User's Guide for R routines to implement assumption adequacy averaging two-group comparison

Stan Pounds
June 10, 2008

Terms of Use

This software may be freely used provided that Pounds and Rai (2008) are cited in all resulting scientific presentations or publications. There is no warranty with this software.

Purpose

This software implements the assumption adequacy averaging (AAA) procedure developed by Pounds and Rai (2008) for the comparison of the median expression of many genes across two experimental groups.

Description of Software

This software consists of routines written for the freely available R statistical computing language (www.r-project.org). It allows the user to implement the procedure as part of an R program or to call the routine with a command-generated graphical user interface (GUI).

Getting Started

You must install R on your computer. Further details on how to install R are available from www.r-project.org. Additionally, you must load the "tcltk" package to use the GUI. Finally, you must download the R code file "AAA-library.R" from www.stjuderesearch.org/depts/biostats and use the source command to make the routines available.

Data Preparation

Data must be in two tab-delimited text files with one row of column headers. The "sample information file" has one row per subject and columns for variables. The sample information file must include a column with sample identifiers and a column with the group labels. The "expression data file" has one row per gene and multiple columns. The first column includes a probe set identifier and subsequent columns have names that must EXACTLY match the sample identifiers in the sample information file.

Example

The website www.stjuderesearch.org/depts/biostats/aaa includes an example R script, example sample information file and example expression data file. You may download the data files and code files to implement the example. The example shows how to implement the method with the command-generated graphical user interface and how to implement the method as part of an R program.

Implementation with Command-Generated GUI

After you have used the source command to make the routines in the library available for use, you may type "gui.aaa.2grp()" at the command prompt to generate a GUI to implement the procedure. The GUI requires the user to specify the name of the sample information file, the names of the sample identifier and group label columns within the sample information file, the expression data file, a results file, an R code file, and whether the expression data should be log-transformed.

First, the user specifies the sample information file by pressing the "Select" button to the right of "Sample Information File." in the top-right corner of the screen. This opens a browser window that allows the user to select the sample information file. After the user selects the file, the file name will appear in the GUI.

Next, the user specifies the sample identifier column of the sample information file by pressing the "Select" button to the right of "Sample ID Column." This will open a window with a list of column names for the user to select from. The user clicks on the column header name and presses "OK." The name of the column header will appear in the GUI after the user has made the selection. Note: Occassionally, windows will pop-up behind other currently open windows. You may need to check the task bar at the bottom to see if a TK window has opend behind other windows.

The user then specifies the group label column of the sample information file in the same manner as the sample identifier column was selected. The expression data file is selected in a manner similar to the selection of the sample information file. The results file and R code file are selected by clicking the respective "Specify" buttons. Finally, the user specifies whether the expression data should be log-transformed by clicking on a radio button.

Now, the user can choose to write the R code to perform the analysis by clicking the "Write R code" button or simply perform the analysis by clicking the "Perform Analysis" button.

Implementation as an R program

The routine "aaa.2grp" can be called as part of an R program. The aaa.2grp routine requires the file path for the sample information file (pheno.file), the name or numeric index of the sample information file with the sample identifiers (sampID.col), the name or numeric index of the group labels in the sample information file (grplbl.col), the name of the expression data file (expr.file), and the name of the result file (result.file) as arguments. The argument “logtrans.expr” is optional with a default of TRUE.

Reference

Pounds S and Rai SN (2008) Assumption Adequacy Averaging as a Concept for Developing More Robust Methods for Differential Gene Ex-pression Analysis. *Computational Statistics and Data Analysis*, to appear.