

RMA

Software for Reduced Major Axis Regression

[HTTP://WWW.BIO.SDSU.EDU/PUB/ANDY/RMA.HTML](http://www.bio.sdsu.edu/pub/andy/rma.html)

Andrew J. Bohonak
San Diego State University

RMA

Software for Reduced Major Axis Regression

RMA is a fast and simple application for reduced major axis regression (RMA). RMA is more appropriate than standard ordinary least squares (OLS) regression when the independent variable x is measured with error (Sokal and Rohlf 1981). In this case, estimates of slope will be biased. McArdle (1988) suggests as a rule of thumb that RMA should be used when the error rate in x exceeds one-third of the error rate in y . (Assuming, of course, that the goal is to estimate the "true relationship" between the variables.) RMA appears to be a more appropriate estimator of slope in population genetic "isolation by distance" analyses than OLS (Hellberg 1994).

Statistical estimators of RMA slope and intercept are outlined below. Error estimates are calculated using three methods:

- 1) standard linear regression approximations
 - 2) jackknifing over cases
 - 3) bootstrapping over cases
-

Operation

Hardware

RMA for Windows requires Windows 95 or later. (I have received one bug report regarding Windows XP.) *RMA* will run on Intel x86, Intel Pentium and compatible processors.

RMA for PPC Macintosh application has modest memory requirements (1-5 MB RAM is usually sufficient) and has been tested using Systems 9.1-9.2.2 on G3 and G4 processors. The program is also expected to run on OS 8.x and OS X under classic mode, although I have not tested this.

Software installation

Installation of *RMA* only requires downloading and expanding the SEA (self expanding archive) available on the *RMA* web site. In Macintosh, default RAM

requirements can be changed manually within the Finder (through the Get Info->Memory menu) if necessary.

Input files

RMA uses a very simple file format. The file can have any name but MUST be in the same folder as the *RMA* application. Each line in the file contains a single case or observation, with the y value first, a tab or space, then the x value. The input file should contain no other information. *RMA* input files can easily be generated by saving a text-only file from a spreadsheet application (e.g., Microsoft Excel). A maximum of 2000 cases (data points) is permitted.

Windows users should note that (depending on system settings), a text file may have a hidden “.txt” suffix.

An example input file named 'RMA.1' is provided. This corresponds to the data set used by Sokal and Rohlf (1981) to illustrate RMA.

Running RMA

After launching the application, enter the # portion of the input file name. Next, enter the number of replicate bootstraps desired (up to 100,000). Following Sokal and Rohlf (1981), the RMA slope b is calculated as:

$$b_{\text{RMA}} = \pm \sqrt{\frac{SSy}{SSx}} = \pm \sqrt{\frac{\sum y^2 - (\sum y)^2 / n}{\sum x^2 - (\sum x)^2 / n}}$$

As with OLS, $a_{\text{RMA}} = \bar{y} - b_{\text{RMA}} \bar{x}$, $r^2_{\text{RMA}} = SS_{xy} / (SS_x)(SS_y)$ and errors are calculated as

$$b_{\text{RMA}}[\text{SE}] = \sqrt{\frac{MSE}{SSx}}$$

$$a_{\text{RMA}}[\text{SE}] = \sqrt{MSE \left(\frac{1}{n} + \frac{\bar{x}^2}{SSx} \right)}$$

Confidence intervals are calculated from these using standard methods (Sokal and Rohlf 1981).

One-delete jackknife estimates of a , b and r^2 and the associated standard errors are calculated following Weir (1990).

Confidence intervals are also calculated by bootstrapping over cases (i.e., creating new data sets with n cases each, by random sampling with replacement). The middle 95% or 99% of the bootstrap replicates constitute the confidence interval.

Programming environment

RMA was written in C and compiled using CodeWarrior v. 8.0 for Macintosh. Source code will be made available upon request.

Published Studies Using IBD

Gonzalez-Solis, J. 2004. Sexual size dimorphism in northern giant petrels: ecological correlates and scaling. *Oikos* **105**: 247-254.

Guillemain, M., H. Fritz, M. Klaassen, A. R. Johnson, and H. Hafner. 2004. Fuelling rates of garganey (*Anas querquedula*) staging in the Camargue, southern France, during spring migration. *Journal of Ornithology* **145**: 152-158.

Hamilton, A. J., N. A. Schellhorn, N. M. Endersby, P. M. Ridland, and S. A. Ward. 2004. A dynamic binomial sequential sampling plan for *Plutella xylostella* (Lepidoptera : Plutellidae) on broccoli and cauliflower in Australia. *Journal of Economic Entomology* **97**: 127-135.

Version History

- v 1.17 3 September 2004 - Additional checks for possible divide by zero errors.
 - New interface and compiler (Windows version).
- v 1.16 24 June 2004 - Minor typecasting fix that may (or may not) have caused crashes or unusual results in very rare instances.
- v 1.14 11 December 2002 - Fixed an overflow error affecting some data sets, particularly the jackknife calculations.
 - Standard linear confidence intervals are now estimated more precisely.
 - Maximum number of cases increased to 2000. The minimum number of bootstraps is lowered to 1.

- v 1.13 18 November 2002 - input file name is no longer constrained
 - If the file is not found, RMA searches for the same file name with “.txt” appended. (In Windows, the extension is often added automatically, and can be hidden from view.)
 - RMA should now correctly parse files with an extra blank line at the end.
- v 1.12 5 June 2002 - fixed rounding error affecting some RMA calculations for very small genetic distances
- v 1.11 23 April 2002 - minor cosmetic fixes, Windows version released
- v 1.1 4 April 2001 - CIs for jackknife removed and replaced with SEs, minor cosmetic changes
- v 1.0 18 March 2001 - initial release

Literature cited

- Hellberg, M. E. 1994. Relationships between inferred levels of gene flow and geographic distance in a philopatric coral, *Balanophyllia elegans*. *Evolution* **48**: 1829-1854.
- McArdle, B. H. 1988. The structural relationship: regression in biology. *Canadian Journal of Zoology* **66**: 2329-2339.
- Sokal, R. R., and F. J. Rohlf. 1981. *Biometry*. 2nd edition. Freeman, NY.
- Weir, B. S. 1990. *Genetic data analysis: methods for discrete population analysis*. Sinauer Associates, Sunderland, MA.
-