Guide to Running TEMAP2.R

The R-program, TEMAP2.R, fits the TE models to the data; it can also construct Monte Carlo simulation of the test statistic, and it can perform bootstrapping to construct distributions of parameter estimates for each model. These features are valuable additions to what the Excel spreadsheet can do, which are especially useful with small samples.

The programming language R is free. See Li and Baron (2011) for an introduction to R with examples of data analysis in psychology. The R program can be downloaded and installed from URL: https://cran.r-project.org/

In addition to the standard installation (current version is 3.4.3), several packages need to be installed: "scales" is used to draw graphs and "boot" for the boot- strapping; "readxl" allows the use of an Excel file as input. To add these to your installed version of R, start R and type the following at the prompt:

- `install.packages("scales")`

You will be asked to select a CRAN mirror site. Choose one near you. Now install "boot" by typing:

- `install.packages("boot")`

Next, install "devtools" and "readxl" as follows:

- `install.packages("devtools")`
- `install.packages("readxl")`

Next, create a folder (i.e., a directory) called TE_results, creating a path (e.g., in Windows) as follows: C:/Users/UserName/Documents/TE_results, which will be the working folder for your input (data) and output (results) of the program. Download the

text file, TEMAP2.txt, from the supplements to this article, and save it as TEMAP2.R in this working folder (which we now call <folder>). Download the Excel file, *Example.xlsx*, and save it in the same folder.

In R, set the folder containing the program to be the working directory, with the appropriate path to your folder:

- setwd("<folder>").

Data Input in Excel File

The Excel file used for input to the program contains three worksheets. The "READ ME" sheet contains information on how to organize the data and information on how to specify the inputs. The "Inputs" sheet contains values that can be adjusted to request that the program analyze TE-4, TE-2, or TE-1, to allow parameters to be free or fixed.

The assumption of response independence assumes that the probability of each of the 16 response patterns is the product of the binary choice probabilities. The statistical test provided is the standard Chi-Square test of this multiplicative independence property. The program provided by Birnbaum (2012) performs tests of other independence properties assumed by random preference choice models. The TE models do not imply these independence properties in either the *i*TET or *g*TET cases (Birnbaum, 2013), except in special cases.

On the "Inputs" worksheet, one can request Monte Carlo simulation and/or Bootstrapping of parameter estimates. Two types of Monte Carlo simulation are performed: The conservative method, which uses the original parameter estimates in each simulated sample to assess fit, and the re-fit method, which re-estimates the parameters

within each new simulated sample before calculating the index of fit. The re-fit method yields better fits of the simulated samples, and therefore leads to smaller estimated p-values; it is thus more likely to reject the null hypothesis that the model fits.

When a model with fixed parameters is specified, one can also request a Monte Carlo simulation of the distribution of difference in fit (between the general model with free parameters and the restricted model with fixed parameters). The difference between the two observed values can then be compared to the simulated distribution of differences to assess statistical significance.

The "participant responses" worksheet contains the data in the form of response frequencies for the 16 possible response patterns. Each line represent a different case, which may represent aggregated data for a group of participants (for *g*TET) or for an individual (*i*TET).

Running the Program

After checking to ensure that the program, TEMAP2.R, and the data file, Example.xlsx, are in the same directory that has been set to the working directory, run the program. As set up in the example file, the program will create 37 new files that contain output from the program.

Three files created by the program contain the original data for each case, parameter estimates of the three TE models (TE-1, TE-2, and TE-4), best-fit predicted values of those data, best-fit Chi-Square index of fit, and the conventional *p*-value for the Chi-Square index. A fourth file contains the same information for the assumption of response independence, which will be violated according to TE models when there is a mixture of true preferences within a group of people or within an individual.

The next three files contain the Monte Carlo simulated p-values for the three TE models by means of the conservative and refit simulation methods. The conservative method uses the best-fit parameters from the original data (which are the population parameters used for the simulations) to measure fit in each new sample, whereas the refit procedure uses the same (simulated) sample data but finds the best-fit parameters in each sample before calculating the index of fit.

The next three files contain results of bootstrapping for the three TE models; it lists the lower and upper values corresponding to the (bootstrapped) 95% confidence intervals on the estimated parameters for the restricted model specified in the data file. Eight pdf files are also created to show the bootstrapped sampling distributions of the parameters of the restricted model in TE-4, and six additional files show the bootstrapped distributions for TE-2.When confidence intervals are desired for the full model as well, one can run the program again with all parameters free, in order to get the bootstrapping results for the unrestricted model(s).

The next three files contain results of Monte Carlo simulations of the distribution of differences in Chi-Square comparing the full model to the restricted model for the three TE models. The files contain the separate Chi-Squares for the full and restricted models, the difference, the conventional p-value, and the Monte Carlo simulated p-value. Figures of the Monte Carlo simulated distributions are also created for the TE-2 and TE-4 models for the first case.

When a restricted model is specified, as in the example file, three files are produced that compare the predictions and index of fit for the full and restricted models.

In addition, a text file containing a summary of the main results of the analysis is generated.

The program is documented by many comments (Text on a line following # are comments in R). The section beginning with line 2320 creates the output of the program. Suggestions are provided in comments for revising this section to access additional information generated by the program and to save it as output. For example, figures for the results of Monte Carlo simulations and bootstrapping could be requested for all cases (instead of just the first case).

The example file has been configured to request only 100 samples to illustrate the program; once the program is running properly, the 100 on the "Inputs" worksheet of *Examples.xlsx* should be changed to 10000 for better accuracy of both the Monte Carlo simulations and the bootstrapping of parameter estimations. The time to run the program may be quite long when the number of samples is large.