Statistical Analysis Programs in *R* for FMRI Data

Gang Chen*, Ziad S. Saad, and Robert W. Cox
Scientific and Statistical Computing Core, National Institute of Mental Health
National Institutes of Health, Department of Health and Human Services, USA
* Contact author: gangchen@mail.nih.gov

Keywords: FMRI; group analysis; mixed-effects meta-analysis; LME; Granger causality.

1. Introduction

Open-source statistical tools are growing and evolving at a fast pace in the *R* community, presenting a valuable opportunity to incorporate frontier methodologies into data analysis in functional magnetic resonance imaging (FMRI). As FMRI data analysis usually involves 3D or 4D (3D+time) datasets in a massively univariate fashion, typical computation is quite intensive, with runtimes ranging from a few seconds to a day or two. With various state-of-art statistical packages in parallel computing using *R* [1] becoming widely available, they can be immediately beneficial to the brain imaging community. Here we present a few FMRI data analysis programs written in *R* during the past couple of years; these are available for download at http://afni.nimh.nih.gov/sscc/gangc.

2. Programs for FMRI data analysis

(1) Mixed-effects meta analysis (MEMA): `3dMEMA`

Data analysis in FMRI is typically carried out in two levels: effect size (linear combination of regression coefficients) and its (within-subject) variance are estimated at the 1st (individual subject) level, and group (population) inferences are performed at a 2nd level. The conventional approach only takes the effect size from the 1st level to the 2nd, assuming cross-subject variance is much larger than within-subject variance, or within-subject variance is uniform in the group [2]. In addition, the data are assumed to follow a normal distribution; thus outliers, when present, are not appropriately handled.

Here we present a computationally efficient frequentist approach [3,4] that incorporates both variance components at the group level. Specifically, we use a REML method to estimate cross-subject heterogeneity in a mixed-effects meta-analysis (MEMA) model and estimate the group effect through weighted least squares. The program handles one-, two-, and paired-sample test types, and covariates are also allowed to control for cross-subject variability. In addition to the group effect estimate and its statistic, the software also provides a cross-subject heterogeneity estimate and a chi-square test for its significance, the percentage of within-subject variability relative to the total variance in the data, and a *Z*-statistic indicating the significance level at which a subject is an outlier at a particular region.

Our approach is computationally economical, and generally more powerful and valid than the conventional method (summary statistics [2]), which ignores the effect size estimate's reliability from individual subjects. It is also relatively robust against outliers in the group data when the Gaussian assumption about the between-subject variability is relaxed by using a Laplace distribution instead [5], whose tails can better model outliers and reduce their impact on group effect and its significance.

(2) Linear mixed-effects (LME) modeling: `3dLME`

Using the **nlme** [6] and **contrast** [7] packages, `3dLME` takes effect size from each subject (at each voxel) as its only input, but allows more flexible designs than `3dMEMA`, such as missing data and unlimited number of categorical variables and covariates. `3dLME` can also handle conditions modeled with multiple time-dependent basis functions, and regression coefficients from all basis functions are analyzed in an LME model with no intercept/constant.

(3) Granger causality analysis: `1dGC` and `3dGC`

These multivariate Granger causality analysis tools use the **vars** package [8], and are suited for (but not limited to) FMRI data. The program identifies patterns of association among brain ROIs, and generates a graphic representation of the identified network. Unlike previous implementations of GC analysis for FMRI [9], this technique preserves information concerning the sign, in addition to the direction, of prior prediction (causality) between ROI time courses from individual subjects all the way to the group level analysis.

(4) Miscellaneous

`3dICA`: independent component (ICA) analysis using **fastICA** [10]
`3dICC`: intraclass correlation analysis for two- and three-way random-effects ANOVA
`3dKS`: Kolmogorov-Smirnov test using `ks.test`
`1dSEMr`: structural equation modeling or path analysis with covariance matrix of multiple ROIs using **sem** [11].

References

[1] R Development Core Team (2009). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.
[2] W D Penny, and A J Holmes (2007). *Random effects analysis*. In: *Statistical Parametric Mapping* (ed. Friston, K. *et al*.). Academic Press.
[3] W Viechtbauer (2005). *Bias and efficiency of meta-analytic variance estimators in the random-effects model*. J Educ Behav Stat, 30, 261-293.
[4] W Viechtbauer (2009). metafor: Meta-Analysis Package for R. R package version 0.5-7. http://CRAN.R-project.org/package=metafor
[5] E Demidenko (2004), *Mixed Models: Theory and Applications*. Wiley-Interscience.
[6] J Pinheiro, D Bates, S DebRoy, D Sarkar and the R Core team (2009). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-96.
[7] Max Kuhn, contributions from Steve Weston, Jed Wing and James Forester (2009). *contrast: A collection of contrast methods*. R package version 0.12.
[8] Bernhard Pfaff (2008). *VAR, SVAR and SVEC Models: Implementation Within R Package vars*. Journal of Statistical Software 27(4).
[9] A Roebroeck, E Formisano, and R Goebel (2005). *Mapping directed influence over the brain using granger causality mapping*. NeuroImage, 25:230–242.
[10] J L Marchini C Heaton, B D Ripley (2009). *fastICA: FastICA Algorithms to perform ICA and Projection Pursuit*. R package version 1.1-11.
[11] John Fox with contributions from Adam Kramer and Michael Friendly (2009). *sem: Structural Equation Models*. R package version 0.9-19.