No Limits: AN(C)OVA for Group Analysis with a Large Number of Factors and Covariates

Gang Chen[1], Ziad Saad[1], Johanna Jarcho[2], Daniel Pine[2], Robert Cox[1]

[1]Scientific and Statistical Computing Core

[2]Section on Development and Affective Neuroscience

NIMH & NINDS / NIH / DHHS / Bethesda MD USA

Contact: gangchen@mail.nih.gov http://afni.nimh.nih.gov

Traditional FMRI Group Analysis Approaches

✧ General Linear Model (GLM)

- Take effect estimates (β's) only from individual subject analyses
- Assume that effect estimates have same precision across subjects
- Models in widespread use
  - Student *t*-tests: one-, two-sample, and paired
  - Simple AN(C)OVA
- **Programs in AFNI**: 3dttest++, 3dANOVA2/3, GroupAna
- **Limitations**
  - Limited number of explanatory variables (factors and covariates)
  - Rigid data structure: no missing data allowed
  - Rigid model: all possible effects included even if not significant
  - Rigid assumptions: sphericity; generic covariance structure for random effects and residuals unavailable

✧ Linear Mixed-Effects (LME) Model

- Take effect estimates (β's) only from individual subject analyses
- Assume effect estimates have same precision across subjects
- ★ Flexibility of specifying covariance structures and heterogeneity
- ★ No limit on number of explanatory variables
- ★ Missing data allowed
- **Program in AFNI**: 3dLME
- **Limitations**: difficult in assigning degrees of freedom

✧ Mixed-Effects Multi-level Analysis (MEMA)

- ★ Consider precision information from individual level
- ★ More accurate group effect estimate and significance test
- **Program in AFNI**: 3dMEMA
- **Limitations**: only able to analyze paired, one-, and two-sample types

Multivariate Modeling (MVM) Paradigm

✧ Repeated Measures (RM) Designs

- Within-subject or RM factor
  - Categorical variable: *e.g.*, 3 levels of emotion (positive, negative, neutral)
  - Effect estimates available at all levels from every subject
  - Traditionally modeled as an explanatory variable (factor)
  - One-way within-subject or RM ANOVA
- RM designs are popular in psychology and FMRI experiments
  - Typically formulated as ANOVA: *e.g.*, one-way within-subject or RM ANOVA

$$y_{ij} = \mu_j + b_i + \epsilon_{ij}$$

where y_{ij} is the response at the jth factor level for the ith subject, μ_j is the effect at the jth factor level, b_i is the deviation of the ith subject at the jth factor level, and ϵ_{ij} is the residual of the ith subject at the jth factor level, $i = 1, ..., n$, $j = 1, ..., m$. The distributional assumptions are $b_i \sim N(0, \tau^2)$ and $\epsilon_{ij} \sim N(0, \sigma^2)$, where τ^2 and σ^2 are between- and within-subject variances respectively.

  - Efficient when sphericity holds, but sensitive to sphericity violation: inflated significance
  - May involve between-subjects (or subject-grouping) factors

✧ Multivariate modeling (MVM) approach

- All levels (or their contrasts) of RM factor: multiple response variables
  - One-way RM ANOVA transforms to a special MANOVA

$$\mathbf{y}_i = \mathbf{u} + \mathbf{e}_i$$

where $\mathbf{y}_i = (y_{i1}, ..., y_{im})^T, \mathbf{u} = (\mu_{i1}, ..., \mu_{im})^T$, and $\mathbf{e}_i = (\epsilon_{i1}, ..., \epsilon_{im})^T$, $i = 1, ..., n$. The distributional assumption is $\mathbf{e}_i \sim N_m(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is a positive definite and symmetric matrix that is constant across subjects, and $\mathbf{e}_i$'s are independent from each other. Unlike ANOVA, the covariance matrix $\mathbf{\Sigma}$ is estimated from the data instead of being assumed spherical.

  - Traditional AN(C)OVA becomes MAN(C)OVA: only between-subjects factors and quantitative variables are treated as explanatory variables
  - Immune to sphericity violation, but may lose power when sphericity holds

Implementation in AFNI: 3dMVM

- Program written in R [1] with package afex [2] for MVM
- Post-hoc tests performed through symbolic coding with labels in R package phia [3]
- Currently implemented as shell scripting with parallel computing capability through package snow [4]

Assessment of MVM Approach

✧ Advantages of MVM: Flexibility in Modeling

- Multivariate aspect: when BOLD response shape is modeled with multiple basis functions, the shape integrity can be maintained at group level using MVM without presuming the covariance structure across the multiple effects
  - Alternative approach: LME with AR covariance matrix for residuals – 3dLME
- Univariate aspect: MVM platform allows for GLM capability
  - Huynh-Feldt and Greenhouse-Geisser corrections available when sphericity is violated
- ★ Allows for unbalanced designs (unequal subjects across groups)
- ★ Explanatory variables can be categorical or quantitative
- ★ No limit on the number of explanatory variables (factors and covariates)
- ★ Post-hoc tests through symbolic coding with labels instead of direct dummy coding

✧ Disadvantages of MVM

- Under some circumstances bigger sample size (subjects) required
- Computationally heavy
- Unable to deal with missing data
  - Solution: 3dLME
- Incapable of handling within-subject quantitative variables
  - Solution: 3dLME

Sample Application

- 6 categorical variables (factors) in FMRI experiment: **6-way ANOVA**
  - 3 between-subjects factors: **age** (adult, child), **diagnosis** (healthy, anxious), and **scanner** (scanners 1 and 2)
  - 3 within-subject factors: subject's **interest in a target** (interest, no interest), target's **interest in a subject** (interest, no interest), **accuracy** (accurate, inaccurate)
- 89 subjects: healthy children (24) and adults (32), anxious children (15) and adults (15)
  - **Unbalanced** design: unequal number of subjects across groups
- 259 post-hoc tests in addition to *F*-stat for main effects and interactions

Acknowledgements

The research was supported by the NIMH Intramural Research Program of the NIH.

References

[1] R Core Team (2013). 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] Henrik Singmann (2013). afex: Analysis of Factorial Experiments. R package version 0.4-57. http://CRAN.R-project.org/package=afex

[3] Helios De Rosario-Martinez (2012). phia: Post-Hoc Interaction Analysis. R package version 0.1-0. http://CRAN.R-project.org/package=phia

[4] Luke Tierney, A. J. Rossini, Na Li and H. Sevcikova (2013). Snow: Simple Network of Workstations. R package version 0.3-12. http://CRAN.R-project.org/package=snow