Variable type

A Comprehensive and Flexible Approach to AN(C)OVA at the Group Level

Gang Chen, Ziad Saad, Robert Cox
Scientific and Statistical Computing Core
NIMH & NINDS / NIH / DHHS / Bethesda MD USA
Contact: gangchen@mail.nih.gov http://afni.nimh.nih.gov

Demo Example: 2 × 3 Mixed ANCOVA

✧ Explanatory variables

- Factor A (Group): 2 levels (patient and control)
- Factor B (Condition): 3 levels (pos, neg, neu)
- Factor S (Subject): 15 ASD children and 15 healthy controls
- Covariate: Age

✧ Piecemeal *t*-tests

- Group comparison + age effect
- Pairwise comparisons among three conditions
- Effects that cannot be analyzed
  - Main effect of Condition
  - Interaction between Group and Condition
  - Age effect across three conditions

✧ ANOVA through sums of squares

- Age cannot be modeled

$$F_{(a-1,a(n-1))}(A)=\frac{MSA}{MSS(A)},$$
$$F_{(b-1,a(b-1)(n-1))}(B)=\frac{MSB}{MSE},$$
$$F_{((a-1)(b-1),a(b-1)(n-1))}(AB)=\frac{MSAB}{MSE},$$

where

$$MSA=\frac{SSA}{a-1}=\frac{1}{a-1}(\frac{1}{bn}\sum_{j=1}^{a}Y_{.j.}^2-\frac{1}{abn}Y_{...}^2),$$

$$MSB=\frac{SSB}{b-1}=\frac{1}{b-1}(\frac{1}{an}\sum_{k=1}^{b}Y_{..k}^2-\frac{1}{abn}Y_{...}^2),$$

$$MSAB=\frac{SSAB}{(a-1)(b-1)}=\frac{1}{(a-1)(b-1)}(\frac{1}{n}\sum_{j=1}^{a}\sum_{k=1}^{b}Y_{.jk}-\frac{1}{bn}\sum_{j=1}^{a}Y_{.j.}^2-\frac{1}{an}\sum_{k=1}^{b}Y_{..k}^2+\frac{1}{abn}Y_{...}^2),$$

$$MSS(A)=\frac{SSS(A)}{a(n-1)}=\frac{1}{a(n-1)}(\frac{1}{b}\sum_{i=1}^{n}\sum_{j=1}^{a}Y_{ij.}^2-\frac{1}{bn}\sum_{j=1}^{a}Y_{.j.}^2),$$

$$MSE=\frac{1}{a(b-1)(n-1)}(\sum_{i=1}^{n}\sum_{j=1}^{a}\sum_{k=1}^{b}Y_{ijk}^2-\frac{1}{n}\sum_{j=1}^{a}\sum_{k=1}^{b}Y_{.jk}-\frac{1}{b}\sum_{i=1}^{n}\sum_{j=1}^{a}Y_{ij.}^2+\frac{1}{bn}\sum_{j=1}^{a}Y_{.j.}^2+\frac{1}{abn}Y_{...}^2)$$

✧ General linear model (GLM)

$$\begin{array}{c} \text{Subj} \\ 1\\1\\1\\2\\2\\2\\3\\3\\3\\4\\4\\4\\5\\5\\5\\6\\6\\6 \end{array}
\begin{pmatrix}\beta_{11}\\\beta_{12}\\\beta_{13}\\\beta_{21}\\\beta_{22}\\\beta_{23}\\\beta_{31}\\\beta_{32}\\\beta_{33}\\\beta_{41}\\\beta_{42}\\\beta_{43}\\\beta_{51}\\\beta_{52}\\\beta_{53}\\\beta_{61}\\\beta_{62}\\\beta_{63}\end{pmatrix}
=
\begin{array}{c}
\begin{array}{cccccccccc} X_0 & X_1 & X_2 & X_3 & X_4 & X_5 & X_6 & X_7 & X_8 & X_9 \end{array}\\
\begin{pmatrix}
1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0\\
1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0\\
1 & 1 & -1 & -1 & -1 & -1 & 1 & 0 & 0 & 0\\
1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0\\
1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0\\
1 & 1 & -1 & -1 & -1 & -1 & 0 & 1 & 0 & 0\\
1 & 1 & 1 & 0 & 1 & 0 & -1 & -1 & 0 & 0\\
1 & 1 & 0 & 1 & 0 & 1 & -1 & -1 & 0 & 0\\
1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & 0 & 0\\
1 & -1 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0\\
1 & -1 & 0 & 1 & 0 & -1 & 0 & 0 & 1 & 0\\
1 & -1 & -1 & -1 & 1 & 1 & 0 & 0 & 1 & 0\\
1 & -1 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 1\\
1 & -1 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 1\\
1 & -1 & -1 & -1 & 1 & 1 & 0 & 0 & 0 & 1\\
1 & -1 & 1 & 0 & -1 & 0 & 0 & 0 & -1 & -1\\
1 & -1 & 0 & 1 & 0 & -1 & 0 & 0 & -1 & -1\\
1 & -1 & -1 & -1 & 1 & 1 & 0 & 0 & -1 & -1
\end{pmatrix}
\end{array}
\begin{pmatrix}\alpha_0\\\alpha_1\\\alpha_2\\\alpha_3\\\alpha_4\\\alpha_5\\\alpha_6\\\alpha_7\\\alpha_8\\\alpha_9\end{pmatrix}
+
\begin{pmatrix}\delta_{11}\\\delta_{12}\\\delta_{13}\\\delta_{21}\\\delta_{22}\\\delta_{23}\\\delta_{31}\\\delta_{32}\\\delta_{33}\\\delta_{41}\\\delta_{42}\\\delta_{43}\\\delta_{51}\\\delta_{52}\\\delta_{53}\\\delta_{61}\\\delta_{62}\\\delta_{63}\end{pmatrix}$$

- Challenging to properly code columns and specify weights for effect testing
- Tedious formulations of statistics (common mistakes in implementations!)
- Difficulty to generalize to handle any number of variables
- Covariates cannot be modeled in the presence of within-subject factors
- No way to correct for sphericity violation or unrealistic assumption (same variance-covariance structure)

✧ Multivariate linear model (MVM)

- Within-subject levels coded as columns
- Each subject associated with one row, but not coded as a column in model matrix

$$\begin{array}{c} Subj\\1\\2\\3\\4\\5\\6\end{array}
\begin{array}{c}
\begin{array}{ccc} Pos & Neg & Neu \end{array}\\
\begin{pmatrix}
\beta_{11} & \beta_{12} & \beta_{13}\\
\beta_{21} & \beta_{22} & \beta_{23}\\
\beta_{31} & \beta_{32} & \beta_{33}\\
\beta_{41} & \beta_{42} & \beta_{43}\\
\beta_{51} & \beta_{52} & \beta_{53}\\
\beta_{61} & \beta_{62} & \beta_{63}
\end{pmatrix}\end{array}
=
\begin{array}{c}
\begin{array}{ccc} Int & Grp & Age \end{array}\\
\begin{pmatrix}
1 & 1 & -6\\
1 & 1 & 10\\
1 & 1 & 4\\
1 & -1 & -4\\
1 & -1 & -1\\
1 & -1 & -3
\end{pmatrix}\end{array}
\begin{array}{c}
\begin{array}{ccc} Pos & Neg & Neu \end{array}\\
\begin{pmatrix}
\alpha_{01} & \alpha_{02} & \alpha_{03}\\
\alpha_{11} & \alpha_{12} & \alpha_{13}\\
\alpha_{21} & \alpha_{22} & \alpha_{23}
\end{pmatrix}\end{array}
+
\begin{array}{c}
\begin{array}{ccc} Pos & Neg & Neu \end{array}\\
\begin{pmatrix}
\delta_{11} & \delta_{12} & \delta_{13}\\
\delta_{21} & \delta_{22} & \delta_{23}\\
\delta_{31} & \delta_{32} & \delta_{33}\\
\delta_{41} & \delta_{42} & \delta_{43}\\
\delta_{51} & \delta_{52} & \delta_{53}\\
\delta_{61} & \delta_{62} & \delta_{63}
\end{pmatrix}\end{array}
\begin{array}{c} Subj\\1\\2\\3\\4\\5\\6\end{array}$$

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]

Variable type

```
3dMVM  -prefix     OutputFile    -jobs 8    -SC                                   \
       -bsVars     'Grp*Age'     -wsVars    'Cond'          -qVars 'Age'          \
       -num_glt 4                                                                 \
       -gltLabel 1 Pat_Pos       -gltCode 1                 'Grp : 1*Pat Cond : 1*Pos'  \
       -gltLabel 2 Ctl_Pos-Neg   -gltCode 2            'Grp : 1*Ctl Cond : 1*Pos -1*Neg' \
       -gltLabel 3 GrpD_Pos-Neg  -gltCode 3   'Grp : 1*Ctl -1*Pat Cond : 1*Pos -1*Neg' \
       -gltLabel 4 Pat_Age       -gltCode 4                       'Grp : 1*Pat Age :'  \
       -dataTabel                                                                 \
Subj   Grp         Age           Cond       InputFile                             \
S1     Ctl         23            Pos        S1_Pos.nii                            \
S1     Ctl         23            Neg        S1_Neg.nii                            \
S1     Ctl         23            Neu        S1_Neu.nii                            \
...
S50    Pat         19            Pos        S50_Pos.nii                           \
S50    Pat         19            Neg        S50_Neg.nii                           \
S50    Pat         19            Neu        S50_Neu.nii
```

Post hoc tests

Data layout

Advantages of MVM

✧ Flexibility in Modeling

- Easy formulation of testing statistics
- No limit on the number of explanatory variables
- Covariates modeled even in the presence of within-subject factors
- Voxel-wise covariate (e.g., SFNR) allowed
- Voxel-wise sphericity correction for univariate testing (UVT)
- Within-subject multivariate testing (MVT) as complementary testing
- MVT: HDR modeled with multiple basis functions

✧ The user only provides information

- Explanatory variable type: between- or within-subject
- Centering options for quantitative covariates
- Post hoc tests through symbolic coding
- Data table listing variables and input files

✧ The user does not need to be involved in specifying

- Dummy coding
- Regressors,
- Design matrix
- Post hoc tests through regressors

Paper

Chen et al., Applications of Multivariate Modeling to Neuroimaging Group Analysis: A Comprehensive Alternative to Univariate General Linear Model, NeuroImage (In Press): 10.1016/j.neuroimage.2014.06.027

Acknowledgements

The research was supported by the NIMH & NINDS Intramural Research Programs 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