

Nested Anova User's Guide

Pierre Legendre
Département de sciences biologiques
Université de Montréal
C.P. 6128, succursale Centre-ville
Montréal, Québec H3C 3J7, Canada

July 2002

E-mail: Pierre.Legendre@umontreal.ca

What does the NesAnova program do?

This program carries out nested multivariate analysis of variance (manova) by canonical redundancy analysis, with permutation testing, using 2 or 3 factors (i.e., one main factor plus 1 or 2 nested factors). If a single factor is present, the program computes simple manova with permutation test. The response subjected to the analysis can be a single variable or a multivariate data table.

The primary interest of this program is to allow analysis of variance to be carried out for community composition data (species presence-absence or abundance data tables) through permutation testing. This type of data can be analysed in three different forms:

- If there are very few zeros in the species response data table, the abundances can be analysed directly without further transformation. This is a rare situation since community composition data are usually obtained across a gradient of some sort, natural or experimental, which aims at maximizing the observed variability, so that many species that have unimodal distributions along the gradient(s) (i.e., those that have an abundance optimum) are represented in the table by many zeros.
- In most cases, the community composition data (species presence-absence or abundance) are transformed in some way. This can be done using one of the transformations proposed by Legendre & Gallagher (2001); these transformations are available in the program, as described below.
- An alternative method is to compute a matrix of ecological distances (**D**) among sites. For species presence-absence data, the one-complements of the indices of Jaccard or Sørensen are often used. For species abundance data, ecologists often rely upon the Steinhaus/Odum/Bray-Curtis distance function. Principal coordinate analysis of that **D** matrix produces a new rectangular data table that can be used as input for nested manova (Legendre & Anderson 1999). The distance functions that are appropriate for various types of data, including species composition data tables, are discussed in Chapter 7 of Legendre & Legendre (1998).

Statistical background

Anova and regression are both expressions of the general linear model. This relationship allows scientists to carry out analysis of variance by regression, or multivariate analysis of variance by canonical analysis, provided that the factors are coded in the correct manner. Coding of the factor in one-way anova can be done using ordinary dummy variables. Coding of the main factors and interaction terms in 2-way or multiway crossed anova or manova requires orthogonal dummy variables; this type of coding is described in Legendre & Anderson (1999, Appendix C). Nested 2-way or 3-way anova does not require the use of orthogonal dummy variables, but it is recommended to use this type of coding, which is fairly easy to understand, because it is unambiguous. In any case, the coding of the nested factors must explicitly express the nesting into the upper factor(s); see examples below.

Nested analysis of variance is used when the classes of an inner factor are unrelated among the classes of the outer factor. For instance, transects are unrelated to each other among sites, even though they may be called transects 1, 2, 3 ... within each site. Nested factors are always random factors. The main

factor may be either fixed or random. This form of anova is discussed in all textbooks of statistics, including Sokal & Rohlf (1995). Examples are given below.

Contrary to crossed 2-way or multiway anova, in nested anova the F -statistics do not all use the residual mean squares as their denominator. The rule for construction of the F -statistics in a 3-way nested anova is described in the following table, where a is the number of classes (groups) in factor A, b the number of classes in factor B, c the number of classes in factor C, and n the number of replicates in each class of factor C:

Source of variation	d.f.	Sum of Squares	Mean Squares	F -statistic
Main factor (A)	$a - 1$	SS_A	$MS_A = SS_A / a - 1$	MS_A / MS_B
1st nested factor (B)	$a(b - 1)$	SS_B	$MS_B = SS_B / a(b - 1)$	MS_B / MS_C
2nd nested factor (C)	$ab(c - 1)$	SS_C	$MS_C = SS_C / ab(c - 1)$	MS_C / MS_{Error}
Within subgroups	$abc(n - 1)$	SS_{Error}	$MS_{Error} = SS_{Error} / abc(n - 1)$	
Total	$abcn - 1$	SS_{Total}		

Programs for canonical analysis, such as CANOCO (ter Braak 1988, ter Braak & Smilauer 1998) or POLYNOMIAL RDACCA (Makarenkov & Legendre 2002), always use the mean squares of the residuals as the denominator of the F -statistic. They are thus unsuitable for nested manova, although special forms of permutation are available in CANOCO v.4 for manova with one nested level.

The number of dummy variables, which is equal to the number of degrees of freedom associated with a factor, necessarily increases between levels of a nested anova. This suggests an easy way of checking the input files: if the number of dummy variables does not increase from one input file to the next, either the files are listed in an incorrect order, or the factors are not coded properly.

Optional tests can be carried out by the program to insure that the files of dummy variables are correctly coded into orthogonal dummy variables. The program checks that the dummy variables within each file sum to 0 and that they are orthogonal.

In all cases, checks for collinearity are performed:

- Check for collinearity of the dummy variables within each factor.
- Check for collinearity of the dummy variables among factors.

Collinearity would prevent inversion of the matrices of sums-of-squares, which is necessary to carry out the regressions.

Nested anova is very costly in the number of dummy variables necessary for coding the factors. So, if they have a choice, users should plan their experiments in such a way that they can be analysed by crossed instead of nested anova. Because they use fewer dummy variables (and degrees of freedom), crossed anova designs have more power to detect small effects than nested designs. Many experiments require, however, the use of nested factors. This is why this program was written (balanced designs only).

Special application: crossed versus nested anova

In some two-factor crossed anovas carried out in community ecology, one of the factors is geographically nested in the other. For instance, consider a study in which fish have been counted along underwater transects on the windward and leeward sides (factor B) of two bays (factor A). Transects are

the replicates in this analysis. A two-way crossed design should be used to test the hypothesis that factor B has an identifiable crossed effect (with respect to A) on the response variables. Crossed anova has more power to detect small effects than a nested analysis. If, however, the interaction between the two factors is significant, one has two options: (1) carry out separate analyses of the two bays, which causes the problem of assembling the results, a posteriori, in a coherent global interpretation. Or (2), consider that the “sides” (factor B) do not actually behave like a crossed factor. One may then approach the problem in a different way by considering that B is nested in the “bays” (factor A) and proceeding with a nested anova or manova. Indeed, in some studies, several kinds of hypotheses can be tested with the same data. The second option presents the advantage of providing a global analysis of the data, capable of showing if the geographically outer factor (bays) and/or the inner factor (sides within bay) are significant.

Input files

The response data table and each of the factors are written to a separate ASCII (text) data file beginning with a header that specifies the number of data rows and the number of variables (columns) in that table. The names of all these files are written to a file that will serve as the main input to the program. The file containing the response data is given first, followed by the file containing the main factor, the first nested factor, and the second nested factor. The easiest way is to place all data files in the folder where the program is found; otherwise, one has to write the full paths to the data files, for the program to find them.

The last-level replicates within groups or subgroups, used to compute the residual sum of squares and residual mean squares, are not coded into a file; coding them would require as many dummy variables as there are residual degrees of freedom: $abc(n - 1)$. These dummy variables are not required since the residual sum of squares can be obtained by subtraction.

Transformations

The program provides the possibility of transforming frequency data, and in particular species abundances, using one of the transformations described by Legendre & Gallagher (2001). These transformations weigh each value in the data table by some function of the values in the same row and/or column (e.g. the sum, or the sum of squares, of the values). The sum of all values in the data table may also be included in the transformation. After Hellinger transformation, for example, computing the Euclidean distance among the rows of the transformed data table is equivalent to computing the Hellinger distance among the rows of the original, untransformed data table. Since the Hellinger distance, as well as the other distances in the proposed set of transformations, are all appropriate for ordination analysis of species data, they are also adequate before manova which, geometrically speaking, is a Euclidean-distance-based method. The dialogue of the program is the following :

```
Transform species abundance data
(Legendre & Gallagher, 2001)
```

- ```
(0) No transformation (Euclidean distance preserved)
(1) Chord distance
(2) Chi-square metric
(3) Chi-square distance
(4) Distance between species profiles
(5) Hellinger distance
```

Answer by giving the number of the selected transformation, or 0 for no transformation.

## **Permutation methods**

Two permutation methods are used in the program, depending on the order of the factor under test:

1. Permutation of the rows of the data matrix **Y**, also called permutation of the raw data, is used to test the significance of the main (or first-order) factor.
2. Permutation of the residuals of the “reduced model” (Freedman & Lane, 1983), which are the residuals of the regression of **Y** on all explanatory variables in the model except those corresponding to the factor under test, is used to test the significance of the nested factors.

The permutations are done without restrictions. Anderson and ter Braak (2002) have shown that the best way to carry out permutation tests in nested anova is to permute the units that are called upon in the denominator mean square of the *F*-statistic, instead of permuting the sampling units (or residuals of some regression model) at random in an unrestricted fashion. For example, in nested anova for one main factor **A** and one nested factor **B**, the blocks of units representing factor **B** should be permuted as whole blocks, when testing factor **A**, since the *F*-statistic is constructed as the ratio of  $MS_A/MS_B$ .

This strategy poses problems in computer programs, however: the permutations are more complex to program than the random permutation of sampling units (or residuals), and the test of each factor has to be done separately since it involves a different method of permutation. This may considerably slow down computer programs. So, a simulation study (Legendre, 2002) was carried out to investigate various unrestricted permutation methods in nested anova and find out, empirically, which methods remain valid and have the best power, under a variety of error conditions, for 2- and 3-factor nested anova and manova. The study showed that the tests used in the present program are valid except under the most extreme error conditions used during the simulation study (cubed exponential random deviates).

The dialogue in the nested anova program is the following:

```
Permutation test:
How many permutations? (e.g. 999, 9999, ...)
No permutation test: type 0
```

One answers by giving the number of permutations to be used, or 0 if no permutation test results are required.

## **Output file**

The main output file “NesAnova.out” contains the results of the anova or manova. An additional output file, called “FittedSc.out”, contains the fitted site scores for each factor in the analysis, as well as a list of the residuals. If the response data matrix contains a single variable, this file contains the fitted and residual site scores that could be obtained by regression analysis. If the response data are multivariate, this file contains the fitted and residual site scores that are linear combinations of the explanatory variables, as found in redundancy analysis (RDA). Note that the response variables are centred at the beginning of the calculations; this is done after transformation if a species data transformations is requested.

## **Disclaimer**

This program is provided without any explicit or implicit warranty of correct functioning. It has been developed as part of a university-based research program. If, however, you should encounter problems with this program, the author will be happy to help solve them. Researchers may use this program for scientific purposes, but the source code remains the property of Pierre Legendre. Users of the program may refer to the present user’s manual as follows:

Legendre, P. 2002. Nested anova user’s guide. Département de sciences biologiques, Université de Montréal. 26 pages. Available from the WWW site <<http://www.fas.umontreal.ca/biol/legendre/>>.

## **Program distribution**

The program to carry out the form of nested anova described in the present user's guide is called NESANOVA. Another program, GENERATE\_X, distributed with NESANOVA, has been written to automatically code into dummy variables the factors of a nested anova comprising one main factor and 1 or 2 nested factors (balanced design). Its use is self-explanatory.

Computer programs written by P. Legendre are available from our base WWWeb site. They include FORTRAN77 source code, documentation, sample files and executable programs. Versions for MacOS (68k or PowerPC), MacOS X, and 32-bit DOS (suitable for DOS sessions under Windows 95/98/NT) are provided. WWWeb address: <<http://www.fas.umontreal.ca/biol/legendre/>>.

The programs are written in FORTRAN77 in order to facilitate diffusion. Indeed, there is a compiler, GNU FORTRAN (or g77; see <http://www.gnu.org/software/fortran/>), that is freely available for this level of FORTRAN, for the DOS/Windows, MacOS X, Unix, and Linux families of operating systems.

Users can modify the parameters ("Parameter" statement), at the beginning of the program, that set the maximum size of the data tables that can be analysed. To run modified versions capable of handling very large data sets, users of MacOS anterior to OS X may have to turn *Virtual memory* on in the *Memory* control panel.

## **References**

- Anderson, M. J. & C. J. F. ter Braak. Permutation tests for multi-factorial analysis of variance. *Journal of Statistical Computation and Simulation* (in press).
- Freedman, D. & D. Lane. 1983. A nonstochastic interpretation of reported significance levels. *Journal of Business & Economic Statistics* 1: 292-298.
- Legendre, P. 2002. NestedAnova simulation report. Département de sciences biologiques, Université de Montréal.
- Legendre, P. & M. J. Anderson. 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs* 69: 1-24.
- Legendre, P. & E. Gallagher. 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia* 129: 271-280.
- Legendre, P., & L. Legendre. 1998. *Numerical ecology, 2nd English edition*. Elsevier, Amsterdam, The Netherlands.
- Makarek, V. & P. Legendre. 2002. Nonlinear redundancy analysis and canonical correspondence analysis based on polynomial regression. *Ecology* 83: 1146-1161.
- Sokal, R. R. & F. J. Rohlf. 1995. *Biometry – The principles and practice of statistics in biological research, 3rd edition*. W. H. Freeman, New York.
- ter Braak, C. J. F. 1988. CANOCO – an extension of DECORANA to analyze species-environment relationships. *Vegetatio* 75: 159-160.
- ter Braak, C. J. F. & P. Smilauer. 1998. *CANOCO reference manual and user's guide to Canoco for Windows: software for canonical community ordination (version 4)*. Microcomputer Power, Ithaca, New York, USA.

**Example 1: Nested anova with one main factor and one nested factor**

The following example is used by Sokal & Rohlf (1995, Box 10.1) to illustrate the calculations involved in nested anova. Two independent measurements (replication level) of the length of the left wing have been made on 4 female mosquitoes *Aedes intrudens* (nested factor) reared in each of 3 cages (main factor). The total number of observations is thus  $3 \times 4 \times 2 = 24$ . The data are the following:

| Wing (mm) | Cage | Female |
|-----------|------|--------|
| 58.5      | 1    | 1      |
| 59.5      | 1    | 1      |
| 77.8      | 1    | 2      |
| 80.9      | 1    | 2      |
| 84.0      | 1    | 3      |
| 83.6      | 1    | 3      |
| 70.1      | 1    | 4      |
| 68.3      | 1    | 4      |
| 69.8      | 2    | 5      |
| 69.8      | 2    | 5      |
| 56.0      | 2    | 6      |
| 54.5      | 2    | 6      |
| 50.7      | 2    | 7      |
| 49.3      | 2    | 7      |
| 63.8      | 2    | 8      |
| 65.8      | 2    | 8      |
| 56.6      | 3    | 9      |
| 57.5      | 3    | 9      |
| 77.8      | 3    | 10     |
| 79.2      | 3    | 10     |
| 69.9      | 3    | 11     |
| 69.2      | 3    | 11     |
| 62.1      | 3    | 12     |
| 64.5      | 3    | 12     |

The information is coded as follows into 3 data files:

| Response.txt | Factor_1.txt | Factor_2.txt         |
|--------------|--------------|----------------------|
| 24 1         | 24 2         | 24 9                 |
| 58.5         | 2 0          | 3 0 0 0 0 0 0 0 0    |
| 59.5         | 2 0          | 3 0 0 0 0 0 0 0 0    |
| 77.8         | 2 0          | -1 2 0 0 0 0 0 0 0   |
| 80.9         | 2 0          | -1 2 0 0 0 0 0 0 0   |
| 84.0         | 2 0          | -1 -1 1 0 0 0 0 0 0  |
| 83.6         | 2 0          | -1 -1 1 0 0 0 0 0 0  |
| 70.1         | 2 0          | -1 -1 -1 0 0 0 0 0 0 |
| 68.3         | 2 0          | -1 -1 -1 0 0 0 0 0 0 |
| 69.8         | -1 1         | 0 0 0 3 0 0 0 0 0    |
| 69.8         | -1 1         | 0 0 0 3 0 0 0 0 0    |
| 56.0         | -1 1         | 0 0 0 -1 2 0 0 0 0   |
| 54.5         | -1 1         | 0 0 0 -1 2 0 0 0 0   |
| 50.7         | -1 1         | 0 0 0 -1 -1 1 0 0 0  |
| 49.3         | -1 1         | 0 0 0 -1 -1 1 0 0 0  |
| 63.8         | -1 1         | 0 0 0 -1 -1 -1 0 0 0 |
| 65.8         | -1 1         | 0 0 0 -1 -1 -1 0 0 0 |
| 56.6         | -1 -1        | 0 0 0 0 0 0 3 0 0    |
| 57.5         | -1 -1        | 0 0 0 0 0 0 3 0 0    |
| 77.8         | -1 -1        | 0 0 0 0 0 0 -1 2 0   |
| 79.2         | -1 -1        | 0 0 0 0 0 0 -1 2 0   |
| 69.9         | -1 -1        | 0 0 0 0 0 0 -1 -1 1  |
| 69.2         | -1 -1        | 0 0 0 0 0 0 -1 -1 1  |
| 62.1         | -1 -1        | 0 0 0 0 0 0 -1 -1 -1 |
| 64.5         | -1 -1        | 0 0 0 0 0 0 -1 -1 -1 |

The first line of each file indicates the number of data rows (24 in this example) and the number of variables (response or dummy variables) to be read in the file. In the files containing the anova factors, the dummy variables are made to sum to 0 and to be orthogonal to one another (cross product = 0).

- Factor 1 (cages) contains 3 classes, hence 2 dummy variables are necessary to code for it.
- Factor 2 (females) contains 4 females per cage. So, 3 dummy variables are necessary to code for them within each cage.

One can easily verify that with this method of coding the factors, the dummy variables are all orthogonal to one another, within as well as across factors. The number of dummy variables necessary to code for a factor is equal to the number of degrees of freedom for this factor in the anova table.

To make file manipulation easier, the various data file names are written to a file that will serve as the main input data file to the program. The “Input\_file.txt” contains the following information, beginning with the number of file names that it contains:

```
Input_file.txt

3
Response.txt
Factor_1.txt
Factor_2.txt

```

This analysis does not require a species data transformation. The dialogue with the program is the following:

(Nested) multivariate analysis of variance  
by canonical redundancy analysis,  
using 1, 2, or 3 factors (one main factor,  
plus 0, 1, or 2 nested factors).

Pierre Legendre  
Département de sciences biologiques  
Université de Montréal  
© Pierre Legendre, 2002

This program can handle a maximum of  
501 objects (rows of all data files)  
500 response variables in matrix Y  
200 dummy var. in all files of anova factors combined  
3 files of anova factors

Name of the file containing the groups of file names?

File containing the file names: Input\_file.txt

Tasks:

- (0) Stop the program
- (1) First, compute the nested anova
- (2) Compute directly the canonical object scores
- (3) Both 1 and 2

**1**

Transform species abundance data  
(Legendre & Gallagher, 2001)

- (0) No transformation (Euclidean distance preserved)
- (1) Chord distance
- (2) Chi-square metric
- (3) Chi-square distance
- (4) Distance between species profiles
- (5) Hellinger distance

**0**

Permutation test:

How many permutations? (e.g. 999, 9999, ...)

No permutation test: type 0

**9999**

Check the coding of factors into orthogonal dummy variables?

- (0) No
- (1) Yes

**1**

```
File of response variables: Response.txt
 24 objects
 1 species
File containing factor 1: Factor_1.txt
 24 objects
 2 dummy variables
File containing factor 2: Factor_2.txt
 24 objects
 9 dummy variables
```

No problem was detected in the orthogonal coding of the dummy variables

No collinearity detected within factor files

No collinearity detected among factor files

Permutation test: please wait...

Computation time for nested anova: 5.92 sec.

*Time: for PPC G4 at 400 MHz*

Results are found in file 'NesAnova.out'

End of the program.

The output file “NesAnova.out” contains the nested anova results that are identical to those reported by Sokal & Rohlf (1995, Box 10.1). They indicate that the main factor (cages) does not have a significant effect but that significant differences are found among flies (nested factor) within cage:

(Nested) multivariate analysis of variance  
by canonical redundancy analysis,  
using 1, 2, or 3 factors (one main factor,  
plus 0, 1, or 2 nested factors).

Pierre Legendre  
Département de sciences biologiques  
Université de Montréal  
© Pierre Legendre, 2002

File containing the file names: Input\_files.txt

File of response variables: Response.txt  
24 objects  
1 species  
No transformation of the species data  
Total sum of squares in Y [transformed] (SSY) = 2401.97333

File containing factor 1: Factor\_1.txt  
24 objects  
2 dummy variables

File containing factor 2: Factor\_2.txt  
24 objects  
9 dummy variables

No problem was detected in the orthogonal coding of the dummy variables

No collinearity detected within factor files

No collinearity detected among factor files

Probabilities after 9999 permutations

| Source of variation | df1 | df2 | SS         | MS        | F         | P(perm)   |
|---------------------|-----|-----|------------|-----------|-----------|-----------|
| Main factor         | 2   | 9   | 665.67583  | 332.83792 | 1.74091   | 0.22970   |
| Nested factor       | 9   | 12  | 1720.67750 | 191.18639 | 146.87815 | 0.00010 * |
| Within subgroups    | 12  |     | 15.62000   | 1.30167   |           |           |
| Total               | 23  |     | 2401.97333 |           |           |           |

Computation time for nested anova: 5.92 sec.



```

4 2 1 1 1 0 0 0 1 0 0 0 0 3 0 1 1 0 0
10 2 5 4 0 0 13 0 0 0 0 0 0 0 50 0 0 0 0
0 2 1 8 0 0 1 0 1 0 0 0 0 2 0 0 2 0 0
1 0 7 9 2 1 7 0 0 1 1 0 0 6 0 1 2 1 2
0 0 4 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 8 0 0 0 3 4 1 2 0 0 0 5 5 4 2 17
1 3 0 3 0 0 9 0 0 0 0 0 0 0 22 0 0 0 0
4 0 7 0 1 0 0 6 1 0 0 0 0 0 92 3 3 1 0
0 0 0 0 0 0 6 0 0 0 0 0 0 0 30 0 0 0 0
0 0 0 1 2 0 0 0 0 0 0 0 0 2 30 0 1 0 8
7 0 1 1 0 0 23 0 0 1 0 0 0 5 85 0 0 0 0
9 2 1 1 0 0 22 1 0 0 1 8 0 0 50 2 1 3 0
2 3 0 2 0 0 12 0 0 0 0 0 0 0 25 0 0 0 1
1 3 0 0 0 0 8 0 0 0 0 0 0 0 22 0 0 0 1
4 3 1 5 2 0 15 4 0 0 0 0 0 5 0 0 0 0 0
3 0 0 0 0 0 3 0 0 0 0 0 0 1 5 4 0 0 0
0 2 2 3 0 0 4 0 0 0 0 0 0 0 0 0 0 1 0
0 0 0 5 0 1 0 10 3 0 0 0 0 3 0 7 2 0 0
15 6 2 13 1 0 5 0 5 0 0 0 10 2 0 2 4 0 0
9 0 1 18 1 1 0 2 2 0 1 0 4 0 0 1 2 1 0
4 7 0 11 0 0 30 0 0 0 0 0 0 4 30 0 0 1 0
7 0 2 4 2 0 16 0 0 0 0 8 0 0 0 0 0 3 0
0 0 0 0 1 0 0 0 0 0 0 0 0 2 0 0 0 1 0
0 4 5 3 0 0 15 1 0 1 0 0 0 0 0 0 0 1 1
3 0 3 10 0 1 10 1 0 0 0 0 0 0 30 1 0 0 2
5 0 9 2 1 2 0 12 0 1 2 0 0 7 0 11 4 0 10
2 3 0 8 0 0 9 0 0 0 0 0 0 0 65 0 0 0 0
0 0 0 33 0 1 0 0 0 1 1 0 46 3 0 0 0 0 31
6 10 5 10 0 1 10 0 1 1 0 0 0 4 20 0 0 0 0
1 12 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 3
3 3 0 5 2 1 46 0 0 0 0 30 0 0 10 0 0 7 0
1 7 0 1 0 0 35 0 0 0 0 25 0 0 7 0 1 0 0
0 0 0 2 0 0 13 0 0 0 1 20 0 0 20 0 2 0 0
2 2 0 1 0 0 33 0 0 0 0 65 0 0 0 0 0 0 0
2 0 0 2 0 0 0 1 0 2 0 10 0 0 0 0 0 0 0
1 0 1 5 0 0 23 3 0 1 0 5 0 12 0 3 2 2 0
8 3 2 0 4 1 0 4 2 1 0 0 0 2 0 8 1 2 0
6 0 3 0 0 0 10 0 1 0 0 0 0 0 0 1 2 1 0
4 4 4 10 0 2 0 0 2 0 0 0 2 1 0 0 0 0 0
0 2 2 7 2 4 0 0 6 0 0 0 6 0 0 0 0 0 0

```

The two islands forming the main factor are coded by a single dummy variable into the file “Islands.txt”:

```

80 1
1
1
1
1
1
1
1
1
1

```













The “Input\_file.txt” contains the following information, beginning with the number of file names that it contains:

```
Input_file.txt

4
Fish.txt
Islands.txt
Sites.txt
Transects.txt

```

This analysis requires a species data transformation; we will first use the chord transformation. The dialogue with the program is the following:

```
(Nested) multivariate analysis of variance
by canonical redundancy analysis,
using 1, 2, or 3 factors (one main factor,
plus 0, 1, or 2 nested factors).
```

```
Pierre Legendre
Département de sciences biologiques
Université de Montréal
© Pierre Legendre, 2002
```

```
This program can handle a maximum of
 501 objects (rows of all data files)
 500 response variables in matrix Y
 200 dummy var. in all files of anova factors combined
 3 files of anova factors
```

```
Name of the file containing the groups of file names?
```

```
File containing the file names: Input_files.txt
```

```
Tasks:
```

- (0) Stop the program
- (1) First, compute the nested anova
- (2) Compute directly the canonical object scores
- (3) Both 1 and 2

```
3
```

```
Transform species abundance data
(Legendre & Gallagher, 2001)
```

- (0) No transformation (Euclidean distance preserved)
- (1) Chord distance
- (2) Chi-square metric
- (3) Chi-square distance
- (4) Distance between species profiles
- (5) Hellinger distance

```
1
```

Permutation test:

How many permutations? (e.g. 999, 9999, ...)

No permutation test: type 0

**999**

Check the coding of factors into orthogonal dummy variables?

(0) No

(1) Yes

**1**

File of response variables:

Fish.txt

80 objects

19 species

Transformation subroutine warning: Row 3 has total species abundance = 0

Transformation subroutine warning: Row 5 has total species abundance = 0

Transformation subroutine warning: Row 7 has total species abundance = 0

Transformation subroutine warning: Row 8 has total species abundance = 0

Transformation subroutine warning: Row 9 has total species abundance = 0

Transformation subroutine warning: Row 10 has total species abundance = 0

Transformation subroutine warning: Row 14 has total species abundance = 0

Transformation subroutine warning: Row 16 has total species abundance = 0

Transformation subroutine warning: Row 19 has total species abundance = 0

Transformation subroutine warning: Row 20 has total species abundance = 0

Transformation subroutine warning: Row 34 has total species abundance = 0

Transformation subroutine warning: Row 35 has total species abundance = 0

Transformation subroutine warning: Row 36 has total species abundance = 0

Transformation subroutine warning: Row 37 has total species abundance = 0

Transformation subroutine warning: Row 38 has total species abundance = 0

Transformation subroutine warning: Row 39 has total species abundance = 0

Transformation subroutine warning: Row 40 has total species abundance = 0

File containing factor 1:

Islands.txt

80 objects

1 dummy variables

File containing factor 2:

Sites.txt

80 objects

6 dummy variables

File containing factor 3:

Transects.txt

80 objects

32 dummy variables

No problem was detected in the orthogonal coding of the dummy variables

No collinearity detected within factor files

No collinearity detected among factor files

Permutation test: please wait...

Permutation no 990

*In large problems, this counter changes every 10 permutations*

Computation time for nested anova: 162.97 sec.

Computation time for canonical object scores: 0.13 sec.

Results are found in file 'NesAnova.out'  
and in file 'FittedSc.out'

End of the program.

The computation times reported by the program were for a PPC G4 processor operating at 400 MHz. The warnings issued by the Transformation subroutine simply indicate that there are data rows in which the sum of the species abundances is zero. This does not impair the transformation nor the analysis of variance. The manova results, found in file "NesAnova.out", follow:

(Nested) multivariate analysis of variance  
by canonical redundancy analysis,  
using 1, 2, or 3 factors (one main factor,  
plus 0, 1, or 2 nested factors).

Pierre Legendre  
Département de sciences biologiques  
Université de Montréal  
© Pierre Legendre, 2002

File containing the file names: Input\_files.txt

File of response variables: Fish.txt

80 objects

19 species

Chord transformation

Total sum of squares in Y [transformed] (SSY) = 51.03347

File containing factor 1: Islands.txt

80 objects

1 dummy variables

File containing factor 2: Sites.txt

80 objects

6 dummy variables

File containing factor 3: Transects.txt

80 objects

32 dummy variables

No problem was detected in the orthogonal coding of the dummy variables

No collinearity detected within factor files

No collinearity detected among factor files

Probabilities after 999 permutations

| Source of variation | df1 | df2 | SS       | MS      | F       | P(perm)   |
|---------------------|-----|-----|----------|---------|---------|-----------|
| Main factor         | 1   | 6   | 6.64825  | 6.64825 | 4.40726 | 0.00100 * |
| 1st nested factor   | 6   | 32  | 9.05086  | 1.50848 | 2.31905 | 0.00100 * |
| 2nd nested factor   | 32  | 40  | 20.81513 | 0.65047 | 1.79203 | 0.00100 * |
| Within subgroups    | 40  |     | 14.51923 | 0.36298 |         |           |
| Total               | 79  |     | 51.03347 |         |         |           |

Computation time for nested anova: 162.97 sec.

The results, obtained using the chord transformation, indicate that all three factors are highly significant. Comparison of the mean squares indicates that the differences between islands are more important than the differences among sites of an island, which are in turn more important than the differences among transects at a site. The Hellinger transformation gave essentially the same results:

| Source of variation | df1 | df2 | SS       | MS      | F       | P(perm)   |
|---------------------|-----|-----|----------|---------|---------|-----------|
| Main factor         | 1   | 6   | 7.08773  | 7.08773 | 5.48380 | 0.00100 * |
| 1st nested factor   | 6   | 32  | 7.75492  | 1.29249 | 2.20990 | 0.00100 * |
| 2nd nested factor   | 32  | 40  | 18.71562 | 0.58486 | 1.61384 | 0.00100 * |
| Within subgroups    | 40  |     | 14.49619 | 0.36240 |         |           |
| Total               | 79  |     | 48.05446 |         |         |           |

The other forms of transformation did not provide as much power to the analysis. The results for the chi-square distance are the following; identical test results are obtained using the chi-square metric. The results of the tests are not as strongly significant as with the chord or Hellinger transformations:

| Source of variation | df1 | df2 | SS        | MS       | F       | P(perm)   |
|---------------------|-----|-----|-----------|----------|---------|-----------|
| Main factor         | 1   | 6   | 23.90553  | 23.90553 | 2.40288 | 0.00200 * |
| 1st nested factor   | 6   | 32  | 59.69211  | 9.94868  | 2.26307 | 0.00100 * |
| 2nd nested factor   | 32  | 40  | 140.67510 | 4.39610  | 1.20637 | 0.00600 * |
| Within subgroups    | 40  |     | 145.76316 | 3.64408  |         |           |
| Total               | 79  |     | 370.03590 |          |         |           |

The results after the transformation into species profiles also display lower power than with the chord or Hellinger transformations:

| Source of variation | df1 | df2 | SS       | MS      | F       | P(perm)   |
|---------------------|-----|-----|----------|---------|---------|-----------|
| Main factor         | 1   | 6   | 2.46526  | 2.46526 | 3.21083 | 0.01400 * |
| 1st nested factor   | 6   | 32  | 4.60676  | 0.76779 | 2.74138 | 0.00100 * |
| 2nd nested factor   | 32  | 40  | 8.96241  | 0.28008 | 1.75807 | 0.00100 * |
| Within subgroups    | 40  |     | 6.37235  | 0.15931 |         |           |
| Total               | 79  |     | 22.40678 |         |         |           |

Carrying out the analysis on untransformed species abundance data produces the least significant results of the lot, in which the variation between islands is weak:

| Source of variation | df1 | df2 | SS          | MS         | F       | P(perm)   |
|---------------------|-----|-----|-------------|------------|---------|-----------|
| Main factor         | 1   | 6   | 7491.96250  | 7491.96250 | 3.67855 | 0.01800 * |
| 1st nested factor   | 6   | 32  | 12219.97500 | 2036.66250 | 2.24889 | 0.00700 * |
| 2nd nested factor   | 32  | 40  | 28980.20000 | 905.63125  | 1.87409 | 0.00100 * |
| Within subgroups    | 40  |     | 19329.50000 | 483.23750  |         |           |
| Total               | 79  |     | 68021.63750 |            |         |           |

In this problem as in many others, application of the chord or Hellinger transformation to the species abundance data, prior to the linear analysis, generated geometric relationships among the points in Euclidean space such that the effect of the factors was more significant than without the transformation.

The file of fitted site scores is not presented because of its length. It is found among the files distributed with the program.

\* \* \* \* \*

## Unix/DOS user's notes prepared by Philippe Casgrain

The Unix (including MacOS X) and DOS versions of this program were built with g77, the GNU FORTRAN compiler. They are command-line tools, which means that they must be started from the command line.

Furthermore, files created by the program, such as our ".out" files, cannot be deleted by the FORTRAN program. If, after launching, the program ends abruptly and a message is displayed, such as:

```
open: 'new' file exists
apparent state: unit 4 named NesAnova.out
lately writing direct unformatted external IO
Abort
```

this means that a file called "NesAnova.out" already exists in the current directory. Rename or remove that file before running the program again. This is a feature, not a bug.

### Unix instructions

**1.** Open a new shell.

MacOS X users: open /Applications/Utilities/Terminal

**2.** At the prompt

(e.g. "[localhost:~] username%") type:

"cd /path/to/the/program/"

where "/path/to/the/program/" represents the directory where the program is found.

Examples: /Applications, ~/Desktop, etc.

Don't forget that Unix systems are case-sensitive: upper- and lowercase letters are different!

**3.** Press the *Return* key.

**4.** Type the name of the program to start it.

Example: ./nesanova

(the prefix "." is essential if the program is not part of your usual path for command-line utilities)

**5.** Press the *Return* key.

**6.** Follow the on-screen instructions.

### DOS instructions

**1.** Open a new shell: from the Start menu, choose Programs Accessories MS-DOS

**2.** At the DOS prompt (e.g., C:\WINDOWS\>),

type "cd c:\path\to\the\program"

where "\path\to\the\program" represents the directory where the program is found.

Examples: c:\tmp, c:\windows\desktop, etc.

Example: nesanova.exe

## **Appendix 1 — Computation method for permutation tests in nested anova**

The calculations involve regressions of the variables forming the response matrix **Y** against blocks of dummy variables coding for either one factor, or all factors. The regressions produce matrices of fitted values (**Fit**) and matrices of residuals (**Res**) of the regressions. The sums-of-squares of these matrices are computed. These sums-of-squares (SS) provide the values from which the *F*-statistics are computed.

Following the notation of variation partitioning (Legendre & Legendre 1998, Section 10.3.5), the total SS in the response matrix **Y** is decomposed into a fraction [a] explained by the main factor A, a fraction [b] explained by the 1st nested factor B, eventually a fraction [c] explained by the 2nd nested factor C, and a fraction [d] of unexplained (i.e., residual) variation. If the factors have been coded correctly, the dummy variables coding for the various factors are orthogonal among factors and the fractions of variation are independent of one another.

- Factor A: Put the variables coding for factor A into matrix **X**.  
 Compute the residualizing matrix **ResidA** =  $\mathbf{X}[\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'$ .  
 Compute matrix **FitA** = **ResidA** × **Y** and its sum-of-squares [a].  
 Compute matrix **ResA** = **Y** – **FitA**.
- Factor B: Put the variables coding for factor B into matrix **X**.  
 Compute the residualizing matrix **ResidB** =  $\mathbf{X}[\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'$ .  
 Compute matrix **FitB** = **ResidB** × **Y** and its sum-of-squares [b].  
 Compute matrix **ResB** = **Y** – **FitB**.
- Factor C: Put the variables coding for factor C into matrix **X**.  
 Compute the residualizing matrix **ResidC** =  $\mathbf{X}[\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'$ .  
 Compute matrix **FitC** = **ResidC** × **Y** and its sum-of-squares [c].  
 Compute matrix **ResC** = **Y** – **FitC**.
- A+B Put the variables coding for factor A and B into matrix **X**.  
 Compute the residualizing matrix **ResidAB** =  $\mathbf{X}[\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'$ .  
 Compute matrix **FitAB** = **ResidAB** × **Y**.  
 Compute matrix **ResAB** = **Y** – **FitAB**.
- A+C Put the variables coding for factor A and C into matrix **X**.  
 Compute the residualizing matrix **ResidAC** =  $\mathbf{X}[\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'$ .  
 Compute matrix **FitAC** = **ResidAC** × **Y**.  
 Compute matrix **ResAC** = **Y** – **FitAC**.
- A+B+C Put the variables coding for factor A, B and C into matrix **X**.  
 Compute the residualizing matrix **ResidABC** =  $\mathbf{X}[\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'$ .  
 Compute matrix **FitABC** = **ResidABC** × **Y** and its sum-of-squares [a+b+c].  
 Compute the sum-of-squares in **Y**, SS<sub>Y</sub>.  
 Check that [a+b+c] = [a] + [b] + [c]. Compute [d] = SS<sub>Y</sub> – [a+b+c].
- Calculate the degrees of freedom associated with each fraction, as shown in the table on page 2.  
 Notation for each factor (as in the tables of manova results produced by the program): df1 = degrees of freedom of the numerator, df2 = degrees of freedom of the denominator.
- Compute  $F(A) = [a] \times df2(A) / [b] \times df1(A)$   
 $F(B) = [b] \times df2(B) / [c] \times df1(B)$   
 $F(C) = [c] \times df2(C) / [d] \times df1(C)$

### Permutation tests

Since there are no covariables in the three tests of significance, a permutation of the data matrix  $\mathbf{Y}$  is equivalent to a permutation of residuals under the reduced regression model (Legendre & Legendre 1998, Table 11.7). Permutation of the residuals of the full regression model is obtained by permuting the residual matrices **ResA**, **ResB** or **ResC** (Anderson & Legendre 1999).

Generate a permutation of the integers  $1 \dots n$  corresponding to the  $n$  objects.

Test of A: Permute **Y**: permute the rows of **Y** to obtain  $\mathbf{Y}^*$ .

Compute matrix  $\mathbf{FitA}^* = \mathbf{ResidA} \times \mathbf{Y}^*$  and its sum-of-squares  $[a^*]$ .

Compute matrix  $\mathbf{FitB}^* = \mathbf{ResidB} \times \mathbf{Y}^*$  and its sum-of-squares  $[b^*]$ .

$F^*(A) = [a^*] \times df2(A) / [b^*] \times df1(A)$

Test of B: Permute **ResAC** to obtain  $\mathbf{Y}^*$ .

Compute matrix  $\mathbf{FitB}^* = \mathbf{ResidB} \times \mathbf{Y}^*$  and its sum-of-squares  $[b^*]$ , unless computed above.

Compute matrix  $\mathbf{FitC}^* = \mathbf{ResidC} \times \mathbf{Y}^*$  and its sum-of-squares  $[c^*]$ .

$F^*(A) = [b^*] \times df2(B) / [c^*] \times df1(B)$

Test of C: Permute **ResAB** to obtain  $\mathbf{Y}^*$ .

Compute the sum-of-squares in  $\mathbf{Y}^*$ ,  $SSY^*$ .

Compute matrix  $\mathbf{FitC}^* = \mathbf{ResidC} \times \mathbf{Y}^*$  and its sum-of-squares  $[c^*]$ , unless computed above.

Compute matrix  $\mathbf{FitABC}^* = \mathbf{ResidABC} \times \mathbf{Y}^*$  and its sum-of-squares  $[a+b+c^*]$ .

Compute  $[d^*] = SSY^* - [a+b+c^*]$ .

$F^*(A) = [c^*] \times df2(C) / [d^*] \times df1(C)$

After  $n_{perm}$  permutations, find the position of  $F(A)$  in the series of values  $\{F(A), F^*(A)_1, F^*(A)_2, \dots, F^*(A)_{n_{perm}}\}$  and calculate the associated probability  $P(A)$  (Legendre & Legendre 1998, Section 1.2). Report it in the manova table. Proceed likewise to estimate the permutational probabilities  $P(B)$  and  $P(C)$ .