

Congruence among distance matrices: Program CADM User's Guide

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

November 2004

Internet: Pierre.Legendre@umontreal.ca

What does program CADM do?

This program computes a test of the congruence among several distance matrices, using Kendall's W statistic. The test, which is meant to be applied prior to using several data matrices together in a combined phylogenetic or other type of multivariate analysis, is described in Legendre & Lapointe (2004).

Principle of the test (H_0 : Incongruence of all matrices):

1. Read in the distance matrices and write them out to a table, each matrix being unfolded onto a row of that table.
2. Transform the distances from each matrix into ranks.
3. Compute Kendall's coefficient of concordance, W .
4. Permute each matrix using a "matrix permutation" procedure, as in the Mantel test, and compute W^* .
5. Repeat step 4 a large number of times.
6. Add the reference value of W to the distribution of W^* values. Calculate the permutational probability associated to W .

The global CADM test is followed by tests of significance in which each distance matrix, in turn, is permuted (H_0 : This distance matrix is incongruent with the other matrices). Complementary Mantel tests, based upon ranks, are also computed; they allow users to determine what are the groups of congruent distance matrices. The choice is left to users to determine if they want the program to compute the tests of congruence only, the Mantel tests based on ranks only, or both:

- (1) Compute and test congruence only
- (2) Compute and test Mantel on ranked distances only
- (3) Both congruence and Mantel

The number of permutations of the two forms of tests can be set separately.

Within a given permutation test, the three statistics W , χ^2 , and S , are monotonic to one another. Since they produce the same permutational probability, they are equivalent for permutation testing. S is a sum-of-squares statistic over the column sums of ranks R_i . Note: W is linearly related to the mean of the Spearman r statistics among all pairs of "judges" (matrices) (Siegel 1956: 232).

The program allows the calculation of the test using asymmetric matrices. This is used only if one of the matrices remains asymmetric, i.e., is not transformed into a symmetric matrix by the program.

The program also allows users to give different weights to the matrices subjected to the analysis. For matrices containing variables that are of the same nature, the matrices may, for instance, be weighted proportionally to the number of characters in each data table from which a distance matrix was computed.

The solution requiring the fewest assumptions is, however, to give equal weights to all matrices subjected to this analysis. This is an application of the principle of parsimony (Ockham's razor).

Input files

There may be a single data file containing all distance matrices, or a different file for each distance matrix:

Data file(s) (containing square distance matrices):
 (1) A single file containing all matrices?
 (2) Separate files for the different matrices?

Each distance matrix is in ASCII, square and with the diagonal. In the case of option 1, all distance matrices are written into a single file, one after the other. An example where separate files are used is presented below.

Output file

See example below.

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. 2001. Congruence among distance matrices: Program CADM user's guide. Département de sciences biologiques, Université de Montréal. 7 pages.

Technical notes

The program is available in a variety of forms:

- FORTRAN source code for Macintosh (file CADM.f), which can be compiled using a FORTRAN compiler. The user may modify the Parameter statement at the beginning of the program, which fixes the size of the largest data matrix that can be analysed (nmax = maximum number of observations, matmax = maximum number of distance matrices).
- FORTRAN source code for DOS or Windows (file CADM.FOR), which can be compiled using a FORTRAN compiler. The user may modify the Parameter statement at the beginning of the program, which fixes the size of the largest data matrix that can be analysed (nmax = maximum number of observations, matmax = maximum number of distance matrices).
- Compiled version for PowerPC processors for Macintosh (file CADM/PPC). The maximum size of the data matrix is 120 objects and 20 distance matrices. The program requires 5.6 Mb RAM for running.
- Compiled version for IBM compatible PC (file CADM.EXE). The maximum size of the data matrix is 120 objects and 20 distance matrices. The program has been compiled for 32-bit operating systems (i.e. Windows95 or WindowsNT) and requires 5.6 Mb RAM for most situations.

References

Legendre, P. and F.-J. Lapointe. 2004. Assessing congruence among distance matrices: single malt Scotch whiskies revisited. *Australian and New Zealand Journal of Statistics* 46: 615-629.

A preprint is available on the WWW site <<http://www.fas.umontreal.ca/biol/legendre/>>. Select “Tirés-à-part / Reprints” in the left-hand column of the introduction page. You will find the title of the paper on the “Tirés-à-part / Reprints” page. Click on the title and you will receive a copy of the preprint in Adobe Acrobat format (pdf).

Siegel, S. 1956. Nonparametric statistics for the behavioral sciences. McGraw-Hill Series in Psychology, McGraw-Hill, New York.

Appendix 1 – Test run

Three distance matrices are presented in the Legendre & Lapointe (2004) paper:

Distance matrix 1: Serology_9x9.txt. This distance matrix is asymmetric. It will be turned into a symmetric matrix within the computer run.

0.00	0.53	0.77	0.82	0.89	0.89	0.86	0.86	0.70
0.59	0.00	0.64	0.55	0.76	0.92	0.85	0.81	0.78
0.92	0.74	0.00	0.38	0.99	0.91	1.07	0.68	0.77
0.66	0.65	0.43	0.00	0.92	0.89	1.03	0.91	0.84
0.85	0.83	0.97	0.97	0.00	0.59	0.70	0.80	0.83
0.88	0.91	0.89	0.87	0.57	0.00	0.54	1.06	0.95
0.80	0.77	1.04	1.00	0.75	0.65	0.00	0.87	0.80
0.88	0.77	0.72	0.95	0.89	0.99	0.78	0.00	0.84
0.75	0.74	0.84	0.80	0.84	0.96	0.81	0.83	0.00

Distance matrix 2: DNA_9x9.txt. This distance matrix is asymmetric. It will be turned into a symmetric matrix within the computer run.

0.00	0.95	0.88	0.86	0.96	0.88	0.92	0.92	0.94
0.95	0.00	0.99	0.98	1.01	0.97	0.97	0.93	0.99
0.90	1.00	0.00	0.81	1.01	0.94	0.96	0.93	0.97
0.84	0.98	0.86	0.00	1.01	0.91	0.94	0.94	0.92
0.96	1.00	0.99	0.99	0.00	0.84	0.78	0.94	0.95
0.90	0.97	0.94	0.95	0.73	0.00	0.63	0.91	0.91
0.92	0.97	0.96	0.94	0.79	0.63	0.00	0.94	0.94
0.90	0.93	0.93	0.93	0.95	0.91	0.94	0.00	0.90
0.92	0.99	0.96	0.92	0.98	0.91	0.94	0.90	0.00

Distance matrix 3: Anatomy_9x9.txt. This distance matrix is symmetric.

0.00	0.58	0.97	0.99	0.79	0.83	0.59	0.81	0.79
0.58	0.00	0.77	0.82	0.68	0.63	0.53	0.55	0.56
0.97	0.77	0.00	0.49	0.92	0.79	0.71	0.98	0.94
0.99	0.82	0.49	0.00	1.12	0.90	0.78	0.99	1.01
0.79	0.68	0.92	1.12	0.00	0.40	0.30	0.75	0.69
0.83	0.63	0.79	0.90	0.40	0.00	0.21	0.72	0.66
0.59	0.53	0.71	0.78	0.30	0.21	0.00	0.58	0.54
0.81	0.55	0.98	0.99	0.75	0.72	0.58	0.00	0.33
0.79	0.56	0.94	1.01	0.69	0.66	0.54	0.33	0.00

The program dialogue window contains the following information:

Français: tapez (1)
 English: type (2)
 2

Test of congruence
 among several distance matrices (CADM)

Pierre Legendre
 Département de sciences biologiques
 Université de Montréal.
 © Pierre Legendre, 2001

Data file(s) (containing square distance matrices):
 (1) A single file containing all matrices?
 (2) Separate files for the different matrices?
 2

Are all distance matrices symmetric?
 (1) At least one matrix is non-symmetric
 (2) All distance matrices are symmetric
 1

How many matrices?
 3
 How many objects?
 9

Name of distance file No. 1
 Input distance file No. 1: Serology_9x9.txt

Name of distance file No. 2
 Input distance file No. 2: DNA_9x9.txt

Name of distance file No. 3
 Input distance file No. 3: Anatomy_9x9.txt

Matrix 1:
 (0) This matrix is symmetric
 (1) Turn it into a symmetric matrix by averaging the
 two triangular half-matrices (recommended solution)
 (2) Turn it into a skew-symmetric matrix $((\text{sup}-\text{inf})/2)$
 (3) Turn it into a skew-symmetric matrix $((\text{inf}-\text{sup})/2)$
 (4) Leave it non-symmetric
 1

Matrix 1 (9 rows and 9 columns):
 (0) Do not print
 (1) Print
 0

Matrix 2:

- (0) This matrix is symmetric
 - (1) Turn it into a symmetric matrix by averaging the two triangular half-matrices (recommended solution)
 - (2) Turn it into a skew-symmetric matrix $((\text{sup-inf})/2)$
 - (3) Turn it into a skew-symmetric matrix $((\text{inf-sup})/2)$
 - (4) Leave it non-symmetric
- 1

Matrix 2 (9 rows and 9 columns):

- (0) Do not print
 - (1) Print
- 0

Matrix 3:

- (0) This matrix is symmetric
 - (1) Turn it into a symmetric matrix by averaging the two triangular half-matrices (recommended solution)
 - (2) Turn it into a skew-symmetric matrix $((\text{sup-inf})/2)$
 - (3) Turn it into a skew-symmetric matrix $((\text{inf-sup})/2)$
 - (4) Leave it non-symmetric
- 0

Matrix 3 (9 rows and 9 columns):

- (0) Do not print
 - (1) Print
- 0

Weighting the distance matrices:

- (1) Give the same weight to all matrices
 - (2) Weight the matrices differently
- 1

Equal weights to all matrices

- (1) Compute and test congruence only
 - (2) Compute and test Mantel on ranked distances only
 - (3) Both congruence and Mantel
- 3

How many permutations for the test of congruence?

(E.g., 999, 9999, ...). No permutation test: type 0.
9999

How many permutations for the Mantel tests?

(E.g., 999, 9999, ...). No permutation test: type 0.
9999

Computation time: 2.55 sec.

Results are found in file 'CADM.out'

End of the program.

Output file

The output file ("CADM.out") contains the following information:

Test of congruence
among several distance matrices (CADM)

Pierre Legendre
Département de sciences biologiques
Université de Montréal.
© Pierre Legendre, 2001

Input distance file No. 1: Serology_9x9.txt
Input distance file No. 2: DNA_9x9.txt
Input distance file No. 3: Anatomy_9x9.txt

Number of permutations = 9999, symmetric matrices

Friedman's Chi-square =	52.29852	<u>Statistics, global CADM test</u>
Kendall's W	= 0.49808	
with correction for tied ranks		

Global test. H0: Incongruence of all matrices.	<u>Global CADM test of significance</u>
All matrices are independently permuted.	

Prob(chi-square, 9999 perm.) = 0.01340

Test for matrix 1. H0: This matrix is incongruent.	<u>A posteriori CADM test, matrix 1</u>
Matrix 1 is permuted.	

Prob(chi-square, 9999 perm.) = 0.02960

Test for matrix 2. H0: This matrix is incongruent.	<u>A posteriori CADM test, matrix 2</u>
Matrix 2 is permuted.	

Prob(chi-square, 9999 perm.) = 0.10400

Test for matrix 3. H0: This matrix is incongruent.	<u>A posteriori CADM test, matrix 3</u>
Matrix 3 is permuted.	

Prob(chi-square, 9999 perm.) = 0.01280

Resemblance among the distance matrices:

Mantel tests

Matrix of Mantel correlations based upon ranks
(Spearman correlations)

1.00000	0.17687	0.36471
0.17687	1.00000	0.19965
0.36471	0.19965	1.00000

Permutation tests of the Mantel statistics:

One-tailed probability in the upper tail ($H_1: r > 0$)

Number of permutations = 9999, symmetric matrices

1.00000	0.13690	0.02510
0.13690	1.00000	0.18570
0.02510	0.18570	1.00000

Computation time: 2.55 sec.

Appendix 2 – Unix and Windows user's notes

The Unix (MacOS X) and DOS versions of this program were built with g77, the GNU FORTRAN compiler.

Click on the program's icon to start it, then answer the questions that appear on the screen. You will be asked to provide the name of the file containing the data. In Windows, make sure the file name is DOS-compatible (maximum 8 characters, no space, etc.). If the input file has an extension, like "txt", you have to type it when you provide the name of the input file, even if the extension does not appear in the file list.

In Unix and Windows, the file "CADM.out" produced by the program cannot be deleted by the FORTRAN program during a following run. If, after launching, the program ends abruptly and a message is displayed, such as:

In Mac OS X: open: 'new' file exists
 apparent state: unit 7 named CADM.out
 last format: list io
 lately reading direct formatted external IO
 Abort trap
 logout

In Windows: Abnormal program termination following which the console window shuts off,
 this means that a file called "CADM.out" already exists in the current directory. Rename or remove that file before running the program again.