canonical correlation analysis).

Conceptually, we can consider $X$ to contain the hypothesis of interest. When using CAP for discriminant analysis in the PERMANOVA+ add-on package for PRIMER, groups are identified by a factor associated with the resemblance matrix. The CAP routine will internally generate an appropriate $X$ matrix specifying the group structure from this factor information, so no additional data are required. However, if a canonical correlation-type analysis is to be done, then a separate data sheet containing one or more $X$ variables (and having the same number and labels for the samples as the resemblance matrix) needs to be identified. The mechanics of performing a CAP analysis are described by the following steps (Fig. 5.2):

1. First, principal coordinates are obtained from the resemblance matrix to describe the cloud of multivariate data in Euclidean space (see Fig. 3.1 for details). The individual PCO axes are not, however, standardised by their eigenvalues. Instead, they are left in their raw (orthonormal) form, which we will denote by $Q^0$. This means that not only is each PCO axis independent of all other PCO axes, but also each axis has a sum-of-squares (or length) equal to 1. So, the PCO data cloud is effectively “sphericised” (see the section entitled Sphericising variables below).

2. From matrix $X$, calculate the “hat” matrix $H = X[X'X]^{-1}X'$. This is the matrix derived from the solutions to the normal equations ordinarily used in multiple regression (e.g., Johnson & Wichern 1992, Neter et al. 1996). Its purpose here is to orthonormalise (“sphericise”) the data cloud corresponding to matrix $X$ as well.

3. If the resemblance matrix is $N \times N$, then there will be, at most, $(N - 1)$ non-zero PCO axes. If we did the canonical analysis using all of these axes, it would be like trying to fit a model to $N$ points using $(N - 1)$ parameters, and the fit would be perfect, even if the points were completely random and the hypothesis were false! So, only a subset of $m < (N - 1)$ PCO axes should be used, denoted by $Q^0_m$. The value of $m$ is chosen using appropriate diagnostics (see the section Diagnostics, below).

---

99 If a resemblance matrix of similarities is available instead, then the CAP routine in PERMANOVA+ will automatically transform these into dissimilarities before proceeding; the user need not do this as a separate step.

100 Orthonormal axes are uncorrelated and have a sum-of-squares and cross-products matrix (SSCP) equal to the identity matrix $I$ (all sums of squares $= 1$ and all cross-products $= 0$).

101 CAP automatically centres the $X$ data cloud.
5. CAP

4. A classical canonical analysis is done to relate the subset of $m$ orthonormal PCO axes to $X$. This is done by constructing the matrix $Q_m^0HQ_m^0$. Eigenvalue decomposition of this matrix yields canonical eigenvalues $\delta_1^2, \delta_2^2, \ldots, \delta_s^2$ and their associated eigenvectors. The trace of matrix $Q_m^0HQ_m^0$ is equal to the sum of the canonical eigenvalues. These canonical eigenvalues are also the squared canonical correlations. They indicate the strength of the association between the data cloud and the hypothesis of interest.

5. The canonical coordinate axis scores $C$, a matrix of dimension $(N \times s)$, are used to produce the CAP plot. These are made by pre-multiplying the eigenvectors by $Q_m^0$ and then scaling (multiplying) each of these by the square root of their corresponding eigenvalue. Thus, the CAP axes are linear combinations of the orthonormal PCO axes.

Fig. 5.2. Schematic diagram of the steps involved in performing a CAP analysis.

The number of canonical axes produced by the analysis ($= s$) will be the minimum of $(m, q, (N-1))$. For a canonical correlation-type analysis, $q$ is the number of variables in $X$. For a discriminant analysis, $q = (g-1)$, where $g$ is the number of groups. If the analysis if based on Euclidean distances to begin with, then CAP is equivalent to classical canonical discriminant analysis (CDA) or canonical correlation analysis (CCorA). In such cases, the number for $m$ should be chosen to be the same as the number of original variables ($p$) in data matrix $Y$, except in the event that $p$ is larger than $(N-1)$, in which case the usual diagnostics should be used to choose $m$.

We will begin with an example provided by Trevor Willis and Chris Denny (Willis & Denny 2000; Anderson & Willis 2003), examining temperate reef fish assemblages at the Poor Knights Islands, New Zealand. Divers have counted the abundances of fish belonging to each of $p = 62$ species in each of nine $25 \times 5$ m transects at each site. Data from the transects were pooled at the site level and a number of sites around the Poor Knights Islands were sampled at each of three different times: September 1998 ($n_1 = 15$), March 1999 ($n_2 = 21$) and September 1999 ($n_3 = 20$). These times of sampling span the point in time when the Poor Knights Islands were classified as a no-take marine reserve (October 1998). Interest lies in distinguishing among the fish assemblages observed at these three different times of sampling, especially regarding any transitions between the first time of sampling (before the reserve was established) and the other two times (after).
The data are located in the file `pkfish.pri` in the ‘PKFish’ folder of the ‘Examples add-on’ directory. An unconstrained PCO on Bray-Curtis resemblances of log(X+1)-transformed abundances did not show a clear separation among the three groups, even though a PERMANOVA comparing the three groups was highly statistically significant (Fig. 5.3, \( P < 0.001 \)). Dr Willis (quite justifiably) said: “I don’t see any differences among the groups in either a PCO or MDS plot, but the PERMANOVA test indicates strong effects. What is going on here?” Indeed, the reason for the apparent discrepancy is that the directions of the differences among groups in the multivariate space, detected by PERMANOVA, are quite different to the directions of greatest total variation across the data cloud, as shown in the PCO plot. The relatively small amount of the total variation captured by the PCO plot (the first two axes explain only 33.1%) is another indication that there is more going on in this data cloud than can be seen in two (or even three) dimensions\(^{102}\).

![Poor Knights Islands Fish](image.png)

**Fig. 5.3.** PCO ordination and one-way PERMANOVA of fish data from the Poor Knights Islands.

To characterise these three groups of samples, to visualise the differences among them and to assess just how distinct these groups are from one another in the multivariate space, a CAP analysis can be done (Fig. 5.4). From the Bray-Curtis resemblance matrix calculated from log(X+1)-transformed data, choose **PERMANOVA+ > CAP > (Analyse against Groups in factor) & (Factor for groups or new samples: Time) & (Diagnostics Do diagnostics)**, then click ‘OK’.

The resulting constrained CAP plot (Fig. 5.5) is very different from the unconstrained PCO plot (Fig. 5.3). The constrained analysis shows that the three groups of samples (fish assemblages at three different times) are indeed distinguishable from one another. The number of CAP axes produced is \( s = \min(m, q, (N - 1)) = 2 \), because in this case, \( q = (g - 1) = 2 \); only 2 axes are needed to distinguish among three groups. The sizes of each of these first two canonical correlations are reasonably large: \( \delta_1 = 0.78 \) and \( \delta_2 = 0.69 \) (Fig. 5.5). These canonical correlations indicate the strength of the association between the multivariate data cloud and the hypothesis of group differences. For these data, the first canonical axis separates the fish assemblages sampled in September 1998 (on the right) from those sampled in March of 1999 (on the left), while the second canonical axis separates fish assemblages sampled in September 1999 (lower) from the other two groups (upper). Also reported in the output file is the choice for the number of orthonormal PCO axes that were used for the CAP analysis: here, \( m = 7 \). If this value is not chosen by the user, then similar indications were apparent using MDS. A two-dimensional MDS of these data does not show clear differences among the groups, but has high stress (0.25). The 3-d solution (with stress also quite high at 0.18) does not show any clear differences among these three groups either, though ANOSIM, like PERMANOVA, gives clear significance to the groups (but with a low \( R \) of 0.13).

\(^{102}\) Similar indications were apparent using MDS. A two-dimensional MDS of these data does not show clear differences among the groups, but has high stress (0.25). The 3-d solution (with stress also quite high at 0.18) does not show any clear differences among these three groups either, though ANOSIM, like PERMANOVA, gives clear significance to the groups (but with a low \( R \) of 0.13).
the CAP routine itself will make a choice, based on appropriate diagnostics (see the section Diagnostics below).

Fig. 5.4. Dialog for the CAP analysis of fish data from the Poor Knights Islands.

Fig. 5.5. Constrained ordination and part of the output file from the CAP analysis of fish data from the Poor Knights Islands.
A natural question to ask is *which* fish species characterise the differences among groups found by the CAP analysis? A simplistic approach to answer this is to superimpose vectors corresponding to Pearson (or Spearman rank) correlations of individual species with the resulting CAP axes. Although such an approach is necessarily *post hoc*, it is quite a reasonable approach in the case of a discriminant-type CAP analysis. The CAP axes have been expressly drawn to separate groups as well as possible, so indeed any variables which show either an increasing or decreasing relationship with these CAP axes will be quite likely to be the ones that are more-or-less responsible for observed differences among the groups.

From the CAP plot, choose **Graph > Special** > (Vectors •Worksheet variables: pkfish > Correlation type: Spearman), then click on the ‘Select’ button and in the ‘Select Vectors’ dialog, choose •Correlation > 0.3 and click ‘OK’. Note that you can choose any cut-off here that seems reasonable for the data at hand. The default is to include vectors for variables having lengths of at least 0.2, but it is up to the user to decide what might be appropriate here. Note also that these correlations are for *exploratory* purposes only and are not intended to be used for hypothesis-testing. For example, it is clearly *not* appropriate to formally test the significance of correlations between individual species and CAP axes that have been derived from resemblances calculated using those very same species; this is circular logic! For more detailed information on how these vectors are drawn, see the section **Vector overlays** in chapter 3.

![Vector overlay of Spearman rank correlations of individual fish species with the CAP axes (restricted to those having lengths > 0.30).](image)

For the fish assemblages from the Poor Knights (Fig. 5.6), we can see that some species apparently increased in abundance after the establishment of the marine reserve, such as the snapper *Pagrus auratus* (‘PAGRUS’) and the kingfish *Seriola lalandi* (‘SERIOLA’), which are both targeted by recreational and commercial fishing, and the stingrays *Dasyatis thetidis* and *D. brevicaudata* (‘DTHET’, ‘DBREV’). Vectors for these species point toward the upper left of the diagram in Fig. 5.6 indicating that these species were more abundant, on average, in the March 1999 samples (the group located in that part of the diagram, see Fig. 5.5). Some species, however, were more abundant before the reserve was established, including leatherjackets *Parika scaber* (‘PARIKA’) and the (herbivorous) butterfish *Odax pullus* (‘ODAX’). These results lead to new ecological hypotheses that might be investigated by targeted future observational studies or experiments.

**Diagnostics**

How did the CAP routine choose an appropriate number of PCO axes to use for the above discriminant analysis (*m* = 7)? The essential idea here is that we wish to include as much of the original variability in the data cloud as possible, but we do not wish to include PCO axes which do
nothing to enhance our ability to discriminate the groups. To do so would be to look at the problem with “rose-coloured” glasses. Indeed, if we were to use all of the PCO axes, then the canonical plot will just show our hypothesis back at us again, which is useless. Recall that in CAP, we are using the PCO axes to predict the groups, and a linear combination of \((N - 1)\) PCO axes can easily be used to plot \(N\) points according to any hypothesis provided in \(X\) perfectly. For example, if we choose \(m = 55\) for the above example (as \(N = 56\), we will see a CAP plot with the samples within each of the three different groups superimposed on top of one another onto three equidistant points.

This tells us nothing, however, about how distinct the groups are in the multivariate space, nor how well the PCO axes model and discriminate among the groups, particularly given a new observation, for example. One appropriate criterion for choosing \(m\) is to choose the number of axes where the probability of misclassifying a new point to the wrong group is minimised.

To estimate the misclassification error using the existing data, we can use a leave-one-out procedure (Lachenbruch & Mickey 1968, Seber 1984). The steps in this procedure are:

1. Take out one of the points.
2. Do the CAP analysis without it.
3. Place the “left-out” point into the canonical space produced by the others.
4. Allocate the point to the group whose centroid (middle location) is closest to it.
5. Was the allocation successful? (i.e., did the CAP model classify the left-out point to its correct group?)
6. Repeat steps 1-5 for each of the points.
7. Misclassification error = the proportion of points that were misclassified. The proportion of correct allocations is called allocation success (= 1 minus the misclassification error).

By repeating this for each of a series of values for \(m\), we can choose to use the number of axes that minimises the misclassification error. Although not a strict requirement, we should also probably choose \(m\) to be at least as large as to include \(\sim 60\%\) of the variation in the original resemblance matrix (if not more). The CAP routine will go through the above outlined leave-one-out procedure for each value of \(m\) and will show these diagnostics in the output file\(^\text{103}\). For discriminant-type analyses, a value of \(m\) is chosen which maximises the allocation success (i.e., minimises the misclassification error). For canonical correlation-type analyses, the value of \(m\) is chosen which minimises the leave-one-out residual sum of squares. The user can also choose the value of \(m\) manually in the CAP dialog, with the option to do the diagnostics for this chosen value of \(m\) alone or for all values of \(m\) up to and including the value chosen (Fig. 5.4).

An output table of diagnostics is shown for the Poor Knights fish dataset in Fig. 5.7. For each value of \(m\), the diagnostic values given are:

- ‘prop.G’ = the proportion of variation in the data cloud described by the resemblance matrix (represented in matrix \(G\)) explained by the first \(m\) PCO axes;
- ‘ssres’ = the leave-one-out residual sum of squares;
- ‘\(d_{1}^2\)’ = the size of the first squared canonical correlation (\(\delta_{1}^2\));
- ‘\(d_{2}^2\)’ = the size of the second\(^\text{104}\) squared canonical correlation (\(\delta_{2}^2\));
- ‘\%correct’ = the percentage of the left-out samples that were correctly allocated to their own group using the first \(m\) PCO axes for the model.

It often helps to visualise these different diagnostics as a function of \(m\), as shown in Fig. 5.8. First, we can see that the proportion of variation in the original data cloud that is explained by the first \(m\) PCO axes (Fig. 5.8a) is a smooth function of \(m\), which is no surprise (as we saw before, the first 2 PCO axes explain 33.1% of the variation, see Fig. 5.3). Next, we can see that the values of the squared canonical correlations (\(\delta_{1}^2\) and \(\delta_{2}^2\)), also just continue to increase, the more axes we

\(^{103}\) The diagnostics will stop when CAP encounters \(m\) PCO axes that together explain more than 100% of the variation in the resemblance matrix, as will occur for systems that have negative eigenvalues (e.g., see Fig. 3.5). In the present case, the diagnostics do not extend beyond \(m = 15\). Examination of the PCO output file for these data shows that the first 16 PCO axes explain 100.58% of the variation in \(tr(G)\).

\(^{104}\) There will be as many columns here as required to show all of the canonical eigenvalues – in the present example, there are only two canonical eigenvalues, so there are two columns for these.
choose to use. However, they do appear to pretty much “level-off” after \( m = 6 \) or 7 axes (Fig. 5.8b). So, we don’t get large increases or improvements in either of the canonical correlations by including more than \( m = 7 \) axes. Unlike the canonical correlations, neither the leave-one-out allocation success, nor the leave-one-out residual sum of squares has a monotonic relationship with \( m \). Although the leave-one-out residual sum of squares is minimised when \( m = 13 \) (Fig. 5.7), we can see that no great reduction in its value is achieved beyond about \( m = 7 \) or 8 (Fig. 5.8c). Finally, the leave-one-out allocation success was maximised at \( m = 7 \), where 41 out of the 56 samples (73.2%) were allocated to the correct group using the CAP model (Fig. 5.8d). The CAP routine has chosen \( m = 7 \) based on this diagnostic alone, but the other diagnostic information also indicates that a value of \( m \) around 7 indeed would be appropriate for these data. These first 7 PCO axes explain about 71.4% of the total variation in \( tr(G) \) (‘prop.G’, Fig. 5.7). Although we would generally like this figure to be above 60% for the chosen value of \( m \), as it is here, this is not a strict requirement of the analysis, and examples do exist where only the first one or two PCO axes are needed to discriminate groups, even though these may not necessarily include a large proportion of the original total variation.

![CAP](image)

**Fig. 5.7.** Diagnostics and cross-validation results for the CAP analysis of the Poor Knights fish data.

Generally, the CAP routine will do a pretty decent job of choosing an appropriate value for \( m \) automatically, but it is always worthwhile looking at the diagnostics carefully and separately in order to satisfy yourself on this issue before proceeding. In some situations (namely, for data sets with very large \( N \)), the time required to do the diagnostics can be prohibitively long. One possibility, in such cases, is to do a PCO analysis of the data first to get an idea of what range of \( m \) might be appropriate (encapsulating, say, up to 60-80% of the variability in the resemblance matrix), and then to do a targeted series of diagnostics for individual manual choices of \( m \) to get an idea of their behaviour, eventually narrowing in on an appropriate choice.
The procedure of pulling out one sample at a time and checking the ability of the model to correctly classify that sample into its appropriate group is also called cross-validation. An important part of the CAP output from a discriminant type of analysis is the table showing the specific cross-validation results obtained for a chosen value of \( m \). This gives specific information about how distinct the groups are and how well the PCO axes discriminate among the groups. No matter what patterns seem to be apparent from the CAP plot, nor how small the \( P \)-value from the permutation test (see the following section), this table of cross-validation results is actually the best way to assess the validity and utility of the CAP model. Indeed, we suggest that when using CAP for discrimination, no CAP plot should be presented without also providing cross-validation results, or at least providing the figure for overall misclassification error (or, equivalently, allocation success). This is because the CAP plot will look better and better (i.e., it will look more and more in tune with the hypothesis) the more PCO axes we choose to use. This does not mean, however, that the predictive capability of the underlying CAP model is improved! Indeed, we have just seen in the previous example how increases in the number of PCO axes (beyond \( m = 7 \)) actually reduces the allocation success of the model. So, the cross-validation provides a necessary check on the potential arbitrariness of the results.

Furthermore, the more detailed cross-validation results provided in the CAP output provide information about which groups are more distinct than others. Although, in this case, the groups had roughly comparable mis-classification errors (\( \sim 70 \text{-} 76\% \), see Fig. 5.7), these errors can sometimes vary quite widely among the groups. The output file also indicates in which direction mistakes are made and for which individual samples this occurred. For example, looking at the cross-validation table for the Poor Knights fish data, 4 of the 15 samples from September 1998 were incorrectly classified as belonging to the group sampled in September 1999, while none were incorrectly classified as belonging to the group sampled in March 1999. Furthermore, the individual samples that were mis-classified (and the group into which they were erroneously allocated) are shown directly under the summary table. For example, the samples numbered 1, 2, 4 and 15 were the particular ones from September 1998 that were mis-classified (Fig. 5.7).

As a rule of thumb, bear in mind that, with three groups, one would expect an allocation success of around 33.3% simply by chance alone. Similarly, one would expect an allocation success of around 50% by chance in the case of two groups, or 25% in the case of 4 groups, etc. If the allocation success is substantially greater than would be expected by chance (as is the case for the Poor Knights data), then the CAP model obtained is a potentially useful one for making future predictions and allocations. Thus, the results of the cross-validation give a direct measure of the
relative distinctiveness of the groups and also the potential utility of the model for future classification or prediction.

CAP can be used to test for significant differences among the groups in multivariate space. The test statistics in CAP are different from the pseudo-$F$ used in PERMANOVA. Instead, they are directly analogous to the traditional classical MANOVA test statistics, so we will demonstrate them here in an example of classical canonical discriminant analysis (CDA). The data were obtained by Edgar Anderson (1935) and were first used by Sir R. A. Fisher (1936) to describe CDA (sometimes also called canonical variate analysis). Data are located in the file `iris.pri` in the ‘Irises’ folder of the ‘Examples add-on’ directory. The samples here are individual flowers. On each flower, four morphometric variables were measured (in cm): petal length (PL), petal width (PW), sepal length (SL) and sepal width (SW). There were 150 samples in total, with 50 flowers belonging to each of 3 species: *Iris versicolor* (C), *Iris virginica* (V) and *Iris setosa* (S). Interest lies in using the morphometric variables to discriminate or predict the species to which individual flowers belong.

A traditional canonical discriminant analysis is obtained by running the CAP routine on the basis of a Euclidean distance matrix and manually choosing $m = p$, where $p$ is the number of variables in the original data file. The first $m$ PCO axes will therefore be equivalent to PCA axes and these will contain 100% of the original variation in the data cloud. Even if the original variables are on different units or scales, there is no need to normalise the data before proceeding with the CAP analysis, as this is automatically ensured by virtue of the fact that CAP uses orthonormal PCO axes. In some situations, the number of original variables will exceed (or come close to) the total number of samples in the data file (i.e., $p$ approaches or exceeds $N$). In such cases, it is appropriate to use the leave-one-out diagnostics to choose $m$, just as would be done for non-Euclidean cases. An alternative would be to choose a subset of original variables upon which to base the analysis (by removing, for example, strongly correlated variables). Here, the usual issues and caveats regarding variable selection for modelling (see chapter 4) come into play.

![Anderson's Iris data](image)

**Fig. 5.9.** Canonical ordination for the discriminant analysis of Anderson’s Iris data.

To proceed with a traditional analysis of the iris data, calculate a Euclidean distance matrix directly from the data and then choose PERMANOVA+ > CAP > (Analyse against ●Groups in factor) & (Factor for groups or new samples: Flower) & (Specify $m$) & (Diagnostics Do diagnostics > ●Chosen $m$ only) & (Do permutation test > Num. permutations: 9999), then click ‘OK’. We have chosen $m = 4$ here, because we wish to obtain the classical analysis and there were 4 original variables (PL, PW, SL and SW). The results show that the first squared canonical correlation is very large ($\delta_1^2 = 0.97$) and indeed the first canonical axis does quite a good job of separating the three iris species from one another (Fig. 5.9). The second canonical axis has a much smaller
eigenvalue ($\delta^2 = 0.22$), and actually there is no clear separation of the groups along this second axis. The role of the original variables can be visualised by superimposing a vector overlay (shown in the inset of Fig. 5.9) using the option Graph > Special > (Vectors • Worksheet variables: iris > Correlation type: Multiple). These vectors show relationships between each of the original individual variables and the CAP axes, taking into account the other three variables in the worksheet (see the section Vector overlays in dbRDA in chapter 4). For these data, petal width and sepal length appear to play fairly important roles in discriminating among the groups. A draftsman plot of the original variables reveals that sepal length and width are highly correlated with one another ($r = 0.96$) and petal length is also fairly highly correlated with each of these ($r > 0.8$), so it is not terribly surprising that PL and SW play more minor roles once SL is taken into account.

Diagnostics show that the choice of $m = 4$ PCO axes includes 100% of the original variation (‘prop.G’ = 1) and that the leave-one-out allocation success was quite high using the canonical model: 93.3% of the samples (140 out of 150) were correctly classified (Fig. 5.10). The most distinct group, which had 100% success under cross-validation, was Iris setosa, whereas the other two species, Iris versicolor and Iris virginica, were a little less distinct from one another, although their allocation success rates were still admirably large (at 92% and 88%, respectively). These results regarding the relative distinctiveness of the groups coincide well with what can be seen in the CAP plot (Fig. 5.9), where the Iris setosa group is indeed easily distinguished and well separated from the other two groups along the first canonical axis.

The results from the permutation tests are shown at the very bottom of the CAP output file (Fig. 5.10). There are two test statistics that are given in the output for the test. The first is a “trace” test statistic. It is the sum of the canonical eigenvalues (i.e., the sum of the squared canonical correlations) or the trace of the matrix $Q_m' HQ_m$ (denoted in the output text by ‘tr(Q_m'HQ_m)’). When a CAP analysis is based on Euclidean distance and $m = p$, then this is equivalent to the traditional MANOVA test statistic known as Pillai’s trace\[105\]. The other test statistic provided in the CAP output is simply the first canonical eigenvalue, which is the first squared canonical correlation, $\delta^2_1$ (denoted in the output text by ‘delta_1^2’). This test statistic is directly related to a

---

\[105\] This equivalence is readily seen by doing a traditional MANOVA using some other statistical package on the Iris data, where the output for Pillai’s trace will be given as 1.1919, the value shown for the trace statistic from the CAP analysis in Fig. 5.10.
statistic called Roy’s greatest root criterion in traditional MANOVA. More specifically, Roy’s criterion is equal to $\delta^2 / (1 - \delta^2)$ when $\delta^2$ is obtained from a CAP based on Euclidean distances and when $m = p$. There are other MANOVA test statistics (i.e., Wilks’ lambda, Hotelling-Lawley trace, see Mardia et al. 1979, Seber 1984, Rencher 1998). Studies show that these different MANOVA test statistics differ in their power to detect different kinds of changes among the group centroids in multivariate space (e.g., Olson 1974, 1975, Rencher 1998). Olson (1974, 1975, 1976) suggested that Pillai’s trace, although not the most powerful in all situations, did perform well in many cases, and importantly, was quite robust to violations of its assumptions, maintaining type I error rates at nominal levels in the face of non-normality or mild heterogeneity of variance-covariance matrices. It is well known that Roy’s criterion, however, will be the most powerful for detecting changes in means along a single axis in the multivariate space (Seber 1984, Rencher 1998). Generally, we suggest that the trace criterion will provide the best approach for the widest range of circumstances, and should be used routinely, while the test using the first canonical eigenvalue will focus specifically on changes in centroids along a single dimension, where this is of interest. Of course, the two test statistics will be identical when there is only canonical axis (e.g., two groups).

The permutation test in CAP assumes only exchangeability of the samples under a true null hypothesis of no differences in the positions of the centroids among the groups in multivariate space (for a given chosen value of $m$). Thus, although the values of the trace statistic and the first squared canonical correlation are directly related to Pillai’s trace and Roy’s criterion, respectively (when Euclidean distance is used), there are no stringent assumptions about the distributions of the variables: tests by permutation provide an exact test of the null hypothesis of no differences in the positions of centroids among groups. Mardia (1971) proposed a permutation test based on Pillai’s trace (e.g., see Seber 1984), which would be equivalent to the CAP test on the trace statistic when based on Euclidean distances. Of course, the CAP routine (like all of the routines in PERMANOVA+ for PRIMER) also provides the additional flexibility that any resemblance measure can be used as the basis of the analysis.

It might seem confusing that both CAP and PERMANOVA can be used to test for differences among groups in multivariate space. At the very least, it begs the question: which test should one use routinely? The main difference between these two approaches is that CAP is designed to ask: are there axes in the multivariate space that separate groups? In contrast, PERMANOVA is designed to ask: does between-group variation explain a significant proportion of the total variation in the system as a whole? So, CAP is designed to purposely seek out and find groups, even if the differences occur in obscure directions that are not apparent when one views the data cloud as a whole, whereas PERMANOVA is more designed to test whether it is reasonable to consider the existence of these groups in the first place, given the overall variability. Thus, in many applications, it makes sense to do the PERMANOVA first and, if significant differences are obtained, one might consider then using the CAP analysis to obtain rigorous measures of the distinctiveness of the groups in multivariate space (using cross-validation), to characterise those differences (e.g., by finding variables related to the canonical axes) and possibly to develop a predictive model for use with future data.

One can also consider the difference between these two methods in terms of differences in the role that the data cloud plays in the analysis vis-à-vis the hypothesis. For CAP, the hypothesis is virtually a given and the role of the data cloud is to predict that hypothesis (the $X$ matrix plays the role of a response). For PERMANOVA, in contrast, the hypothesis is not a given, and the role of the data cloud is as a response, with the hypothesis ($X$ matrix) playing the role of predictor. This fundamental difference is also apparent when we consider the construction of the among-group SS in PERMANOVA and compare it with the trace statistic in CAP. For PERMANOVA, the among-group SS can be written as $tr(HQH)$ or, equivalently (if there were no negative eigenvalues), we could write: $tr(HQH)$. This is a projection of the variation in the resemblance matrix (represented by $Q$ here) onto the $X$ matrix, via the projection matrix $H$. For PERMANOVA, all of the PCO axes

---

107 The data cloud in the space of the appropriate resemblance measure is represented either by matrix $G$, Gower’s centred matrix, or by matrix $Q$, the matrix of principal coordinate axes.
are utilised and they are each standardised by their eigenvalues ($\lambda$'s, see chapter 3). On the other hand, the trace statistic in CAP is $\text{tr}(Q^0HQ^0)$. The roles of these two matrices ($Q$ and $H$) are therefore swapped. Now we have a projection of the $X$ variables (albeit sphericised, since we are using $H$) into the space of the resemblance matrix instead. Also, the PCO axes in CAP are not scaled by their respective eigenvalues, they are left as orthonormal axes ($SS = 1$). The purpose of this is really only to save a step. When predicting one set of variables using another, the variables used as predictors are automatically sphericised as part of the analysis\textsuperscript{108}. Since the eigenvector decomposition of matrix $G$ produces orthonormal PCO axes $Q^0$, we may as well use those directly, rather than multiplying them by their eigenvalues and then subsequently sphericising them, which would yield the same thing!

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig511.png}
\caption{Schematic diagram of the difference in the scaling of PCO axes for PERMANOVA vs CAP.}
\end{figure}

The difference in the scaling of PCO axes, nevertheless, highlights another difference between CAP and PERMANOVA. In PERMANOVA, the relative importance of each PCO axis is in proportion to its eigenvalue (its variance), whereas in CAP, each PCO is given equal weight in the analysis (Fig. 5.11). The latter ensures that directions that otherwise might be of minor importance in the data cloud as a whole, are given equal weight when it comes to searching for groups. This is directly analogous to the situation in multiple regression, where predictor variables are automatically sphericised (normalised to have $SS = 1$ and rendered independent of one another) when the hat matrix is calculated.

Another way of expressing the ‘sphericising’ being done by CAP is to see its equivalence with the classical MANOVA statistics, where the between-group variance-covariance structure is scaled by a (pooled) within-group variance-covariance matrix (assumed to be homogeneous among groups). The distances between points being worked on by CAP therefore can also be called Mahalanobis distances (e.g., Seber 1984) in the multivariate PCO space because they are scaled in this way. In contrast, to obtain the $SS_A$ of equation (1.3) for PERMANOVA one could calculate the univariate among-group sum of squares separately and independently for each PCO axis and then sum these up\textsuperscript{109}. Similarly, the $SS_{Res}$ of equation (1.3), may be obtained as the sum of the individual within-group sum of squares calculated separately for each PCO axis. PERMANOVA’s pseudo-$F$ is effectively therefore a ratio of these two sums\textsuperscript{110}.

\textsuperscript{108}The reader may recall that, for the very same reason, it makes no difference to a dbRDA whether one normalises the $X$ variables before proceeding or not (see chapter 4). The dbRDA coordinate scores and eigenvalues will be the same. Predictor variables are automatically sphericised as part of multiple regression, as a consequence of constructing the $H$ matrix.

\textsuperscript{109}Note that the $SS$ from PCO axes corresponding to negative eigenvalues would contribute negatively towards such a sum.

\textsuperscript{110}Other test statistics could be used. For example, a sum of the individual $F$ ratios obtained from each PCO axis (e.g., Edgington 1995, which one might call a “stacked” multivariate $F$ statistic), could also be used, rather than the ratio of sums that is PERMANOVA’s pseudo-$F$. 

---

170
The relative robustness and power of PERMANOVA versus CAP in a variety of circumstances is certainly an area warranting further research. In terms of performance, we would expect the type I errors for the CAP test statistic and the PERMANOVA test statistic to be the same, as both use permutations to obtain P-values so both will provide an exact test for the one-way case when the null hypothesis is true. However, we would expect the power of these two different approaches to differ. In some limited simulations (some of which are provided by Anderson & Robinson 2003, and some of which are unpublished), CAP was found to be more powerful to detect effects if they were small in size and occurred in a different direction to the axis of greatest total variation. On the other hand, PERMANOVA was more powerful if there were multiple independent effects for each of a number of otherwise unrelated response variables. Although there is clearly scope for more research in this area, the take-home message must be to distinguish and use these two methods on the basis of their more fundamental conceptual differences (Fig. 5.12), rather than to try both and take results from the one that gives the most desirable outcome!

![Fig. 5.12. Schematic diagram of the conceptual and practical differences, in matrix algebra terms, between the CAP analyses and PERMANOVA (or DISTLM).](image)

**Caveats on using CAP**

When using the CAP routine, it should come as no surprise that the hypothesis (usually) is evident in the plot. Indeed the role of the analysis is to search for the hypothesis in the data cloud. However, once faced with the constrained ordination, one might be tempted to ask the question whether it is useful to examine an unconstrained plot at all. The answer is most certainly ‘yes’! We would argue that it is always important to examine an unconstrained plot (either a PCO or an MDS if dealing with general resemblances, or a PCA if Euclidean distances are appropriate). There are several reasons for this.

First, a CAP plot (or other constrained plot, such as a dbRDA) views the data cloud through the filter of our hypothesis, so-to-speak, which is a bit like viewing the data through “rose-coloured glasses”, as mentioned before. This will have a tendency to lead us down the path towards an over-emphasis on the importance of the hypothesis, and we can easily neglect to put this into a broader perspective of the data cloud as a whole. Although a restriction on the choice of $m$ and the use of cross-validation to avoid over-parameterising the problem will help to reduce this unfortunate instinct in us (i.e., our intrinsic desire to see our hypothesis in our data), the unconstrained plot has the important quality of “letting the data speak for themselves”. As the unconstrained plot does not include our hypothesis in any way to draw the points (but rather uses a much more general
5. CAP

criterion, like preserving ranks of inter-point resemblances, or maximizing total variation across the cloud as a whole), we can trust that if we do happen to see our hypothesis playing a role in the unconstrained plot (e.g., to separate groups), then it is probably a pretty important feature of the data. So, first of all, the unconstrained plot helps to place our hypothesis within a broader perspective. Of course, if we see the separation of groups clearly in the unconstrained plot, then it should come as no surprise that the CAP routine will also have no trouble finding axes that are good at discriminating the groups. Diagnostics (like cross-validation) are crucial, however, for telling us something about the actual distinctiveness of the groups in multivariate space.

**Fig. 5.13.** MDS and CAP plots for a subset of three years’ data on percentage cover of coral species from South Tikus Island, Indonesia.

Second, the purpose of the CAP routine is to seek out separation of the group centroids. As a consequence, it effectively completely ignores, and even destroys, differences in dispersions among the groups. A case in point serves as a useful example here. We shall re-visit the Tikus Island corals dataset (tick.pri in the ‘Corals’ folder of the ‘Examples v6’ directory), consisting of the percentage cover of \( p = 75 \) coral species in each of 6 years from 10 replicate transects in Indonesia (Warwick et al. 1990b). Taking only the years 1981, 1983 and 1985, we observed previously that the dispersions of these groups were very different when analysed using Bray-Curtis resemblances (Fig. 5.13a). However, when we analyse the data using CAP, the first two squared canonical correlations are very large (0.90 and 0.84, respectively), the cross-validation allocation success is impressively high (29/30 = 97%), and the plot shows the three groups of samples as very distinct from one another, with not even a hint of the differences in dispersions that we could see clearly before in the unconstrained MDS plot (Fig. 5.13b). These differences in dispersions are quite real (as verified by the PERMDISP routine), yet might well have remained completely unknown to us if we had neglected to examine the unconstrained plot altogether. An important caveat on the use of the CAP routine is therefore to recognise that CAP plots generally tell us absolutely nothing about relative dispersions among the groups. Only unconstrained plots (and associated PERMDISP tests) can do this.

A new utility of the windows-based version of the CAP routine in PERMANOVA+ is the ability to place new samples onto the canonical axes of an existing CAP model and (in the case of a discriminant analysis) to classify each of those new samples into one of the existing groups. This is done using only the resemblances between each new sample and the existing set of samples that were used to develop the CAP model. First, these inter-point dissimilarities are used to place the new point onto the (orthonormal) PCO axes. It is then quite straightforward to place these onto the canonical axes, which are simply linear combinations of those PCO axes (see Anderson & Robinson 2003 for more details). The only requirement is that the variables measured on each new
sample match the variable list for the existing samples and also that their values occur within the
general multivariate region of the data already observed. For example, suppose we have three new flowers which we suspect belong to one of the three species of irises analysed by CAP in the above section named Test by permutation. Suppose the values of the four morphometric variables for each of these new flowers are:

<table>
<thead>
<tr>
<th></th>
<th>PL</th>
<th>PW</th>
<th>SL</th>
<th>SW</th>
</tr>
</thead>
<tbody>
<tr>
<td>New1</td>
<td>6.3</td>
<td>2.8</td>
<td>5.4</td>
<td>1.9</td>
</tr>
<tr>
<td>New2</td>
<td>4.8</td>
<td>3.5</td>
<td>1.4</td>
<td>0.2</td>
</tr>
<tr>
<td>New3</td>
<td>6.6</td>
<td>3.0</td>
<td>5.7</td>
<td>2.1</td>
</tr>
</tbody>
</table>

Open the file iris.pri (in ‘Examples add-on\irises’) and add these three new samples into the data file (use, for example, Edit > Insert > Row), giving them the names of ‘New1’, ‘New2’ and ‘New3’ and typing in the appropriate values for each variable (Fig. 5.14). Choose Edit > Factors and for the factor named ‘Flower’, we clearly do not know which species these three flowers might belong to yet, so give them the level name of ‘New’, to distinguish them from the existing groups of ‘S’, ‘C’ or ‘V’ (Fig. 5.14). One can enter new samples into an existing data file within PRIMER in this fashion, or include the new samples to be read directly into PRIMER with the original data file. The essential criterion for analysis is that the new samples must have a different level name from the existing groups for the factor which is going to be examined in the CAP discriminant analysis. To add new samples to a canonical correlation-type analysis (see below), a factor must be set up which distinguishes the existing samples from new ones (one can use a factor to distinguish ‘model’ samples from ‘validation’ samples, for example).

---

111 This latter criterion may be very difficult to check. The CAP routine currently does not attempt to identify data points as “outside previous experience” and the development of an appropriate criterion for doing this would be a worthwhile subject for future research.

---

**Fig. 5.14.** Dialog in CAP showing the addition of three new samples (new individual flowers), to be classified into one of the three species groups using the CAP model developed from the existing data.
Once the new samples have been entered and identified as such, the resemblance matrix for all of the samples together must be calculated. For the iris data set, calculate a Euclidean distance matrix. Proceed with the CAP analysis by choosing: PERMANOVA+ > CAP > (Analyse against •Groups in factor) & (Factor for groups or new samples: Flower) & (√ Add new samples > Factor level for new samples New) & (Specify m 4) & (Diagnostics √ Do diagnostics •Chosen m only), then click ‘OK’ (Fig. 5.14). The CAP plot shows the three Iris groups, as before, but the new samples are shown using a separate symbol (Fig. 5.15). The only other difference between the CAP plot in Fig. 5.15 compared to Fig. 5.9 is that the y-axis has been flipped. As with PCA, PCO, dbRDA or MDS ordination plots in PRIMER, the signs of the axes are also arbitrary in a CAP plot. The visual representation of the points corresponding to each of the new flowers have been labeled and from this one might make a guess as to which species each of these new samples is likely to belong.

More detailed information is given, however, in the CAP results file under the heading of ‘New samples’ (Fig. 5.16). First are given the positions of each of the new samples on the canonical axes, followed by the classification of each of the new samples according to these positions. In the present case, the samples ‘New1’ and ‘New3’ were allocated to the group Iris virginica, while the
sample ‘New2’ was allocated to the group *Iris setosa*. Each new sample is allocated to the group whose centroid is the closest to it in the canonical space. For reference, the output file includes these distances to centroids for each sample upon which this decision was made (Fig. 5.16).

So far, the focus has been on hypotheses concerning groups and the use of CAP for discriminant analysis. CAP can also be used to analyse how well multivariate data can predict the positions of samples along a continuous or quantitative gradient. As an example, we shall consider a study of the meiofauna and macrofauna in soft-sediment habitats from five creeks in the Fal estuary along a pollution gradient (Somerfield *et al.* 1994). These creeks have different levels of metal contamination from long-term historical inputs. Original data are in file Fa.xls in the ‘Fal’ folder of the ‘Examples v6’ directory. This file splits the biotic data into separate groups (nematodes, copepods and macrofauna), but here we will analyse all biotic data together as a whole. A single file containing all of the biotic data (with an indicator to identify the variables corresponding to nematodes, copepods and macrofauna) is located in falbio.pri in the ‘FalEst’ folder of the ‘Examples add-on’ directory. Also available in this folder are the environmental variables in file falenv.pri, which includes concentrations for 10 different metals, % silt/clay and % organic matter.

Previous study (e.g., Somerfield *et al.* 1994) indicated a high degree of correlation among the different metals in the field. Based on this high correlation structure, we might well consolidate this information in order to obtain a single gradient in heavy metal pollution among these samples, using principal components analysis (PCA). A draftsman plot also suggests that log-transformed metal variables will produce more symmetric (less-skewed) distributions for analysis. Select only the metals, highlight them and choose **Tools > Transform (individual) > (Expression: log(V)) & (√Rename variables (if unique))** in order to obtain log-transformed metal data. Rename the data sheet log.metals for reference. Next choose **Analyse > Pre-treatment > Normalise variables** and re-name the resultant sheet norm.log.metals. A draftsman plot of these data demonstrates reasonably even scatter and high correlation structure. Choose **Analyse > PCA > (Maximum no of PCs: 5) & (√Plot results) & (√Scores to worksheet)**. The first 2 PC axes explain 93% of the variation in the normalised log metal data. The first PC axis alone explains 77% of the variability, and with approximately equal weighting on almost all of the metals (apart from Ni and Cr). This first PC axis can clearly serve as a useful proxy variable for the overall gradient in the level of metal contamination across these samples. In the worksheet of PC scores, select the single variable with the scores for samples along PC1, duplicate this so it occurs alone in its own data sheet, rename this single variable **PC1** and rename the data sheet containing this variable **poll.grad**.

**Fal Estuary, PCA on metal concentrations (log-transformed, normalised)**

**Fig. 5.17.** PCA of metal concentrations from soft-sediment habitats in creeks of the Fal estuary.
Our interest lies in seeing how well the biotic data differentiate the samples along this pollution gradient. Also, suppose a new sample were to be obtained from one of these creeks at a later date, could the biota from that sample alone be used to place it along this gradient and therefore to indicate its relative degree of contamination, from low to high concentrations of metals? Although the analysis could be done using only a subset of the data (e.g., just the nematodes, for example), we shall use all of the available biotic information to construct the CAP model. Open the file \texttt{falbio.pri} in the same workspace. As indicated in Clarke & Gorley (2006, see pp. 40-41 therein), we shall apply dispersion weighting (Clarke et al. 2006a) to these variables before proceeding. Choose \texttt{Analyse > Pre-treatment > Dispersion weighting} > (Factor: Creek) & (Num perms: 1000). Next transform the data using square-roots and calculate a Bray-Curtis resemblance matrix from the transformed data. We are ready now to run the CAP routine to relate the pollution gradient to the biotic resemblance matrix. Choose \texttt{PERMANOVA+ > CAP > (Analyse against \textbullet Variables in data worksheet: poll.grad) & (Diagnostics \checkmark Do diagnostics)} (Fig. 5.18).

![Fig. 5.18. Dialog and excerpts from the output file of a CAP analysis relating biota from sites in the Fal estuary to the pollution gradient (as represented by PC1 from Fig. 5.17).](image)

In the results file, we should first examine the diagnostics. This is not a discriminant analysis, so there are no groups and thus no cross-validation. Instead, the leave-one-out residual sum of squares is the criterion that was used to decide upon an appropriate value for \( m \) here. The CAP routine has chosen to use \( m = 7 \) PCO axes for the analysis, which encapsulates 81.1\% of the variability in the resemblance matrix, and which indeed minimises ‘ssres’. We can see from the diagnostics that no further substantial increases in the canonical correlation occurs if more PCO axes are included, and the model actually gets worse (the leave-one-out residual SS increases) if we were to include more PCO axes (Fig. 5.18). The choice of \( m = 7 \) therefore appears reasonable.

As there is only one variable in the data file (PC1), there is only one canonical axis. The squared canonical correlation is very high (\( \delta_1^2 = 0.94 \)), suggesting we have a very good model here (Fig. 5.19). When there is only 1 canonical axis, CAP will plot this axis with the original variable in a two-dimensional scatter-plot. It is appropriate that the variable (PC1 in this case, which is a proxy
for overall metal contamination) be positioned on the y-axis, with the canonical scores on the x-axis, as the purpose of CAP is to find an axis through the cloud of multivariate data that is best at predicting this variable. Given a new sample, we can re-run CAP to place that new sample into the canonical space, yielding a position for it along this x-axis, which in turn (via the CAP model) allows a prediction for the position of that point onto the y-axis (the pollution gradient).

![Fal Estuary Biota](image)

**Fig. 5.19.** CAP analysis relating biota from sites in the Fal estuary to the pollution gradient.

**Fig. 5.20.** Dialog and portion of the output for the placement of new points into the canonical analysis.
To see how this works, let us suppose that the samples labeled R2, P2 and E2 had unknown metal concentrations. Accordingly, the environmental data for them would be empty or missing. To place these as ‘new samples’ into the model and predict their position along the pollution gradient, we first need to have a factor that identifies them as new samples. Go to the faibio.pri data sheet and create a new factor called ‘Model’, which has ‘model’ for all of the samples except R2, P2 and E2, which will be identified instead as ‘new’ (Fig. 5.20). Next, go to the poll.grad data sheet and select all of the samples except for R2, P2 and E2 (which we are presuming for the moment to have no value for this variable). Now, from the resemblance matrix (which includes all of the samples) choose PERMANOVSA+ > CAP > (Analyse against •Variables in data worksheet: poll.grad) & (✓ Add new samples > Factor for groups or new samples: Model > Factor level for new samples: new) & (✓ Specify m 7) & (Diagnostics ✓ Do diagnostics > •Chosen m only). Clearly, there is no need to re-do all of the diagnostics here, and we can specify an appropriate value for m (= 7 in this case) directly (Fig. 5.20).

The CAP model for the reduced data set (i.e. minus the three samples that are considered ‘new’) is not identical, but is very similar to the model obtained using all of the data, having the same high canonical correlation of 0.94 for the choice of m = 7. In the CAP output, we are given values for the positions of the new samples along the canonical axis and also, as predicted, along the PC1 gradient (Fig. 5.20). By glancing back at the original plot, we can see that the model, based on the biotic resemblances among samples only, has done a decent job of placing the new samples along the pollution gradient in positions that are pretty close to their actual positions along PC1 (compare Fig. 5.21 with Fig. 5.19).

As highlighted in the section on discriminant analysis, it is instructive also to view the unconstrained data cloud, which helps to place the CAP model into a broader perspective. An unconstrained MDS plot of the biotic data shows that communities from these 5 creeks in the Fal estuary are quite distinct from one another (Fig. 5.22). The differences between assemblages in Restronguet Creek and those in the other creeks are particularly strong. This is not terribly surprising, as this creek has the highest metal concentrations and is directly downstream of the Carnon River, which is the source of most of the heavy metals to the Fal estuary system (Somerfield et al. 1994). Distinctions among the other creeks might include other factors, such as grain size characteristics or %organics. The MDS plot shows differences among these other creeks occur in a direction that is orthogonal (perpendicular) to their differences from Restronguet Creek,
which is an indication of this. Therefore, although it is likely that the strongest gradient in the system is caused by large changes in metal concentrations (seen in both the MDS and the CAP plot), there are clearly other pertinent drivers of variation across this system as well. Although an important strength of the CAP method is its ability to identify and “pull out” and assess useful relationships, even in the presence of significant variation in other directions, this is also something to be wary of, in the sense that one can be inclined to forget about the other potential contributors to variation in the system. A look at the unconstrained ordination is, therefore, always a wise idea.

\[ \text{Fal Estuary Biota} \]

Dispersion weighting
Transform: Square root
Resemblance: S17 Bray Curtis similarity

\[ \text{Creek} \]
\[ \text{Restronguet} \]
\[ \text{Mylor} \]
\[ \text{Pill} \]
\[ \text{St Just} \]
\[ \text{Percuil} \]

2D Stress: 0.11

Fig. 5.22. Unconstrained MDS ordination of all biota from the Fal estuary.

In some cases, interest lies in finding axes through the cloud of points so as to maximise correlation with not just one $X$ variable, but with linear combinations of multiple $X$ variables simultaneously. In such cases, neither of these two sets of variables (i.e. the PCO’s arising from the resemblance matrix, on the one hand, and the $X$ variables, on the other) have a specific role in the analysis – neither set is considered to be either predictors or responses. Rather, when there are several $X$ variables, CAP can be conceptually described as sphericising both sets of variables, and then rotating them simultaneously against one another in order to find axes with maximum inter-correlations between these two sets.

As the agenda here is neither to explain nor predict one set using the other set, canonical correlation analysis with multiple $X$ variables is a method for simply exploring relationships between two sets of variables. As such, its utility is perhaps rather limited for ecological applications, but certainly can be useful for generating hypotheses. CAP does canonical correlation between the PCO axes $Q_m (N \times m)$ and $X (N \times q)$, where $m$ is generally chosen so as to minimise the leave-one-out residual sum of squares (see the section on Diagnostics, above), and the number of canonical axes generated will be $\min(m, q, (N – 1))$. If $p < (N – 1)$ and the measure being used is Euclidean embeddable (e.g., see the section on Negative eigenvalues in chapter 3 herein and Gower & Legendre (1986) regarding the geometric properties of dissimilarity measures), then it makes sense to manually set $m = p$ in the CAP dialog *a priori*, as the dimensionality of PCO’s in such cases is known to be equal to $p$.

Note also that, as CAP will search for linear combinations of $X$ variables that are maximally correlated with the PCO’s, it therefore makes sense to spend a little time examining the distributions of the $X$ variables first (just as in dbRDA) to ensure that they have reasonably symmetric distributions and even scatter (using a draftsman plot for example), to transform them if necessary and also to consider eliminating redundant (very highly correlated) variables. The two sets of variables are treated symmetrically here, so CAP also *simultaneously* searches for linear combinations for the PCO’s that are maximally correlated with the $X$ variables. Thus, one might also consider examining the distributions of the PCO’s (in scatterplot ordinations or a draftsman plot), to ensure that they, too, have fairly even scatter (although certainly no formal assumptions in this regard are brought to bear on the analysis).
It was previously stated that CAP effectively “sphericises” the data clouds as part of the process of searching for inter-correlations between them (e.g., Fig. 5.11). The idea of “sphericising” a set of variables, rendering them orthonormal, deserves a little more attention here, especially in order to clarify further how CAP differs from dbRDA as a method to relate two sets of variables. Let us begin by thinking about a single variable, X. To centre that variable, we would simply subtract its mean. Furthermore, to normalise that variable, we would subtract its mean and divide by its standard deviation. A normalised variable has a mean of zero, a standard deviation of 1 and (therefore) a variance of 1. Often, variables are normalised in order to put them on an “equal footing” prior to multivariate analysis, especially if they are clearly on different scales or are measured in different kinds of units. When we normalise variables individually like this, their correlation structure is, however, ignored.

Now, suppose matrix X has q variables\(^{112}\) that are already centred individually on their means. We might wish to perform a sort of multivariate normalisation in order to standardise them not just for their differing individual variances but also for their correlation (or covariance) structure. In other words, if the original sums-of-squares and cross-products matrix (\(q \times q\)) for X is:

\[
SSCP_X = XX'
\]

we might like to find a sphericised version of X, denoted \(X^0\), such that:

\[
SSCP_{X^0} = X'^0X^0 = I
\]

where I is the identity matrix. In other words, the sum of squares (or length, which is the square root of the sum of squares, Wickens 1995) of each of the new variables in \(X^0\) is equal to 1 and each of the cross products (and hence covariance or correlations between every pair of variables) is equal to zero. In practice, we can find this sphericised matrix by using what is known as the generalised inverse from the singular value decomposition (SVD) of X (e.g., Seber 2008). More specifically, the matrix X (\(N \times q\)) can be decomposed into three matrices using SVD, as follows:

\[
X = UWV'
\]

If \(N > q\) (as is generally the case), then U is an \((N \times q)\) matrix whose columns contain what are called the left singular vectors, \(V\) is a \((q \times q)\) matrix whose columns contain what are called the right singular vectors, and W is a \((q \times q)\) matrix with eigenvalues \((w_1, w_2, \ldots, w_q)\) along the diagonal and zeros elsewhere. The neat thing is, we can now use this in order to construct the generalised inverse of matrix X. For example, to get the inverse of X, defined as \(X^{-1}\) and where \(X'X = I\), we replace W with its inverse \(W^{-1}\) (which is easy, because W is diagonal, so \(W^{-1}\) is a diagonal matrix with \((1/w_1, 1/w_2, \ldots, 1/w_q)\) along the diagonal and zeros elsewhere\(^{113}\)) and we get:

\[
X^{-1} = UW^{-1}V'
\]

Similarly, to get X to the power of zero (i.e., sphericised X), we calculate:

\[
X^0 = UW^0V'
\]

where \(W^0\) is I, the identity matrix, a diagonal matrix with 1’s along the diagonal (because for each eigenvalue, we can write \(w_i^0 = 1\) and zeros elsewhere. Another useful result, which also shows how the hat matrix (whether it is used in dbRDA or in CAP) automatically sphericises the X matrix is:

\[
H = X^0X'^0
\]

If matrix algebra makes you squirm (and don’t worry, you are not alone!), then an actual visual example should help do the trick. Suppose I randomly draw \(N = 20\) samples from a bivariate normal distribution with means of \([10, 7]\), variances of \([10, 2]\) and a covariance of 3. Beginning with a simple scatterplot, we can see a clear positive relationship between these two variables (Fig. 5.23a, the sample correlation here is \(r = 0.7447\)). Now, each variable can be centred on its mean

\(^{112}\) and suppose also that X is of full rank, so none of the q variables have correlation = 1; all are linearly independent

\(^{113}\) If X is of full rank, as previously noted, then none of the w eigenvalues in the SVD will be equal to zero. The CAP routine will take care of situations where there are linear dependencies among the X variables (resulting in some of the eigenvalues being equal to zero) appropriately, however, if these are present.
(Fig. 5.23b), and then the cloud can be sphericised using equation 5.5 (Fig. 5.23c). Clearly, this has resulted in a kind of “ball” of points; the resulting variables (X') each now have a mean of zero, a sum of squares of 1 (and hence a length of 1) and a correlation of zero. In summary, sphericising a data cloud to obtain orthonormal axes can be considered as a kind of multivariate normalisation that both standardises the dispersions of individual variables and removes correlation structure among the variables. CAP sphericises both of the data clouds (X and Q) before relating them to one another. Similarly, a classical canonical correlation analysis sphericises both the X and the Y data clouds as part of the analysis.

Fig. 5.23. Bivariate scatterplot of normal random variables as (a) raw data, (b) after being centred and (c) after being sphericised (orthonormalised).

**CAP vs dbRDA**

So, how does CAP differ from dbRDA for relating two sets of variables? First, dbRDA is directional. Each set of variables has a role as either predictor variables (X) or response variables (Q), while for CAP (when there are multiple variables in X), the two sets of variables are essentially treated symmetrically. Canonical correlation analysis (on the basis of Euclidean distances) finds linear combinations of Y and linear combinations of X that are maximally correlated with one another. In contrast, RDA finds linear combinations of X that are best at explaining or predicting Y (the latter set of variables are not sphericised). Note that canonical correlation sphericises both of the data clouds using X' and Y' in the calculations. This ensures that the correlations among the variables within either set are taken into account. In contrast, RDA uses sphericised X' (by constructing matrix H), but the variables in Y are not sphericised. Now, CAP generalises canonical correlation to any resemblance measure by replacing Y' with Q_m', while dbRDA generalises RDA to any resemblance measure by replacing Y with Q. Furthermore, no matter how many variables occurred in the original Y matrix, over-parameterisation in CAP can be avoided by a prudent choice for m (generally obtained using diagnostics).

The decision of which method to use should always be based on the conceptual goals of the analysis. When there are multiple X variables, CAP can be helpful for exploring relationships between these variables and a multivariate data cloud expressed by a resemblance matrix. In contrast, dbRDA is more appropriate when one wishes to explicitly model the variability in the multivariate data cloud using a set of X predictor variables. Table 5.1 provides a summary of the differences between these two approaches.
Table 5.1. Summary of differences between CAP and dbRDA for the analysis of the relationships between two sets of variables. Matrices are defined in the text and are also defined in the index of mathematical notation.

<table>
<thead>
<tr>
<th>Analysis of:</th>
<th>CAP</th>
<th>dbRDA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Q^\circ$ and $X^0$</td>
<td>$Q$ and $X^0$</td>
</tr>
<tr>
<td>Standardisation:</td>
<td>PCO axes ($Q^0$) are unit-normed (sphericised) to SSCP = I. X variables are sphericised.</td>
<td>PCO axes ($Q$) are standardised to SSCP = $\Lambda$. X variables are sphericised.</td>
</tr>
<tr>
<td>Roles of data clouds:</td>
<td>Symmetric: $Q^\circ$ ↔ $X^0$</td>
<td>Directional: Project $Q$ (response data cloud) onto $X^0$ (predictors)</td>
</tr>
<tr>
<td>Avoiding over-parametrisation:</td>
<td>A subset of $m$ PCO axes are used; $m$ chosen using ‘leave-one-out’ diagnostics. All X variables are used.</td>
<td>All PCO axes are used. A subset of X variables can be chosen (optionally) using model selection criteria.</td>
</tr>
<tr>
<td>Purpose:</td>
<td>If $q &gt; 1$, explore correlations between two data clouds. If $q = 1$, $Q^\circ$ can be used to predict X, a single gradient.</td>
<td>Use X to explicitly model, explain or predict variation in Q.</td>
</tr>
<tr>
<td>Test-statistic:</td>
<td>$tr(Q^\circ H Q^\circ)$</td>
<td>$tr(HQH)$ (best to calculate directly as $tr(HGH)$ for when there are negative eigenvalues)</td>
</tr>
<tr>
<td>Interpretation of eigenvalues:</td>
<td>Squared canonical correlation ($\delta^2$)</td>
<td>A portion of the explained variation ($\gamma^2$), which can be expressed as a proportion of the total explained variation ($\Sigma \gamma^2$).</td>
</tr>
<tr>
<td>Interpretation of axes:</td>
<td>A linear combination of $Q^\circ$ variables having maximum correlation with a linear combination of X variables.</td>
<td>A linear combination of X variables that explains the greatest amount of variation in the response data cloud Q.</td>
</tr>
<tr>
<td>Default for vector overlay:</td>
<td>Linear combinations of $X^0$ having maximum correlation with CAP axes.</td>
<td>Direct projection of $X^0$ onto dbRDA axes.</td>
</tr>
</tbody>
</table>

The relationship between dbRDA and CAP can also be seen if we consider their formulation using singular value decomposition (SVD). For simplicity, but without loss of generality, suppose that each of the variables in matrices Y and X are centred on their means. A classical RDA is then obtained by the SVD of matrix $Y'X^0$, namely:

$$Y'X^0 = U_R W_R V_R'$$  \hspace{1cm} (5.7)

where $W_R = \Gamma$, a diagonal matrix of eigenvalues ($\gamma_1, \gamma_2, \ldots, \gamma_s$).\textsuperscript{114} The RDA coordinate scores are then linear combinations of the fitted values, $\hat{Y} = HY$, where the coefficients are contained in the left-singular vectors $U_R$. That is,

$$Z = \hat{Y} U_R$$  \hspace{1cm} (5.8)

Now, a classical CCorA is obtained by SVD of matrix $Y^0'X^0$, namely:

$$Y^0'X^0 = U_C W_C V_C'$$  \hspace{1cm} (5.9)

\textsuperscript{114} Note: these are the square root of the eigenvalues that would be obtained from running the dbRDA routine on the data using Euclidean distances.
where $W_C = \Delta$, a diagonal matrix of eigenvalues ($\delta_1, \delta_2, \ldots, \delta_r$) that are the classical canonical correlations.\(^{115}\) As CCorA is symmetric, we can plot coordinates either in the space of $X$, or in the space of $Y$. Canonical coordinate scores in the space of $X$ are:

$$B = X^\prime V_C \Delta$$

(5.10)

while canonical coordinate scores in the space of $Y$ are:

$$C = Y^\prime U_C \Delta$$

(5.11)

The correlation between the variables $B_1$ and $C_1$ will be equal to $\delta_1$, between $B_2$ and $C_2$ will be $\delta_2$, and so on. Note also that the scaling of each of these new sets of variables by their corresponding eigenvalues (i.e., multiplying by $\Delta$ as shown in equations 5.10 and 5.11) is optional.

Next, we can generalise the above to any resemblance measure by replacing $Y$ with $Q$, thus dbRDA is obtained by an SVD as follows:

$$Q^\prime X^\prime = U_R \Gamma V_R$$

(5.12)

with fitted values $HQ$, dbRDA coordinate scores $Z = HQU_R$ and with an appropriate vector overlay for the plot (eigenvector coefficients for normalised $X$ variables) being contained in $V_R$.

Similarly, CAP is obtained by an SVD of:

$$Q_m^\prime X^\prime = U_C \Delta V_C$$

(5.13)

with canonical coordinate scores $Q_m^\prime U_C \Delta$ and an appropriate vector overlay for the plot (eigenvector coefficients for normalised $X$ variables) being contained in $V_C$.

The purpose of this section is not to throw some scary matrix algebra around and generate fear! Rather, it is intended to further highlight the conceptual differences and similarities between these two approaches (as outlined in Table 5.1) and also to provide some algebraic conventions for formulating and discussing these methods which (hopefully) complements existing literature describing the classical (Euclidean-based) versions of them.

---

A study by van der Aart & Smeek-Enserink (1975) explored the relationships between two sets of variables: the abundances of hunting spiders (Lycosidae) obtained in pitfall traps and a suite of environmental variables for a series of sites across a dune meadow in the Netherlands. A subset of these data ($p = 12$ spider species, $q = 6$ environmental variables and $N = 28$ sites) are provided in the ‘Spiders’ folder of the ‘Examples add-on’ directory. Open up the files containing the spider data (hspi.pri) and the environmental variables (hspienv.pri) in PRIMER. Transform the spider data using an overall square-root transformation, then calculate a resemblance matrix using the chi-squared distance measure (D16). By using chi-squared distances as the basis for the analysis, we are placing a special focus on the composition of the spider assemblages in terms of proportional (root) abundances. Next, see the description of the environmental data by clicking on hspienv.pri and choosing Edit > Properties. The variables measured and included here are water content, bare sand, moss cover, light reflection, fallen twigs and herb cover, all on a log scale. A draftsman plot (including the choice Correlations to worksheet) shows that no additional transformations are necessary. Also, the maximum correlation observed is between fallen twigs and light reflection ($r = -0.87$), so it is not really necessary to remove any of these variables. From the chi-squared resemblance matrix of square-root transformed spider data, choose PERMANOVA+ > CAP > (Analyse against Variables in data worksheet: hspienv) & (Diagnostics Do diagnostics) & (Do permutation test > Num. permutations: 9999), then click OK.

The results show that there were some very strong and significant correlations between the spider abundance data cloud (based on chi-squared distances) and the environmental variables ($P = 0.0001$). The first two canonical correlations are both greater than 0.90 (Fig. 5.24, $\delta_1 = 0.9809$, $\delta_2 = 0.9256$). Diagnostics revealed that the first $m = 4$ PCO axes (which together explained 92.7% of the total variability in the resemblance matrix) resulted in the smallest leave-one-out residual sum of squares, so there was no need to include more PCO axes in the analysis.

\(^{115}\) Note: these are the square root of the eigenvalues that would be obtained from running the CAP routine on the data using Euclidean distances.
5. CAP

Fig. 5.24. Excerpts from the output file of the CAP analysis of the hunting spider data.

Fig. 5.25. CAP ordination plot relating hunting spiders to environmental variables.

The CAP axes (‘Canonical coordinate scores’) given in the output file and also shown graphically in the plot are new variables (matrix C in Fig. 5.2) that are linear combinations of the PCO’s (based on the resemblance measure of choice) that have maximum correlation with the X’s. Also given in
the output file are the weights, labeled ‘Canonical eigenvectors in the space of X’. These are the coefficients for linear combinations of the normalised X variables that will produce axes that have maximum correlation with the CAP axes. For example, the following linear combination of normalised X variables (produced using the weights given under ‘CAP1’ in the output file, Fig. 5.24):

\[ B_1 = -0.303(X_{\text{norm, WaterCon}}) + 0.453(X_{\text{norm, BareSand}}) - 0.512(X_{\text{norm, FallTwig}}) + 0.469(X_{\text{norm, CoreMoss}}) - 0.046(X_{\text{norm, CoreHerb}}) + 0.467(X_{\text{norm, ReLux}}) \]  

(5.14)

produces a new variable \( B_1 \) that has maximum correlation with CAP axis 1 \( (C_1) \). Furthermore (and the reader is encouraged to verify this by hand, it is perfectly safe!), the Pearson correlation between these two variables \( (B_1 \text{ and } C_1) \) is precisely the first canonical correlation of \( \delta_1 = 0.98 \). Similarly, the weights given for the normalised X variables for ‘CAP2’ will produce a second new variable \( B_2 \), which is independent of (perpendicular to) the first variable \( B_1 \) and has maximum correlation with CAP axis 2 \( (C_2) \), which is \( \delta_2 = 0.93 \), and so on. These eigenvector weights are also able to be seen visually on the CAP plot, as the default vector overlay for the X variables (Fig. 5.25).

One thing to be aware of here is that the CAP axes shown in the graphic and given in the output file as canonical coordinate scores are not a linear combination of the X variables, but of the PCO’s. Therefore, the default vector overlay shown in the CAP plot is not the same as what would be obtained by a direct projection of the X variables (as multiple partial correlations) onto these axes (i.e., using the ‘Multiple’ option as the correlation type in the ‘Configuration Plot’ dialog of Graph > Special). This contrasts with the dbRDA plot, where the relationships between the X variables and the dbRDA axes shown by the default vector overlay and the projected multiple partial correlations are indeed the same thing (see the section Vector overlays in dbRDA in chapter 4).

![Hunting spider data](image)

**Fig. 5.26.** CAP ordination plot relating hunting spiders to environmental variables, but with a vector overlay consisting of the multiple partial correlations of the original species variables (spider abundances, square-root transformed) with the canonical axes.

For the spiders dataset, we can see a fundamental shift in the structure of the assemblage that is strongly associated with the environmental variable of log percentage cover of fallen leaves and twigs (Fig. 5.25, see the samples numbered 16, 8, 17, 19, 21, 15, 20 and 18 at the bottom lower-left of the diagram and the associated vector labeled ‘FallTwig’). In addition, a gradient in community
composition is evident among the other samples (stretching from the upper left to the lower right of the canonical plot), which is strongly related to log percentage of soil dry mass (‘WaterCon’) and log percentage cover of the herb layer (‘CoveHerb’) on the one hand, and log percentage cover of bare sand (‘BareSand’), moss cover (‘CoveMoss’) and light reflection (‘RefLux’) on the other.

Although the purpose here is to do little more than explore relationships, some clear patterns have emerged. Another vector overlay that can elucidate patterns, particularly for the spiders dataset, as we have just a few original species variables ($p = 12$), is to project the multiple partial correlations of these original variables (suitably transformed, in this case located in the worksheet named ‘Data1’) onto this plot (e.g., Fig. 5.26). Choose Graph > Special > (Vectors: $\bullet$Worksheet variables: Data1 > Correlation type: Multiple). Certain species, such as *Pardosa lugubris* (‘Pardlugu’) and *Trochosa terricola* (‘Trocterr’) are associated with fallen leaves and twigs, while others, such as *Arctosa perita* (‘Arctperi’), *Alopecosa fabrilis* (‘Alopfabr’) and *Alopecosa accentuata* (‘Alopacec’), are associated with bare sand. This type of vector overlay, as outlined previously (see the section on Vector overlays in dbRDA), projects the (orthonormal) Y variables as multiple partial correlations onto the CAP axes. The cautions and caveats associated with interpreting vector overlays should be kept in mind for CAP, as for other ordination techniques in the PERMANOVA+ add-on package.
Acknowledgements

We wish to thank our many colleagues, whose ongoing research has supported this work by providing ideas and datasets. We trust that our citations in the text and associated with datasets provides ample evidence of the many researchers who have contributed towards the development of methods and this software. We extend special thanks to those who have organised courses and workshops with the earlier DOS and beta versions of this software, testing these methods. The software and the scope of the manual were greatly improved by these trials, especially through questions, comments and suggestions offered by participants. We offer special thanks to Antonio Terlizzi and Euan Harvey, who, by organising combined workshops at the University of Lecce and at the University of Western Australia, respectively, were largely responsible for bringing us together, leading to this joint endeavour. Thanks are also due to the University of Auckland, Plymouth Marine Laboratory and Massey University for their recognition and support of this work. KRC would like to acknowledge his Honorary Fellowships of the Plymouth Marine Laboratory and the Marine Biological Association of the UK, and his Adjunct Professorship at Murdoch University, Western Australia.

References

References


Godsell PJ & Connell SD. 2002. Can habitat loss be treated independently of habitat configuration?

Implications for rare and common taxa in fragmented landscapes. Marine Ecology Progress Series 239: 37-44.


Pinet EJG. 1937b. Significance tests which may be applied to samples from any populations. Biometrics Bulletin 1: 191-193.

Pinet EJG. 1937a. Significance tests which may be applied to samples from any populations. III. The correlation coefficient test. Biometrics Bulletin 3: 119-130.


References


Index to mathematical notation and symbols

Matrices and vectors
A = matrix containing elements \( a_{ij} = \frac{1}{2} d_{ij}^2 \)
B = matrix of variables \( (N \times s) \) that are linear combinations of normalised \( X \) variables having maximum correlation with CAP axes
C = matrix of CAP axes \( (N \times s) \), standardised by the square root of their respective eigenvalues
D = matrix containing elements \( d_{ij} \) corresponding to distances or dissimilarities
G = Gower’s centred matrix, consisting of elements \( g_{ij} = a_{ij} - \overline{a}_i - \overline{a}_j + \overline{a} \)
H = ‘hat’ matrix = \( X[X'X]^{-1}X' \), used as a projection matrix for regression models
I = identity matrix, with 1’s along the diagonal and 0’s elsewhere
Q = matrix of PCO axes, standardised by the square root of their respective eigenvalues
Q' = matrix of PCO axes, orthonormalised to SSCP = \( I \) (‘sphericised’)
U = matrix whose columns contain the left singular vectors from a singular value decomposition (SVD) of a matrix (e.g., \( X = U W V' \)); if \( X \) is \( (N \times q) \) and \( q < N \), then \( U \) is \( (N \times q) \)
V = matrix whose columns contain the right singular vectors from a singular value decomposition (SVD) of a matrix (e.g., \( X = U W V' \)); if \( X \) is \( (N \times q) \) and \( q < N \), then \( V \) is \( (q \times N) \)
W = diagonal matrix of eigenvalues from a singular value decomposition (SVD) of a matrix (e.g., \( X = U W V' \)); if \( X \) is \( (N \times q) \) and \( q < N \), then \( W \) is \( (q \times q) \)
X = matrix of predictor variables \( (N \times q) \) (often a set of environmental variables)
X' = matrix of \( X \) variables, orthonormalised to SSCP = \( I \) (‘sphericised’)
Y = matrix of response variables \( (N \times p) \) (often a set of species variables)
Y0 = matrix of \( Y \) variables, orthonormalised to SSCP = \( I \) (‘sphericised’)
\( \hat{Y} \) = matrix of fitted values \( (N \times p) \)
y = vector of \( p \) response variables for the \( j \)th observation in the \( i \)th group
\( \overline{y}_i \) = the centroid vector of \( p \) response variables for group \( i \)
Z = matrix of dbRDA canonical axes \( (N \times s) \)

Letters
\( a, b, c, \) etc… = number of levels of factor A, B, C, etc… in an ANOVA experimental design
\( AIC \) = multivariate analogue to Akaike’s ‘An information criterion’
\( AIC_c \) = multivariate analogue to the small-sample-size corrected version of \( AIC \)
\( B_i \) = the \( i \)th variable in the space of normalised \( X \) variables that has maximum correlation with the \( i \)th coordinate axis (\( C_i \)) from a CAP analysis
\( BIC \) = multivariate analogue to Schwarz’s ‘Bayesian information criterion’
\( C_i \) = the \( i \)th coordinate axis scores from a CAP analysis
\( d_{ij} \) = distance or dissimilarity between sample \( i \) and sample \( j \)
\( df \) = degrees of freedom
\( F \) = pseudo-\( F \) statistic for testing hypotheses in PERMANOVA or DISTLM
\( i \) = index used for samples (i.e., \( i = 1, \ldots, N \)) or index used for groups (\( i = 1, \ldots, a \))
\( j \) = second index used for samples (i.e., \( j = 1, \ldots, N \)) or index used for replicates within a group (\( j = 1, \ldots, n \))
\( k \) = index used for variables (i.e., \( k = 1, \ldots, p \) or else \( k = 1, \ldots, q \))
\( \ell \) = index used for canonical axes or eigenvalues for either dbRDA or CAP (i.e., \( \ell = 1, \ldots, s \)) or the abbreviation for ‘log-likelihood’ or the ‘length’ of a vector (depending on context).
\( m \) = number of PCO axes chosen as a subset for analysis by CAP
\( MC \) = Monte Carlo
\( MS \) = mean square
\( N \) = total number of samples
\( n \) = number of samples (replicates) within a group or cell in an experimental design
\( P \) = \( P \)-value associated with the test of a null hypothesis
\( p \) = number of multivariate response variables in matrix \( Y \)
\( q \) = total number of predictor variables in matrix \( X \)
\( r \) = Pearson correlation coefficient
Index to symbols

\( R = \) the ANOSIM \( R \) statistic (see Clarke 1993)
\( R^2 = \) proportion of explained variation from a model
\( s = \) number of canonical eigenvalues and associated canonical axes obtained from either a dbRDA
\( \text{or} \) a CAP analysis
\( SS = \) sum of squares
\( SSCP = \) sum of squares and cross products
\( SVD = \) singular value decomposition
\( t = \) pseudo-\( t \) statistic = \sqrt{\text{pseudo-}F}
\( tr = \) ‘trace’ of a matrix = the sum of the diagonal elements
\( X_k = \) the \( k \)th predictor variable
\( Y_k = \) the \( k \)th response variable
\( z_{ij} = \) distance to group centroid for the \( j \)th replicate within the \( i \)th group.

Greek symbols and matrices
\( \alpha = \) significance level chosen for a test (usually \( \alpha = 0.05 \)).
\( \delta^2 = \) the \( \ell \)th eigenvalue from a CAP analysis, a squared canonical correlation
\( \Delta = \) diagonal matrix containing the square roots of the eigenvalues from a CAP analysis (a capital delta)
\( \gamma^2 = \) the \( \ell \)th eigenvalue from a dbRDA analysis, a portion of the explained (regression) sum of squares from a dbRDA model.
\( \Gamma = \) diagonal matrix containing the square roots of the eigenvalues from a dbRDA analysis (a capital gamma)
\( \lambda_i = \) the \( i \)th eigenvalue from a PCO analysis
\( \Lambda = \) diagonal matrix of eigenvalues from a PCO analysis (a capital lambda)
\( v = \) number of parameters in a particular model during model selection
\( \rho = \) Spearman rank correlation (rho)
\( \Sigma = \) sum over the relevant index
## Index to data sets used in examples

Below is an index to the data sets used in examples, listed in order of appearance in the text. With each dataset are given the name and location of the data file, the original reference, a description of its use as an example in the manual and the page number where this can be found (italicised and in parentheses).

1. Ekofisk oil-field macrofauna (ekma.pri in Examples v6\Ekofisk, Gray et al. 1990) – demonstrate one-way PERMANOVA (22), model selection procedures, diagnostics and building models in DISTLM (136) and visualising models using dbRDA (145).
2. Victorian avifauna (vicsurv.pri in Examples add-on\VictAvi, Mac Nally & Timewell 2005) – demonstrate Monte Carlo P values (28). Also used at the level of individual surveys (vicsurv.pri) to demonstrate a repeated measures design (65) and also PCO (107), negative eigenvalues (109), scree plots (111) and vector overlays (112).
3. Subtidal epibiota (sub.pri in Examples add-on\SubEpi, Glasby 1999) – demonstrate a two-way crossed design (31) and contrasts (40) in PERMANOVA.
4. Tasmanian meiofauna (tas.pri in Examples add-on\TasMei, Warwick et al. 1990a) – demonstrate fixed versus random factors (42), components of variation (43), expected mean squares (44), constructing F from EMS (45), exchangeable units (46), inference space and power (46), and testing the design (49).
5. Holdfast invertebrates (hold.pri, holdenv.pri and Mollusca.agg in Examples add-on\HoldNZ, Anderson et al. 2005a) – demonstrate a nested design (50), estimating components of variation (53), and pooling or excluding terms (54). Also used later to demonstrate analyses with covariates in PERMANOVA (74) and marginal and conditional tests with DISTLM (130).
6. Plankton net study (plank.pri in Examples add-on\Plankton, Winsor & Clarke 1940) – demonstrate designs that lack replication (59) and increased power as a result of blocking (60).
7. Woodstock plants (wsk.pri in Examples add-on\Woodstock, Prober et al. 2007) – demonstrate a split-plot design (61).
11. Bumpus’ sparrows (spar.pri in Examples add-on\BumpSpar, Bumpus 1898) – demonstrate test of dispersion in Euclidean space (88).
12. Tikus Island corals (tick.pri in Examples v6\Corals, Warwick et al. 1990b) – demonstrate test of dispersion for ecological data (92) and how choice of dissimilarity measure matters (92). Also used later to demonstrate how CAP tells you nothing about relative within-group dispersions (172).
13. Norwegian macrofauna (norbio.pri and norenv.pri in Examples add-on\NorMac, Ellingsen & Gray 2002) – demonstrate use of the test of dispersion to investigate beta diversity (95).
14. Okura macrofauna (okura.pri, in Examples add-on\Okura, Anderson et al. 2004) – demonstrate tests of dispersion in nested designs (97). Also used to demonstrate PCO of distances among centroids (118) and PCO versus MDS when samples are split into groups (121).
15. Cryptic fish assemblages (cryptic.pri in Examples add-on\Cryptic, Willis & Anderson 2003) – demonstrate PERMDISP for a two-factor crossed design, in conjunction with PERMANOVA (100).
17. Thau lagoon bacteria (thbac.pri and thevsp.pri in Examples add-on\Thau, Amanieu et al. 1989) – demonstrate analysing variables in sets using DISTLM (147).
Index to data sets


20. Poor Knights Islands fish (pkfish.pri in Examples add-on\PKFish, Willis & Denny 2000) – demonstrate discriminant analysis based on Bray-Curtis using CAP (160).

21. Iris data (iris.pri in Examples add-on\irises, Anderson 1935) – demonstrate classical discriminant analysis and MANOVA test statistics using CAP (167). Also used later to show how the positions of new samples are added into a discriminant-type analysis, with prediction of group membership (172).

22. Fal estuary biota (Fa.xls in Examples v6\Fal; falbio.pri and falenv.pri in Examples add-on\FalEst, Somerfield et al. 1994) – demonstrate canonical correlation analysis with CAP based on the Bray-Curtis measure relating biota to a single environmental gradient (175).

General index

A

A matrix (in PCO), 106

a posteriori comparisons. See pair-wise comparisons

a priori

choice of m in CAP, 179

contrasts. See contrasts
groups, 6, 7, 104, 157

number of dimensions in MDS, 105

random allocation, 48

removal of terms, 55

scale of observation, 121

significance level. See significance level

abundance data, 17, 115

accuracy of P-values, 24, 83

adding new samples, 172–75

additional assumptions, complex designs, 40

additivity, 37–38

affected by ranking, 38

affected by resemblance measure, 38

affected by transformation, 37

in the space of the resemblance measure, 39

of treatment effects, 21

adequacy

of a CAP model, 166–67

of a dbRDA plot, 143, 145

of a DISTLM/dbRDA model, 146

of a PCO plot, 107, 112

of an MDS plot, 105

adjusted

AIC. See AIC,

Bray-Curtis resemblance, 108, 109, 111

$R^2$, 135, 149, 152

aerobic heterotrophs, 148

AIC, 135, 141

AICc, 135

Akaike. See AIC

algebraic conventions, 183

all cells filled, 69, 71

allocation

of new observations to groups, 157, 159, 172–75

success, 157, 164, 166

among-group sum of squares, 16, 18, 169, 170

among-group variation, 31

an information criterion. See AIC

analysing variables in sets (DISTLM), 147–50

ANCOVA model, 74

ANOSIM

affinity with MDS, 30, 106

flexible and robust, 5

non-parametric approach, 5, 155, 156

power relies on number of permutations, 31

R statistic, 30

ranking, 30

sensitivity to differences in dispersion, 21, 87

vs PERMANOVA, 30–31

ANOVA

and methods of permutation, 38

and quasi $F$ ratios, 81

as a linear model, 123

as a regression on categorical variables, 151

assumptions, 21

classical, using PERMANOVA, 19, 21, 56

estimators of variance components, 53, 54

expectations for mean squares, 44

experimental designs, 18

for Levene’s test, 88

identity, 17

logic for testing interactions, 35

mixed models, 45

models, 38, 41

one-way partitioning, 16

over-parameterisation, 45

partitioning, 56

robust to heterogeneity, 22

transformations to comply with assumptions, 38

used in PERMDISP, 87

approximate $P$-value

for traditional quasi $F$ ratios, 81

obtained using Monte Carlo asymptotics, 30

obtained using permutations, 25, 38

arbitrary signs of axes in ordination plots, 174

arithmetic averages, 90

asserting a component of variation is equal to zero, 54

assumption-free. See non-parametric assumptions

exchangeability under a true null hypothesis, 21, 48

fixed effects sum to zero, 45

for analysing randomised blocks, 58

for analysing repeated measures, 65

for analysing split-plot designs, 62, 63

for complex designs, 39–40

for designs lacking replication, 61

for univariate ANOVA, 21

for using Monte Carlo $P$-values, 28

general, 6

homogeneity

of dispersions, 21

of slopes in ANCOVA, 77

in CAP, 169, 179

in DISTLM, 133, 132–34, 156

in observational studies, 48

in PERMANOVA, 22

in traditional MANOVA, 32, 169

independent and identically distributed, 21

of conditional tests, 133

random sampling, 48

asymmetrical designs, 82–85

asymptotic

approximation. See approximate $P$-value

permutation distribution, 28

asymptotically exact test, 38, 129

average

distance-to-centroid, 99

similarities/dissimilarities, 22, 28, 31, 101

taxonomic distinctness, 56

avifauna. See Victorian avifauna

avoiding normality assumption, 21

axes. See number of axes

axis

maximising total variation, 157

related to quantitative variable, 159

separating groups, 157

B

B matrix (in CAP), 183

BACI designs, 85

backward elimination, 134

bacteria. See Thau lagoon bacteria

balanced design

definition, 32, 68

equality of types of SS, 69
fully nested model, 50
independent terms, 69
multipliers on components of variation, 68
one-way model, 18
split-plot, 64
two-way crossed model, 32
Venn diagram, 69
with covariates, 73
barrens habitat, 79, 100
Bayesian information criterion. See BIC
beetles. See flea-beetles
before/after, control/impact. See BACI designs
benthic, 22, 95, 97, 115, 126, 136
BEST, 19, 134, 137, 154, 155, 156
best (selection procedure in DISTLM), 134
best result for each number of variables, 140
beta diversity, 94–96
between-group sum of squares. See among-group sum of squares
beyond BACI, 85
BIC, 135, 141
binary data
categorical predictor variables, 151
presence/absence, 37, 94
binomial deviation, 17, 28, 65, 66, 79
BIENV, 19, 154, 155, 156
birds from Borneo, 71
blocks
in a randomised block design, 59
in a two-way mixed model, 43
in split-plots, 61
treated as fixed, 47
Bonferroni, 27
bootstrapping, 85
Bray-Curtis resemblance
adjusted, 108, 109, 111
after applying a transformation, 38
and menu instructions, 12
and sparse data, 92, 93
as a semi-metric measure, 109
as multivariate variation, 133
as similarity or dissimilarity, 23
as variation in community structure, 23
defines a non-Euclidean space, 90, 99, 120
distance-to-centroid, 92, 99
in CAP, classification, 159
in PCO, distances among centroids, 119
in PCO, non-linear relationships, 117
in PERMDISP, 93
interpretation as a percentage, 53
on birds from Borneo, 71
on cryptic fish data, 100
on Ekofisk macrofauna, 23, 138
on Fal estuary biota, 176
on Mediterranean molluscs, 83
on Okura macrofauna, 97, 118
on oribatid mites, 152
on Poor Knights Islands fish, 161
on presence/absence data, 95
on species abundance data, 17
on subtidal epiibiot, 33
on Tasmanian meiofauna, 43
on Tikus Island corals, 92, 172
on Woodstock plants, 63
brief description of routines, 6
bubble plots, 10, 75, 108, 144
Bumpus’ sparrows, 88
BVSTEP, 19, 154

C

C matrix (in CAP), 160, 183
CA
description, 105
scaling method 1, 117
unimodal relationships, 117
vs PCO, 117
calculating sums of squares from a distance matrix, 18
canonical
analysis of principal coordinates. See CAP
coordinate axis scores (C), 160
correlation analysis. See CCorA
correlations, 160, 161
correspondence analysis. See CCA
discriminant analysis. See CDA
eigenvalues
in CAP, 160
in RDA, 143
partitioning, 19
variate analysis, 167
CAP
adding new samples, 172–75
adequacy of the model, 166–67
allocation success, 166
analysis
of Fal estuary biota, 176
of hunting spiders, 183–86
of iris data, 168
of Poor Knights Islands fish, 160–63
of Tikus Island corals, 172
canonical
analysis of
a single gradient, 175–79
multiple X’s, 179
coordinate axes, 160, 184
correlations, 160, 161
eigenvalues, 160
eigenvectors in the space of X, 185
caveats, 171–72
characterise group differences, 161
cross-validation, 166–67
diagonstics, 163–65, 179
discriminant analysis, 160–63
exploring relationships, 179
mechanics of, 159–60
model placed into perspective, 178
sphericising variables, 181
symmetric roles of variables, 179, 181
tells nothing of relative dispersions, 172
test by permutation, 167–69
vs dbRDA, 181–82
vs PERMANOVA, 171
weights, 185
cartouche, 12
categorical predictor variables, 150–54
causation, 83, 115, 136, 142
caveats on using CAP, 171–72
CCA, 133, 136
CCorA, 160, 181, 182
CDA, 160, 167
cells
as exchangeable units, 46
in a design, 32
centering groups in PERMDISP, 91
central limit theorem, 28, 30
centred variable (definition), 180
centroid
  in the space of the resemblance measure, 90
overall, 16, 91
  visualisation, 105
change the order of terms
  in DISTLM, 131
  in PERMANOVA, 58
characterise group differences, 161
check box, 12
chi-square, used in MC asymptotic result, 28
chi-squared distance, 17, 105, 117, 133, 183
choice of resemblance measure
  affects relative dispersions, 93
  defines multivariate variation, 6
  in PERMDISP, 92–94
meaningful for PERMANOVA, 30
  choices made, given in results file, 24
classical methods. See traditional methods
classical scaling. See PCO
classification, 158, 172–75
CLUSTER, 7
cluster analysis, 157, 159
Clyde
  environmental data, 115, 126
  macrofauna, 126
combinations
  of predictor variables, 134
  of treatment levels, 32
combining
  factors, 43, 98, 102
  results from two PERMANOVA tables, 64
  sums of squares. See pooling or excluding terms
community data, 17
calibration of methods using SVD, 182–83
complex designs, additional assumptions, 39–40
criteria
  of variation, 43–44
  for repeated measures, 65
correlation
  for multiple tests, 27, 41, 131
  for negative eigenvalues, 123
  to AIC for small samples, 135
  to df for repeated measures, 65
covariates
  and linear combinations of mean squares, 79
  and sequential tests, 132
  and types of sums of squares, 77
  expected mean squares, 77
  in DISTLM, 129
  for multiple tests, 158
  for model selection, 134
  for unconstrained ordination, 105, 158

crossed factors, 31, See two-way crossed design

cryptic fish, 100

doctor 

d (distance/dissimilarity) matrix
  in CAP, 159
  in DISTLM/dbRDA, 124, 142
  in PCO, 106
  in PERMANOVA, 18
data sets. See index to data sets
dbRDA. See DISTLM/dbRDA
defining multivariate variation. See choice of

degrees of freedom
  affect pseudo-F, 30
  affect R², 149
  after pooling, 57
General index

and power, 31
associated with the denominator, 45
associated with the numerator, 45
available for contrasts, 41
for a split-plot design, 62
for categorical predictor variables, 151
in multivariate systems, 126
in PERMANOVA, 19
in sets of predictor variables, 149
not whole numbers, 78
running out of, 126, 155
test the design, 49
delta+, 56
denominator
df increases power, 47
for the F ratio, 45
identifies exchangeable units, 38, 46
linear combination of mean squares, 78
no single term, 78
density, 130, 150
deposition of sediment, 97
depth, 130
design file
create, 23
eat, 23
factor information, 23
icon, 23
new development, 9
save in its own format, 24
to test a design, 49
designs. See PERMANOVA: experimental designs
details of the expected mean squares, 45
deviations
from fitted values, 39
from means, 16, 91
from medians, 91
from normality, 65
of cells means from expected values, 37
of group means from the overall mean, 16
df. See degrees of freedom
diagnostics
for CAP, 163–65
for DISTLM/dbRDA, 133, 132–34, 136
for PCO, 112
proposed for PERMANOVA, 39
dialog
for calculating distances among centroids, 119
for calculating taxonomic distinctness, 55
for CAP
adding new samples, 173, 177
canonical correlation, 176
discriminant analysis, 162
for creating a design file, 24
for dbRDA, 144
for DISTLM
excluding certain variable sets, 152
fitting sets of predictor variables, 148
simple regression, 127
two ordered predictor variables, 131
for PCO, 108
for PERMANOVA
contrasts, 40
include covariates, 76
nested terms, 51
one-way, 25
pair-wise tests, 27
pooling, 58
repeated measures, 67
split-plots, 63, 64
for PERMDISP, 89
for selecting a subset of variables, 51
for vector overlays, 113
dialog box, 12
differences
between impact and controls, 83
between time points, 65
in beta diversity, 95
in cell means, 37
in correlation structure, 22
in dispersion, 22, 87, 103
in location, 22, 87
in sample sizes, 69
in sizes of components of variation, 85
dimensionality reduction. See ordination
direction of effects, 32
directories for example data, 12
discriminant analysis, 160–63, 167
discriminate among groups, 158
dispersion
among centroids, 97
among samples, 87
as beta diversity, 94–96
effects, 101
in crossed designs, 100–103
in nested designs, 96–100
dispersion weighting, 176
dissimilarity
automatically calculated from similarity, 23, 124
being modeled directly, 15, 18, 30, 106, 123, 155
between time points, 66
preserved by PCO, 107, 112
properties, 109
distance-based
linear models. See DISTLM/dbRDA
redundancy analysis. See DISTLM/dbRDA
distances
among centroids
calculated from PCO axes, 105, 106, 118–21
used to analyse split-plot designs, 62
from observations
to centroids, 87, 88
to spatial medians, 87
distinctiveness of groups, 161, 166–67
distinguishing a priori groups, 157
DISTLM/dbRDA
adequacy of the model, 146
analysing variables in sets, 147–50
analysis
of Clyde macrofauna diversity, 128
of Ekofisk macrofauna, 136–42
of holdfast invertebrates, 130–32
of oribatid mites, 153
of Thau lagoon bacteria, 149
as RDA on PCO axes, 133
assumptions and diagnostics, 132–34
building models, 134–36
categorical predictor variables, 150–54
cautionary notes, 136
classical regression with Euclidean distance, 126
conditional tests, 128–30
constrained ordination, 123
description, 123
drop redundant variables, 133
excluding variables, 152
explain/model/predict variability, 124
limitations, 155
limits on inferences, 136
matching sample names, 127
multi-collinearity, 133
non-linear models, 133
ordination
  adequacy of the plot, 145
  compare patterns with PCO plot, 143, 146
  of Ekofisk macrofauna, 145–47
  of fitted values, 143
  of oribatid mites, 154
  relative importance of predictor variables, 146
partitioning, 123–26
rationale, 123
relationships between \( Y \) and \( X \) depend on
  resemblance measure, 133
  roles of variables, 124
  sequential tests, 132
  simple linear regression, 126–28
test by permutation, 125
test statistic, 124, 129
variables have roles, 124
visualise fitted model, 142–43
\( \hat{\theta} \) vs BEST/BIOENV, 154–56
\( \hat{\theta} \) vs CAP, 181–82
distribution
  asymptotic d. under permutation, 15, 25
  linear form in chi-square variables, 28
  normal, 180
  of categorical predictor variables, 152
  of covariates, 75
  of environmental variables, 75, 115
  of original variables, 5, 21, 148, 169
  of PCO's, 179
  of points in the data cloud, 91
  of predictor variables, 148
  of pseudo-\( F \), 19, 20, 38, 65, 78, 124
  of pseudo-\( t \), 26
  of quasi \( F \) ratios, 81
  of residuals vs fitted values, 39
  of species' counts, 5
  of the traditional \( F \) statistic, 19, 56
  of variables for PCA, 59
  of \( X \) variables, 133, 179
  of \( z \)'s, 91
disturbance, 43
DOS routines, 9
dotNet environment, 4
draftsman plot
  bivariate scatterplot, 127
  of Bumpus' sparrow data, 88
  of Clyde environmental variables, 115
  of covariates, 75
  of Ekofisk sediment variables, 137, 138
  of holdfast environmental variables, 74
  of hunting spiders environmental data, 183
  of iris data, 168
  of metals from the Fal estuary, 175
  of oribatid mite environmental data, 152
  prior to CAP, 179
  prior to DISTLM/dbRDA, 133, 148
  prior to PCA, 59
dummy data to test a design, 49
dummy species, 12, 92

E
effect
  definition in PERMANOVA, 42
  fixed, 42
effect size
  additivity, 35
  in PERMANOVA, 31
  in two-way crossed design, 34
  shift in location, 34
eigenvalues
  canonical, 160
  from CAP, 160
  from CCorA, 183
  from dbRDA, 142
  from PCO, 106, 159, 170
  negative, 109, 123
  used in MC asymptotics, 28
  from RDA, 182
  from SVD, 180
  interpretation, 182
eigenvectors
  in CAP, 160, 185
  in dbRDA, 142
  in PCA, 115, 116, 117
  in PCO, 106
Ekofisk macrofauna, 22, 26, 68, 136, 145
El Niño, 92
e-mail, 4
EMS. See expected mean squares
environmental
  data. See index to data sets
  heterogeneity, 95
  impacts, 82, 85
  variables. See index to data sets
epibiota. See subtidal epibiota
equality of methods to traditional counterparts. See Euclidean distance
equivalence of methods to traditional counterparts
  and canonical correlation, 181
  and discriminant analysis, 167
  and Huygens' theorem, 17
  and Levene's test, 88, 91
  and RDA, 126, 142
  and scree plots, 117
  and simple linear regression, 126
  and traditional MANOVA, 168
  and variance components, 53, 56
equivalence of methods to traditional counterparts
  from observations to centroids, 87, 90
  linear fit in designs with covariates, 75
  linear relationships in ordination, 75
  in Bumpus' sparrow data, 89
  in Clyde environmental variables, 115
  in holdfast diversity, 131
  in iris data, 167
  on PCO axes, 90, 106, 112
  on plankton data, 59
  on species abundance data, 17
  on Thau lagoon bacteria, 148
  with CAP, 160
  with DISTLM, 125
  with pair-wise tests, 26
  with PCO, 105, 107
  with PERMANOVA, 17, 21, 30, 44, 68
  with PERMDISP, 87
Euclidean metric properties, resemblance measures, 18
  exact test
  definition, 20, 129
one-way design, 24, 25
example data. See index to data sets
Excel version, 4
exchangeability
assumed for observational studies, 21, 48
assured under a priori random allocation, 21, 48
in DISTLM, 133
of appropriate units, 48
of errors, 25, 39
of levels of a random factor, 48
of samples, 21, 65, 169
exchangeable units, 46
for a one-way model, 20
for linear combinations of mean squares, 78, 80, 81
for more complex designs, 45
excluding a term from the model, 54, 61, 63, 66, 77
excluding variables in DISTLM, 131, 152
expectations of SS and MS
excluding variables in DISTLM, 131, 152
expected mean squares, 44–45
affected by choice of fixed vs random factors, 48
Cornfield-Tukey rules, 45
details and notation, 45
equivalent for pooled terms, 54
follow univariate analogues, 15, 41, 48
for different types of SS, 72
given in results file, 25
in designs with covariates, 77
in unbalanced designs, 68
re-evaluated after pooling, 55
sum-to-zero constraints, 45
test the design, 49
to estimate components of variation, 53
to identify denominator for pseudo-F, 44, 45
used to construct linear combinations of mean
squares for tests, 78
experimental design
and logical inference, 48
shown in results file from PERMANOVA, 24
specification. See design file
visualisation, 105
explained sum of squares, 124, 129
exploratory
power of models, 134
variables, 124, 136, 147, 156
explorer tree, 24
extrapolate, 152
extreme values. See outliers

F

F statistic. See also pseudo-F
for complex designs, 45
in DISTLM, 124, 129
in PERMANOVA, 19
infinite, 96
used in PERMDISP, 87
factor
as binary codes in DISTLM, 152
definition, 15
fixed vs random, 41–43
information in the design file, 23
labels on samples, 23
nested, 49
order, 33
partitioned into contrasts, 40
to specify
groups in CAP, 159
new samples in CAP, 173, 178
Fal estuary biota, 175
family-wise error, 28
fencing, 62
fire disturbance, 62
first canonical eigenvalue, 168
fish. See New Zealand f., cryptic f., Poor Knights
Islands f. and index to data sets
Fisher, R.A., 19, 48, 167
fishing, 163
fitted values, 125, 142
fixed effects sum to zero, 45, 79
fixed factor, 42
fixed vs random factors, 41–43
flea-beetles, 157
flexibility
choice of resemblance measure, 17
of the general approach, 5
flowchart of routines, 8
flowering intensity. See Victorian avifauna
flowers, 167
forest, 71
forward selection, 134, 138
fourth root. See transformation
full model, 54, 129
full model residuals. See methods of permutation
further partitioning. See contrasts

G

G (Gower’s) matrix
in CAP, 160
in DISTLM/dbRDA, 124, 142
in PCO, 106, 142
Gaussian, 115
generalisation to dissimilarities (in PERMDISP), 90–91
generalised inverse, 180
generality
of methods using SVD, 182
test of, 5, 43, 46, 47
geographical
area, 94
coordinates, 148, 150
distance, 147
transition, 96
gometric partitioning, 16, 53
Gosset, 26
Gower’s matrix. See G
gradient analysis in CAP, 175–79
grazers, 62
groups. See levels

H

H ("hat") matrix
in CAP, 159
in DISTLM/dbRDA, 124, 125
sphericises X variables, 137, 170, 180
habitat, 74, 79, 100, 103, 150
hat matrix. See H
heavy metals, 137, 175
Hellinger resemblance measure, 17
help menu, 4
heterogeneity
in multivariate dispersions, 93
of variances, 93
hierarchical design. See nested design
high leverage points, 155
high-dimensional space, 101, 105, 122, 123, 158
highest-order
interaction, 58
term approach, 78, 80
holdfast invertebrates, 50, 53, 56, 74, 130
homogeneity
and test of sphericity, 67
of errors, 39
of levels of random factors, 87
of multivariate dispersions, 21, 87, 90, 91, 94, 99, 103
of residuals, 87
of slopes in ANCOVA, 77
of variance-covariance matrices, 33
of variances (classical), 65, 81, 87
Hotelling-Lawley trace, 169
hunting spiders, 183
Huygens’ theorem, 17–18
hydrocarbons, 136
hydrology, 97
hypotheses and construction of F ratio, 45
and inference space, 48
for fixed vs random factors, 44, 47
used in constrained ordination, 157
hypothesis. See null hypothesis

I
I (identity) matrix, 124, 180
i.i.d. See independent and identically distributed
icon for a design file, 23
identities of species, 94
identity matrix. See I
ignored regions of overlap in unbalanced designs, 70
imaginary axes, 110
space, 111
impact vs controls. See asymmetrical designs
importing factors, 15
including interactions in designs with covariates, 76
independence
of axes in CAP, 185
of axes in dbRDA, 143
of axes in PCO, 106
of orthogonal contrasts, 41
of samples in DISTLM/dbRDA, 126
of samples in PERMANOVA, 21
of terms in asymmetrical designs, 82, 85
of terms in balanced designs, 73
of terms in Type III SS, 70
independent and identically distributed, 28
errors in DISTLM models, 133
ersors in multi-way PERMANOVA, 39
levels of random factors, 48
samples in one-way PERMANOVA, 21
index
to data sets used in examples, 197
to mathematical notation and symbols, 195
indicator, 147, 151
individual deviations, 90, 96
inference
after transformation, 93
follows univariate analogue, 48
for fixed factors, 42
for random factors, 42
for significant impacts, 85
from PERMANOVA and PERMDISP (joint), 104
from random sampling, 48
from small samples, 30, 83
test by permutation for mixed/random models, 48
inflation
error rates, 22
total sums of squares, 110
total variation, 112
information
content of multivariate data, 126
criteria (AIC, AICc, BIC), 135
required to detect effects, 47
installation, 4
interaction
confounded with the residual, 60, 61
definition of, 32
interpretation of, 34–37
involving a random factor, 44
involving contrasts, 40
involving covariates, 76
no exact test, 38
patterns in an ordination plot, 35
patterns in ordination plot, 36
to detect environmental impact, 85
intercept for regression models, 124
interpretation
affected by assumptions, 39
affected by order of terms in the model, 74
does not require mathematics, 13
from a PCO plot, 107
of a test for interaction, 35
of components of variation, 53
of dispersion effects in presence of interaction, 102
of eigenvalues, 182
of ordination axes, 182
of PCO axis units, 105
of PERMANOVA and PERMDISP (jointly), 94, 104
of P-values from multiple runs, 21
of P-values from small-sample sizes, 30
of resemblance measures as percentages, 53
of vector overlays, 114
interpreting interactions, 34–37
invariance to rotation, 91
invertebrates. See holdfast invertebrates and index to
data sets
iris data, 167, 173
isolate the component of interest, 38, 44, 45, 46, 81, 129

J
Jaccard resemblance measure, 17, 52, 53, 95, 109
joint absence, 93

K
kelp forest, 79, 100, 130
Kendall rank correlation, 155
Kulczynski resemblance measure, 17

L
label matching, 33, 127, 173
large-sample approximations, 30
latent roots. See eigenvalues
latin square, 58, 61
latitude, 95, 147
least-absolute-deviation (LAD) residuals, 91
least-squares (LS) residuals, 91
least-squares scaling. See metric MDS
leave-one-out
allocation success, 157, 164
procedure, 164
residual sum of squares, 157, 164
left singular vectors, 180
levels
General index

as treatments in a crossed design, 32
definition, 15
finite for a fixed factor, 42
definition in nested designs, 97
of nested factors, 50
random subset from a population, 42
Levene's test, 88
linear combinations of mean squares, 78–82
linearity
and Pearson correlation, 115
constraint in metric MDS, 122
for covariates, 75, 134
for vector overlays, 112
in CAP, 160, 172, 179
in DISTLM/dbRDA, 123, 133, 136
in RDA, 133, 181
in simple regression, 126
in the space of the resemblance measure, 39
of ANOVA models, 37
of eigenvectors in PCA, 116
linkage trees, 136
location effects, 101
log(x+1). See transformation
logging, effects of, 71
logic
circular in confirming correlations, 163
dependence, 15
fixed vs random factors, 41
following the test for interaction, 35
for tests of dispersion in complex designs, 96
in pooling or excluding terms, 55
of hypotheses in mixed models, 48
of pair-wise comparisons
for fixed but not random factors, 42
in general, 26
of PERMDISP as complement to PERMANOVA, 87, 104
of tests extending to populations, 48
to infer causation, 83, 115, 136, 142
log-likelihood, 135
longitude, 147
M
macrofauna. See Ekofisk m., Norwegian m., Okura m., Clydes m.
and index to data sets
Mahalanobis distance, 33, 170
Manhattan distance, 93, 109
MANOVA
assumptions, 32
geometrical approach, 16
lack of robustness, 5
relative power, 31
sensitivity to differences in correlation, 22
test statistics, 169
Mantel test, 19, 126
marginal
means, 37
tests, 127, 128, 130, 131, 132
marinas, 31
marine agar, 148
marine reserve, 100, 160, 163
matching sample/variable names in CAP, 173
in DISTLM/dbRDA, 127
mathematical notation and symbols, 195
matrix algebra descriptions of methods, 182–83
maximise
allocation success, 164, 165
correlation, 179
group differences, 157
total variation, 105, 106, 157
MDS
2-stage, 19
affinity with ANOSIM, 30, 106
and additivity, 38
definition of stress, 37
description, 105
flexibility and robustness, 5, 106, 121
metric, 122
of birds from Borneo, 72
of Ekofisk macrofauna, 22
of Fal estuary biota, 179
of Mediterranean molluscs, 83
of Norwegian macrofauna, 96
of NZ cryptic fish, 101
of NZ holdfast fauna, 75
of Okura macrofauna, 98, 99, 121
of subtidal epibiota, 34
of Tasmanian meiofauna, 43
of Tikus Island corals, 92, 94, 172
of Victorian avifauna, 29, 66
references, 23
'streching and squeezing', 122
to visualise patterns, 106
unconstrained ordination, 157, 158
vs PCO, 105, 106, 121–22
mean-variance relationships, 93
measures of variability, 44
mechanics
of CAP, 159–60
of dbRDA as a constrained ordination, 142
of DISTLM, 125
of PCO, 106–7
median for multivariate data, 91
mediancentre. See spatial median
Mediterranean molluscs, 82
meiofauna. See Tasmanian meiofauna
memory requirements, 4
methods described in matrix algebra, 182–83
methods of permutation
description for multi-factor designs, 39, 38–39
permutation of residuals
in DISTLM, 129
in PERMDISP, 91
under a reduced model, 38
under the full model, 39
unrestricted permutation of raw data, 38
metric MDS, 122
metric resemblance measure, 109
minimise
leave-one-out residual SS, 164, 165
misclassification error, 164
residual variation, 105, 106
stress, 122
sum of distances from observations to the spatial
median, 91
misclassification error, 164, 166
mis-labeled factor levels, 59
missing
bits of explained variation in unbalanced designs, 70
observations. See unbalanced designs
mites. See oribatid mites
mixed models, 25, 43, 45, 46, 48, 79, 87, 152
mixtures of variable types, 150
Model II regression, 152
model selection
balancing severity with generosity, 142
bias, 136
criteria
adjusted $R^2$, 135
$AIC$, 135
$AIC_c$, 135
$BIC$, 135
$R^2$, 134
procedures
all specified, 134
backward elimination, 134
best, 134
forward selection, 134
step-wise, 134
modified Gower resemblance, 17, 75, 76
molluscs. See Mediterranean m., holdfast invertebrates
and index to datasets
monitoring programs, 68
monotonicity
and Spearman rank correlation, 115
for vector overlays, 112, 163
in ANOSIM, 30
Monte Carlo $P$-values, 28–30
too few permutations, 25
vs permutation $P$-values, 30
morphometric data, 88, 157, 167
multi-collinearity, 133, 137, 142
multiple comparisons, 27
partial correlations, 117, 144, 146, 185, 186
regression, 69, 123, 124, 125, 133, 159, 170
multiplicative effects, 37
multipliers
on components of variation, 45, 68, 69, 78
on the mean squares, 45
multivariate control charts, 68
data cloud, 16, 19, 38, 68, 123, 124, 143, 155, 158, 159, 161, 181
dispersion, 21, 87, 93, 94
Levene’s test, 88–90
location, 87
models, 5
multiple regression, 123, 125
observations. See samples
variation, 5, 6, 115, 124, 133, 156

N

negative
components of variation, 54
correlations
among variables, 151
of variables with axes, 113
effects, 35, 100
eigenvalues, 109–12
eigenvalues, corrections for, 123
sums of squares, 31
values of $AIC$, $AIC_c$, or $BIC$, 136
values of pseudo-$F$, 31, 79
variance, 31, 110
nested design, 49–53
and tests of dispersion, 97
and Type I SS, 70
distinguished from crossed design, 50
nested factor, definition, 49
new features in PERMANOVA+, 9–11
new samples. See prediction
New Zealand fish, 79
no replication, 58, 65

no test. See linear combinations of mean squares
nominal variables. See categorical predictor variables
non-concept, 95
non-independence
among values in a resemblance matrix, 126
in partial Mantel tests, 19
of factors. See interaction of pair-wise comparisons, 28
of predictor variables, 128
of response variables, 21
of samples through time, 65
of terms in an unbalanced design, 69
terms in designs with covariates, 73
through time (or space), 68
non-linear models, 39, 68, 133
relationships of axes to variables, 117
non-metric multidimensional scaling. See MDS
non-negative
eigenvalues, 117
resemblance measures, 18, 109
non-normality, 40, 91, 169
non-parametric
advantages, 155
cannot test interactions, 85
flexibility, 156
lack of a “yardstick”, 122
less powerful, 31
limitations, 5
methods in PRIMER, 5
methods work on ranks, 106
normal
equations, 124, 159
level of detail in the results, 134
theory, 21, 135
normalisation
automatic in multiple regression, 137
of variables on different scales/units, 88, 149
normalised variable (definition), 180
normality assumption, 5, 19, 21, 33, 56, 65, 75, 81
Norwegian macrofauna, 95, 121
notation
for components of variation, 44
for expected mean squares, 45
mathematical. See index to mathematical notation
null hypothesis
and construction of $F$ ratio, 45, 79
and exchangeable units, 46
for conditional tests, 129
in CAP, 169
in DISTLM, 123, 124, 133
in PERMANOVA, 19, 22
in PERMDISP, 87
number
of axes
from PCO to use in CAP, 161, 164
in CAP, 160, 161, 179
in dB-RDA, 142
in PCO, 105, 111
of possible pair-wise comparisons, 27
permutations
in DISTLM, 125
in PERMANOVA, 28
nutrient agar, 148

O

observations. See samples
oil platform, 22
Okura macrofauna, 97, 118, 122
one df contrast. See contrasts
one response variable. See univariate
one sample per cell, 58
one-way
design, 15
example in PERMANOVA, 22–23
order
data in worksheets, 9
factors
in a design file, 33
in a nested hierarchy, 50
of set of variables, 149
of samples in files, 76
of terms
and Type I SS, 70
in a design with covariates, 74
in an unbalanced design, 69
in balanced designs, 34
in DISTLM, 130
in PERMANOVA, 58
in regression models, 128
ordered selection, 74, 102
ordinal data, 50, 150
ordinalization, 105
oribatid mites, 150
orthogonal
axes in CAP, 185
axes in dBRDA, 143
axes in PCO, 106, 142
contrasts. See contrasts
factors. See crossed factors
hypotheses in Type III SS, 70
perpendicular, 178
terms in asymmetrical designs, 85
orthonormal axes, 159
orthonormālize. See sphericising variables
other test statistics, 19
outliers
in covariates, 38, 39, 75
in MDS, 121
in PCA, 59, 115
in predictor variables, 133, 155
in tests of dispersion, 91
outline of routines, 6
output individual deviation values, 90
overall best solutions (in DISTLM), 140
overlap in explained variation
among sets of predictor variables, 149
explore using Type I SS, 70
in DISTLM, 155
in multiple regression models, 128
in unbalanced designs, 69
overlay vectors. See vector overlays
over-parametrization
in ANOVA, 45, 151
in CAP, 159, 171, 181

**P**
P(MC). See Monte Carlo P-values
pair-wise comparisons
after significant interaction, 36
after significant main effects, 36
in PERMANOVA, 26–28
in PERMDISP, 99
not logical for random factors, 42
yield average similarities/dissimilarities, 101
parsimonious model, 123, 126, 136, 140, 155
partial Mantel test, 19, 126
partial tests. See conditional tests
partitioning
for a latin square design, 61
for a one-way model, 16
for a split-plot design, 62, 64
for a two-way model, 32
for dissimilarity matrices, 15
for repeated measures, 68
geometric, 16
in DISTLM/dBRDA, 123–26
in PERMANOVA, 15–17
of a factor by specifying contrasts, 40
distance matrices, discussed by others, 18
of multivariate data clouds, 19
of multivariate variability, 25, 31
PCA
adequacy of the plot, 143
as a projection, 105
description, 105, 115, 116
eigenvectors, 117
equivalent to PCO on Euclidean distances, 105, 107, 167
of Bumpus’ sparrow data, 89
of Clyde environmental variables, 115
of metals from Fal estuary, 175
of plankton net study, 59
unconstrained ordination, 157
vector overlay, 116
vs PCO, 105, 115–18
PCO
adequacy of the plot, 107
affinity with PERMANOVA and PERMDISP, 122
as a projection, 105, 106, 122
axes used
in canonical correlation, 179
in CAP, 157, 159, 163
in dBRDA, 133
in PERMANOVA, 170
axis scores, 108
compare patterns with dBRDA plot, 143, 146
corrections to negative eigenvalues, 112
description, 90, 105
diagnostics, 112
eigenvalues, 106, 143
flexibility, 117
full set of axes preserves dissimilarities, 90, 106, 107, 112
imaginary axes, 110
mechanics of, 106–7
negative eigenvalues, 109–12
on chi-squared distances, 105
on Euclidean distances, 105, 107
ordination of
Clyde environmental variables, 115
Ekofisk macrofauna, 146
Okura macrofauna (centroids), 120
Poor Knights Islands fish, 161
Victorian avifauna (subset), 109, 114
plotting options, 108
reflection of axes, 106
scaling of axes, 106, 170
scree plot, 111
unconstrained ordination, 157, 158
used to calculate centroids, 90, 99
vector overlays, 115
vs CA, 117
vs MDS, 121–22
vs PCA, 115–18
Pearson correlation, 113, 115, 117, 127, 144, 163, 185
penalty term
identical for a given number of variables, 140
included in selection criteria, 135
percentage
cover, 32, 92, 172, 185, 186
of samples correctly allocated in CAP, 164
of the fitted variation, 143
of the total sum of squares, 54
of the total variation, 109
of unshared species, 52, 95
percentage of variation explained
by a model ($R^2$), 124
by dbRDA axes, 143, 146
by PCO axes, 106, 107, 108, 111, 112
by predictor variables, 126, 149
PERMANOVA
additivity, 37–38
affinity
with DISTLM, 152
with PCO, 122
analysis
of birds from Borneo, 73
of cryptic fish, 101
of Ekofisk macrofauna, 26
of holdfast invertebrates, 77, 78
of Mediterranean molluscs, 84
of molluscs in holdfasts, 52, 56, 58
of New Zealand fish, 80, 81
of Okura macrofauna, 100, 118
of plankton net data, 60
of Poor Knights Islands fish, 161
of subtidal epibiota, 33, 36, 41
of Tasmanian meiofauna, 44, 47
of Victorian avifauna, 29, 67
of Woodstock plants, 63, 64
assumptions, 21–22, 39–40, 48
components of variation, 43–44
constructing $F$ from EMS, 45–46
contrasts, 41
creating a design file, 23–24
description, 15
dialog, 25
estimating components of variation, 53–54
exchangeable units, 46
expected mean squares, 44–45
experimental designs
asymmetrical, 82–85
BACI/beyond BACI, 85
Latin square, 61
multi-way, 39, 79, 118
nested (hierarchical), 49–53
one-way, 22–23
randomised blocks, 59
repeated measures, 64–68
split-plots, 61–64
that lack replication, 58–61
to detect environmental impacts, 85
two-way crossed, 31–34
unbalanced, 68–69
with covariates, 73–78
features, 15
fixed vs random factors, 41–43
inference space and power, 46–48
interpretation (jointly) with PERMDISP, 94, 104
interpreting interactions, 34–37
linear combinations of mean squares, 78–82
methods of permutation, 38–39
Monte Carlo $P$-values, 28–30
pair-wise comparisons, 26–28
partitioning, 16
pooling or excluding terms, 54–58
table of results, 25
test by permutation, 19–21
test statistic, 19
testing the design, 49
types of SS, 69–73
vs a Mantel test, 19
vs ANOSIM, 30–31
vs CAP, 167, 169–71
PERMANOVA+
advantages shared with PRIMER v6, 9
changes from DOS routines, 9
complements PRIMER tools, 5
general applications, 6
new features of the software, 9–11
PERMDISP
affected by
choice of resemblance measure, 92–94
choice of transformation, 93
affinity with PCO, 122
analysis
of Bumpus’s sparrow data, 89
of cryptic fish, 102, 103
of Norwegian macrofauna, 96
of Okura macrofauna, 98
of Tikus Island corals, 93
as a multivariate Levene’s test, 88–90
description, 87
for crossed designs, 100–103
for nested designs, 96–100
interpretation (jointly) with PERMANOVA, 94, 104
pair-wise tests, 99
$P$-values by permutation, 91
rationale, 87
test based on medians, 91
test of homogeneity in the space of the resemblance
measure, 90–91
to compare beta diversity, 94–96
uses PCO axes to calculate $z$ values, 90–91
with small sample sizes, 96
permutable units. See exchangeable units
permutation
frequency distribution is discreet, 20
in unbalanced designs, 69
inferences at a given level, 24
methods of, 38–39
number of possible $s$, 20
number of unique values obtained, 25
of raw data, 20, 24, 38, 43, 75
of residuals
in DISTLM, 129
in PERMDISP, 87, 91
requirements, 39
under a reduced model, 34, 38, 39
under the full model, 39
power and precision of, 24
random subset of $s$, 20, 125
test. See test by permutation
perpendicular
axes in CAP, 185
axes in dbRDA, 143
axes in PCO, 107
projection in PCA, 105
Pillai’s trace, 168
pitfall traps, 183
plankton net study, 59, 61
planned comparisons. See contrasts
plants. See Woodstock plants and index to data sets
plot factor labels, 23
plots (experimental), 61, 100
point-intercept technique, 62
pollution gradient, 175
Pool... button, 55, 56, 57, 58
pooled terms, 57
pooling
  combining df and SS, 55
distinguished from excluding a term, 54
pooling or excluding terms, 54–58
Poor Knights Islands fish, 160
population
  of control locations, 84
  of levels, 47, 48
to which inferences apply, 48
post hoc, 163
power
  compromised by ranking, 31
denominator degrees of freedom, 31, 47, 49, 57
  increased (perhaps) by pooling, 81
  increased by blocking, 61
  increased by reduced inference space, 47
different methods of permutation, 38
of distance-based tests, 31
of fixed vs random models, 47
of MANOVA statistics, 31, 169
of methods in PERMDISP, 91
of PERMANOVA vs ANOSIM, 31
of PERMANOVA vs CAP, 171
of PERMDISP compromised in crossed designs, 103
of permutation tests, 24, 31
of restricted permutations, 38
to detect interactions, 35
precision of permutation tests, 24
predators, 100
prediction
  of group membership, 158, 164, 166–67, 172–75
  of positions along a gradient, 159, 175–79
predictor variables
  all possible combinations, 134
  building models, 134–36
categorical, 150–54
choose order of fit, 131
conditional on observed values, 152
diagnostics, 148
examine distributions, 133
force inclusion/exclusion, 151
in separate data sheet, 127
in sets, 147
limited by sample size, 126
multi-collinearity, 133
normalised in DISTLM, 137
relative importance and overlap, 130
X matrix, 124
presence/absence, 37, 52, 94, 95
preserve inter-point resemblances, 90
PRIMER v6, 4
  advantages shared by PERMANOVA+, 9
PRIMER-E Ltd, 4
principal components analysis. See PCA
principal coordinates analysis. See PCO
principal response curves (PRC), 68
proc GLM (in SAS), 45
procedural control, 31
procedures for model selection, 134
projection
  in CA, 105
  in PCA, 105
  in PCO, 105, 106, 122
in regression and RDA, 125
matrix (H), 142, 169
of variables onto ordination axes, 185
proximity, 133, 136, 175, 177
pseudo-F
  compared to ANOSIM R, 30
  construction, 19
  finding the right denominator for the test, 45
  from linear combinations of mean squares, 78
  in conditional tests, 129
  in DISTLM, 124, 129
  in PERMANOVA, 19
  negative values of, 31
pseudo-F*, 19
pseudo-replication, 43
pseudo-t, 26, 30
P-value
  being conservative by inclusion of observed test statistic, 20
differences with multiple runs, 21, 24
obtained by Monte Carlo asymptotics, 28
obtained by permutation, 20, 91, 125, 167–69
strength of evidence, 31
Q
Q matrix
  in dbRDA, 181
  in PCO, 106
Q0 matrix (in CAP), 159, 181
quasi F ratios, 81
R
R statistic, 30, 156
R2
  as a model selection criterion, 134
  defined, 124
  increases with the no. of predictor variables, 126
radio button, 12
random
  factor, 42
  sampling, 48
subset of values of F∗, 20
random allocation
  assures exchangeability, 21, 48
in split-plot designs, 61
randomised block design
  assures exchangeability, 21, 48
  in split-plot designs, 61
randomization test. See test by permutation
ranking
  affects additivity, 37, 38
  in ANOSIM, 30
  in BEST/BIOENV, 154
  in MDS, 30, 38, 105, 121
  in ordinal data, 150
  in Spearman correlations, 113, 115, 144, 155, 163
  leads to non-metricity, 31, 111
  loss of information, 31
  non-parametric approaches, 5, 106
randomization test. See test by permutation
rare species, 37, 38
rationale
  for CAP, 157
  for DISTLM/dbRDA, 123
  for PCO, 105–6
  for PERMANOVA, 15
  for PERMDISP, 87
  for ranking in ANOSIM, 30
  for removing terms, 54
  for split-plot designs, 62
RDA
and principal response curves, 68
linear relationships between \( Y \) and \( X \), 133
multivariate multiple regression, 125
real Euclidean axes, 111
reduce dimensionality. See ordination
reduced model, 129
reduced model residuals. See methods of permutation
redundancy analysis. See RDA
redundant variables, 133, 137, 151
reflection of ordination axes, 106
regression
Model II, 152
multiple, 124, 125, 133
multivariate multiple, 123, 125
simple, 125, 126
sum of squares, 124
RELATE, 19, 137
relative abundance information, 95
relative contribution of species, 37
relative importance
of components of variation, 54
of predictor variables, 130, 146, 156
removing terms. See pooling or removing terms
repeated measures, 64–68
repeated values under permutation, 25
replicates. See samples
resemblance matrix
as the basis for PERMANOVA+ routines, 6
resemblance measures
affect additivity, 38
affect dispersions, 93, 94
and beta diversity, 94
and distances to centroids, 17
and ordination, 121
and the triangle inequality, 111
meaning, 112
non-Euclidean, 17, 19, 44, 85, 90, 118, 122, 124, 136
non-metric, 18, 31
properties, 18, 109
semi-metric, 18, 31, 112
residual sum of squares
in a two-way model, 32
in adjusted \( R^2 \), 135
in \( AIC \), \( AIC_c \), and \( BIC \), 135
in DISTLM, 124
in PERMANOVA, 16, 18
leave-one-out, 157, 164
residuals. See methods of permutation
response variables
as dimensions, 16, 105
too many for traditional analysis, 5, 6
right singular vectors, 180
robustness
of ANOVA to heterogeneity, 22
of asymptotic theory, 30
of the general approach, 5
roles of variables, 181
in CAP vs DISTLM, 159
in CAP vs PERMANOVA, 169
in DISTLM/dbRDA, 124
rose-coloured glasses, 164, 171
routines in PERMANOVA+, 6
Roy’s greatest root criterion, 169
rules for deriving EMS, 45
running PERMANOVA, 24–26
S
S(*). See components of variation
sample size
and asymptotic MC \( P \)-value, 83
and correlation among residuals, 39
and correlation of \( z \) values, 96
and methods of permutation, 38
and type I error for asymptotic permutation tests, 129
differences in unbalanced designs, 69
effect on \( BIC \) relative to \( AIC \), 135
equal. See balanced design unequal. See unbalanced designs
warning in PERMDISP, 96
samples, 15
sandflat, 42
sandstone plates, 32
sandstone rocky reefs, 31
SAS computer program, 45, 69
saving a design file, 24
scale of sampling unit, 121
scales of variation, 50
scaling of axes
in CA, 117
in CAP, 160
in CCorA, 183
in dbRDA, 142
in PCA, 117
in PCO, 106, 170
scatter-plot of \( BIC \) vs \( AIC \) values, 141
schematic diagram of
a bivariate system with three groups, 16
a design
asymmetrical (Mediterranean molluscs), 82
nested (holdfast invertebrates), 50
split-plot (Woodstock plants), 61
two-way crossed (subtidal epiobiota), 32
unbalanced (Borneo birds), 71
a violated triangle inequality, 110
additive or multiplicative effects, 37
apparent differences in dispersion caused by an interaction, 103
CAP vs PERMANOVA, 170, 171
differences
in dispersion or correlation structure, 21
in location and/or dispersion, 87
different types of SS, 70
distances to centroids, 88
geometric partitioning, 16
Huygens’ theorem, 17
multivariate interactions, 35
overlap in SS for unbalanced vs balanced designs, 69
partitioning
in regression, 124, 128
showing SS for a conditional test, 129
with different outcomes for marginal/conditional tests, 130
permuting labels, 20
routines in PERMANOVA+, 8
sphericising variables, 181
SS from a distance matrix, 18
the mechanics
of CAP, 160
of dbRDA as a constrained ordination, 142
of DISTLM, 125
of PCO, 107
unconstrained vs constrained ordination, 158
vector overlay (raw Spearman correlations), 114
scree plot, 111
General index

seasons, 97
sediment deposition, 97
selecting
  predictor variables. See building models
  response variables, 50
terms to be pooled, 56
selection
criteria, 134
  procedures, 134
semi-metric, 109
semi-parametric, 6
sensitivity of methods to violated assumptions, 21, 39
separating groups, 157
sequential fit
  in designs with covariates, 74
  in DISTLM, 130
  in unbalanced ANOVA. See Type I SS
sets of variables, 128, 147, 179
sewage outfall, 82
shading, 31
Shannon diversity \((H')\), 126
shift in location/direction. See effect size
shrubs, 150
shuffling of labels. See permutations
sign of ordination axes, 106
significance level, 20, 24, 27, 28, 129
similarity. See dissimilarity
simple linear regression, 126–28
SIMPROF, 7
simulation studies, 39, 91, 96, 129, 135, 171
singular value decomposition (SVD), 180
sizes of effects. See effect size
skewness
  in covariates, 75
  in data cloud, 91
  in environmental variables, 137
  in PCA, 115
  in predictor variables, 133
  in response variables, 5, 93
small sample sizes
  in PERMANOVA. See Monte Carlo P-values
in PERMDISP, 96
snapper, 100, 163
soft-sediment, 22, 95, 98, 175
soldier crab, 42
some cells empty, 71
Sørensen resemblance measure, 95
space of the resemblance measure, 21, 39, 75, 90, 99,
  105, 122, 134, 158, 170
sparsely distributed data
  for CAP, 157
  for PCO, 111
  in DISTLM, 124, 129
  in PERMANOVA, 19–21
spatial
  coordinates, 147
  median, 87, 91
  scales, 43, 50, 52, 53, 57, 61, 64, 76, 82, 84, 94
  variability, 47, 50, 64
  variables, 148, 150
Spearman rank correlation, 113, 114, 115, 117, 144,
  155, 163
species
  abundance data, 17, 115
  characterising group differences, 163
  composition, 94
  species-area relationship, 74
  sphericising variables, 181
  sphericity in repeated measures, 65
spiders. See hunting spiders
spill-over effects, 28
split-plot designs, 58, 61–64
spot-mapping, 71
SPSS, 45
square root. See transformation
squared dissimilarity, 53
stabilising
  relative dispersions, 94
  selection, 88
stacked \(F\) statistic, 170
standard deviation, 53, 59, 180
standardisation of axes, 182
states of a fixed factor, 42
statistical inference. See inference
  step-wise (selection procedure in DISTLM), 134
stress
  definition, 37
  in metric and non-metric MDS, 122
strip transect method, 28
Student, 26
sub-plots, 62
subset of PCO axes for CAP, 157
substratum, 150
subtidal epi-benthos, 31, 36, 40
sum of canonical eigenvalues, 168
sum of squared fixed effects, 42, 43, 46
sum of squared inter-point distances. See Huygens' theorem
sums of squares
  and cross-products (SSCP), 142, 180
  for contrasts, 41
  from a distance matrix, 18–19
  in a one-way model, 16
  in a two-way crossed model, 32
  in regression, 124
symmetric resemblance measures, 18, 109
synchronised permutations, 38
system requirements, 4

T

t statistic, in PERMANOVA, 26
Tasmanian meiofauna, 42, 44, 46, 49
taxonomic distinctness, 56
Terms... button, 55, 57, 58, 61, 63, 74, 77
test based on medians in PERMDISP, 91
test by permutation
  conditional on order statistics, 48
  in CAP, 167–69
  in DISTLM, 125, 129
  in PERMANOVA, 19–21
  in PERMDISP, 91
  logic in mixed/random models, 48
validation, 48
test statistic
  in ANOSIM, 30
  in CAP, 168
  in DISTLM, 124, 129
  in pair-wise comparisons, 26
  in PERMANOVA, 19
testing the design, 49
Thau lagoon bacteria, 148
Tikus Island corals, 92, 172
time as a factor in repeated measures, 65
total sum of squares
  in a two-way model, 32
  in DISTLM/dbRDA, 124
  in PCO, 111
  in PERMANOVA, 15, 18
trace of a matrix, 124
trace statistic, 160, 168
traditional methods
  ANOVA, 19, 21, 22, 56, 81, 88

212
view of the data cloud, 158
violation
  of assumptions
    in MANOVA, 169
    in PERMANOVA, 39
    in repeated measures, 65
  of the triangle inequality, 110
visualise
  fitted model, 142–43
  overall patterns, 142
volume, 74, 130

W
W matrix (from SVD), 180
warning
  no replication at the lowest level, 59
  small sample size in PERMDISP, 96
water displacement, 74
website, 4
weights
  in CAP, 185
  in CCA, 133
  in dbRDA, 146
whole plots, 61
Wilks’ lambda, 169

Windows™ version, 4
within-group sum of squares. See residual sum of squares
within-group variation, 30, 31, 96
Woodstock plants, 62

X
X matrix
  in CAP, 159, 181
  in DISTLM/dbRDA, 124
X⁰ matrix (sphericed), 180

Y
Y matrix
  in CCorA, 181
  in DISTLM/dbRDA, 123
  in RDA, 142
Y⁰ matrix (sphericed), 181

Z
Z matrix (in dbRDA), 142, 183
z values in PERMDISP, 88