A Manual for the Multivariate Permutation Test for Correlations

Jennifer Urbano Blackford, Geunyoung Kim, Niels Waller & Paul Yoder

Table of Contents

1 Introduction

This manual is a guide for the multivariate permutation test (MPT) for correlations using the R program, as demonstrated in Yoder, Blackford, Waller, and Kim (2003). This manual will be most useful for people who understand the basic concept of multivariate permutation tests, regardless of programming skills. Conceptual ideas related to multivariate permutation testing will not be discussed here. Please refer to Westfall & Young (1993) and Yoder, Blackford, Waller, and Kim (2003) for background.

2 Software Requirements

R is a programming environment that lends itself well to novel statistical techniques, especially those that require matrix manipulation. This makes R well suited for multivariate permutation testing. The R program is freeware and is available for download from http://r-project.org/. Recommend resources for learning R are listed in Appendix A.

3 MPT.Corr Function

Description

The MPT.Corr function will perform a multivariate permutation test of correlations for a data matrix with one criterion variable and multiple predictor variables.

Usage

```
MPT.Corr(observed.data, criterion, predictor.start,
    predictor.end, output.file="Results.txt", test=2,
    tail=2, alpha=.05, output=0, permutations=10000)
```

Arguments

`observed.data`	Matrix containing data to be analyzed.
`criterion`	Column number of criterion variable.
`predictor.start`	Column number for start of predictor variables.
`predictor.end`	Column number for end of predictor variables.
`output.file`	File and path for results output. The default is "Results.txt".

`test`	1 = one tailed test, 2 = two tailed test (default)
`tail`	-1 = test negative correlation 1 = test positive correlation 2 = two tailed test (default)
`alpha`	Alpha level; the default is .05.
`output`	0 = Minimal: correlations, MPT results (default) 1 = Full: predictor data, criterion data, correlations, empirical test distribution, MPT results
`permutations`	Number of permutations; the default is 10,000.

Usage Example

```
MPT.Corr(observed.data=test.data,
     criterion=1,predictor.start=2,
     predictor.end=9,output.file="c:\\results.txt",
     test=1, tail= -1, alpha=.05, output=0,
     permutations=10000)
```

4 Data Requirements

This program assumes that data are arranged with subjects or cases in rows and variables in columns. It is strongly suggested that variable names be included in the data file as this information is used in the results output. Variables will be automatically assigned labels V1…Vn if variable names are not provided. Missing data must be specified when the data are imported into R. Pairwise deletion will be used for missing data. The user is responsible for importing the data into R, and instructions for various file formats are shown below.

Sample data is provided for practice importing data and running the MPT program. See Appendix B for sample data or download the data file from http://www.vanderbilt.edu/quantmetheval/downloads.htm.

5 Data Import

Data are most easily imported into R from a text file. Data can usually be exported from other formats into a text file and then imported into R. The '`read.table`' function provides a useful way for importing data from a text file. However, for very small datasets, the data may be directly entered into R using the '`edit`' function. Functions also exist for importing data directly from common statistical packages and relational

databases. Details about these functions can be found in the "R Data Import/Export" manual (see Appendix A).

If the data are in a comma-delimited (.csv) or tab-delimited (.txt) text file, '`read.table`', provides an effective way of importing data. This function has many optional arguments; more details can be found in "The R Reference Index manual" (see Appendix A). The following commands are directly applicable in R for Windows. Use of one of these will satisfy most requirements for importing a data file into an R object called '`test.data`'.

Comma-delimited file with headers

```
test.data<-read.table("c:\\TestData.csv", header=TRUE,
           sep=",")
```

Comma-delimited file with headers and missing data specified as '.'

```
test.data<-read.table("c:\\TestData.txt", header=TRUE,
           sep=",", na.strings=".")
```

Comma-delimited file without headers

```
test.data<-read.table("c:\\TestData.csv", header=FALSE,
           sep=",") #read in a comma-delimited file
```

Tab-delimited file with headers

```
test.data<-read.table("c:\\TestData.txt", header=TRUE,
           sep="")
```

You can also create a blank matrix or data frame and then use '`edit`' to invoke a spreadsheet-like editor in R. Data can be directly entered into the spreadsheet.

Create matrix of 0s with 3 rows and 6 columns

```
test.data<-matrix(0,3,6)
```

Invoke the editor and save changes

```
test.data<-edit(test.data)
```

5 Program Execution

1. Save the program file and data file onto your c: drive.

2. Start R by running the R executable file.

3. Source the program file. This will read in the program file and load the function called MPT. You can source the program file by selecting "source R code" from the file menu or by typing the command '`source`' at the command line with the path and name of the program file in parentheses. This example will read the program called MPT.r found on the c: drive.

```
source("C:\\MPTCorr.r")
```

4. Import data (see section 5). This example will read in the comma-delimited text file called TestData.csv (see Appendix B) into an R object called "test.data".

```
test.data<-read.table("c:\\TestData.csv",
           header=TRUE, sep=",")
```

5. Run the program by calling the function and including all arguments. This example specifies a one-tailed test for negative correlations. The default values of alpha=.05 and permutations=10,000 did not need to be specified. This will only take a few seconds to run; however, larger data files will take longer to run and speed will be dependent on processor speeds.

```
MPT.Corr(observed.data=test.data,
    criterion=1,predictor.start=2,
    predictor.end=9,output.file=
    "c:\\results.txt", test=1, tail= -1)
```

6. The output of the program will be written to the file specified in the arguments. The argument 'output' specifies content and format of the output file. For the "Observed Correlations & p-values" section, variable names are in column 1, observed r in column 2, and the standard p-value associated with the correlation in column 3. For the "Results of MPT for Correlations" section, the test number is in the first column, followed by the variable name, correlation value, standard p-value, and the exact p-value calculated by the multivariate permutation test. Only variables with exact p-value less than or equal to the specified alpha will be printed.

```
Multivariate Permutation Test for Correlations
   Yoder, Blackford, Waller, and Kim (2003)

Observed Correlations & P-values

         r           p
E1       -0.653      0.015
E2       -0.878      0.000
E3       -0.660      0.014
E4       -0.760      0.003
E5       -0.790      0.002
E6       -0.376      0.128
E7       -0.699      0.008
E8       -0.256      0.224


Results of MPT for Correlations

        Variable     r          p        MPT exact p
1       E2           -0.878     0        0.016
2       E5           -0.79      0.002    0.046
```

References

Westfall, P., & Young, S. (1993). *Resampling-Based Multiple Testing.* New York: Wiley & Sons.

Yoder, P. J., Blackford, J. B., Waller, N. G., & Young, G. (in press). Enhancing Power While Controlling Family-Wise Error: An Illustration of the Issues Using Electrocortical Studies. Journal of Clinical and Experimental Neuropsychology.

Appendix A

R Resources

The R programming language is very similar to S and S-Plus. Any resources for learning S or S-Plus can be easily used for learning R. The R project website has resources available on-line for learning R.

Website:

www.r-project.org

R Manuals:

1. An Introduction To R
2. R Language Definition
3. Writing R Extensions
4. R Data Import/Export
5. R Installation And Administration
6. The R Reference Index

Contributed Documentation:

http://cran.r-project.org/other-docs.html

Textbooks:

1. Venables, W. N. & Ripley, B. D. (2002). Modern applied statistics with S (4th edition).
2. Chambers, J. M. (1998). Programming with data: A guide to the S language. New York: Springer.
3. Spector, P. (1994). An introduction to S and S-Plus. Belmont, California: Duxbury Press.

Appendix B

Sample Data

Sample data are provided so that you can test run this program. The sample data represent electrophysiological signals measured from 9 different electrodes placed on the scalps of 11 subjects. The research question is: Is there a significant relation between electrophysiological signals from one or more of the electrodes and a behavioral measure? The 9 electrode signals (E1-E9) are the predictor variables and the behavioral measure (behavior) is the criterion variable.

Behavior	E1	E2	E3	E4	E5	E6	E7	E8	E9
4	5.58	5.2	5.47	5.29	5.19	5.08	5.27	5.01	5.22
2	5.16	5.02	5.25	5.08	5.22	5.1	5.28	5.28	5.3
3	5.24	4.87	5.12	5.09	5.21	5.2	5.31	5.23	5.24
1	5.26	5.16	5.28	5.24	5.21	5.16	5.39	5.2	5.23
2	5.26	5.17	5.3	5.21	5.16	5.12	5.26	5.17	5.16
1	4.98	5.25	5.01	4.94	5.28	5.2	5.25	4.98	4.91
14	4.47	4.16	4.48	4.38	4.78	4.96	4.77	4.98	4.98
6	5.05	4.93	5.09	4.94	5.13	5.04	5.33	5.31	5.31
3	5.39	5.24	5.62	5.33	5.46	5.42	5.59	5.56	5.71
6	5.09	4.88	5.25	5.15	5.02	5.03	5.25	5.32	5.38
1	4.99	4.94	5.09	5.02	5.13	4.9	5.15	5.24	5.28