

Group Analysis: Hands-On

Gang Chen

SSCC/NIMH/NIH/HHS



Make sure you have the files!

- Under directory `group_analysis_hands_on/`
 - Slides: **GroupAna_HO.pdf**
 - Data: **AFNI_data6/GroupAna_cases/**
 - In case you don't have the data
`wget http://afni.nimh.nih.gov/pub/dist/edu/data/AFNI_data6.tgz`
- Require R installation
 - Google R, and then download proper binaries
 - Install a few R packages: `install.packages('afex')`
 - `afex`
 - `phia`
 - `snow`
 - `nlme`
 - `lme4`
 - `contrast`

Preview: choosing programs

- Program list
 - 3dttest++, 3dMEMA, 3dANOVAx, 3dMVM, 3dLME
 - 3ttest, 3dRegAna, GroupAna almost retired
 - Voxel-wise approach
 - ROI analysis **not** discussed: R, Matlab, Excel, SAS, SPSS
 - **uber_ttest.py**: for 3ttest++ and 3dMEMA only
 - Other programs: scripting (too hard? Rick Reynolds!)
gen_group_command.py
 - Typical mistakes
 - Extra spaces after the continuation character **BACKSLASHES** (\)
 - **file_tool -test -infile ...**
 - Typos
 - Model specifications, misuses of options, ...

Preview: choosing programs

- Data layout should not always be the only focus
 - Experiment design: number of explanatory variables (factors and quantitative variables), levels of a categorical variable
 - Balance: equal number of subjects across groups?
 - Missing data: throw out those subjects, or keep the partial data?
 - List all the tests you would like to get out of the group analysis
- If computation cost is of concern
 - Super fast programs: 3dttest++, 3dANOVAx, 3dttest, 3dRegAna
 - Super slow programs: 3dMEMA, 3dMVM, 3dLME, GroupAna
- Special features of 3dMEMA
 - Weights subjects based on reliability
 - Models and identifies outliers at voxel level
 - Handles missing data at voxel level (*e.g.* ECoG data)
 - Cross-subjects variability measures (τ^2 , H, I^2 , ICC) and group comparisons in τ^2

Road Map: Choosing a program?

- ✧ Starting with HDR estimated via shape-fixed method (SFM)
 - One β per condition per subject
 - It could be significantly underpowered
- ✧ Two perspectives
 - Data structure
 - Ultimate goal: list **all** the tests you want to perform
 - Possible to avoid a big model
 - Use a piecemeal approach with 3dttest++ or 3dMEMA
- ✧ Most analyses can be done with 3dMVM and 3dLME
 - Computationally inefficient
 - Last resort: not recommended if alternatives available

Road Map: Student's t -tests

✧ `3dttest++` and `3dMEMA`

✧ Not for F -tests except for ones with 1 DF for numerator

- All factors are of two levels, e.g., 2×2 , or $2 \times 2 \times 2$

✧ Scenarios

- One-, two-sample, paired
- Multiple regression: one group + one or more quantitative variables
- ANCOVA: two groups + one or more quantitative variables
- ANOVA through dummy coding: all factors (between- or within-subject) are of **two** levels
- AN(C)OVA: multiple between-subjects factors + one or more quantitative variables
- One group against a whole brain constant: `3dttest -base1 C`
- One group against a voxel-wise constant: `3dttest -base1_dset`

Road Map: Between-subjects ANOVA

✧ One-way between-subjects ANOVA

- 3dANOVA
- Two groups: 3dttest++, 3dMEMA (OK with > 2 groups too)

✧ Two-way between-subjects ANOVA

- 3dANOVA2 -type 1
- 2 x 2 design: 3dttest++, 3dMEMA (OK with > 2 groups too)

✧ Three-way between-subjects ANOVA

- 3dANOVA3 -type 1
- 2 x 2 design: 3dttest++, 3dMEMA (OK with > 2 groups too)

✧ *N*-way between-subjects ANOVA

- 3dMVM

Road Map: With-subject ANOVA

✧ One-way within-subject ANOVA

- 3dANOVA2 -type 3
- Two conditions: 3dttest++, 3dMEMA

✧ Two-way within-subject ANOVA

- 3dANOVA3 -type 4
- 2 x 2 design: 3dttest++, 3dMEMA

✧ *N*-way within-subject ANOVA

- 3dMVM

Road Map: Mixed-type ANOVA and others

- ✧ One between- and one within-subject factor
 - 3dANOVA3 -type 5 (requiring **equal** # subjects across groups)
 - 3dMVM (especially unequal # subjects across groups)
 - 2 x 2 design: 3dttest++, 3dMEMA
- ✧ Other scenarios
 - Multi-way ANOVA: 3dMVM
 - Multi-way ANCOVA (between-subjects covariates only): 3dMVM
 - HDR estimated with multiple basis functions: 3dMVM
 - Missing data: 3dLME
 - Within-subject covariates: 3dLME
 - Subjects genetically related: 3dLME
 - Trend analysis: 3dLME

Preview: learning by 7 examples

- BOLD responses estimated with one basis function
 - 1 group, 3 conditions with missing data
 - 3 groups, 1 numeric variable (between-subjects)
 - ANOVA
 - ANCOVA
 - Within-subject covariate
- BOLD responses estimated with multiple basis functions
 - 1 group
 - 2 groups

Case 1: three conditions

- Run command line
 - `tcsch -x LME.txt`
 - `tcsch -x LMEtable.txt`
- MEG data
 - **3 conditions**: Baseline, Ket, Placebo
 - 17 subject with missing data: 11 with full data
- Analysis approaches
 - One-way within-subject **ANOVA**
 - Worst: wasting 6 subjects
 - 3 pairwise comparisons with **t-test**
 - Better: partially wasting
 - **LME**
 - Best: all data fully utilized
 - Overall F-stat plus 3 pairwise contrasts

Subj	Baseline	Ket	Placebo
S101	1	1	0
S102	1	1	1
S105	1	1	1
S107	1	1	1
S108	1	1	1
S109	1	1	1
S110	1	1	1
S111	1	1	0
S112	0	1	1
S113	1	1	1
S115	0	1	1
S116	1	1	0
S118	1	1	1
S120	1	1	1
S121	1	1	0
S122	1	1	1
S123	1	1	1

Case 1: three conditions

- Put the data table in a separate text file
 - Unix issue (“Arg list too long): the whole command line beyond the system allows
 - Same dataset can be used for different models
 - Not all columns have to be used
- Navigate the output dataset

Case 2: three groups

- Data information
 - COMT (catechol-*O*-methyl transferase) gene with a Val/Met (valine-to-methionine) polymorphism for schizophrenia
 - 3 genotypic groups: Val/Val (12), Val/Met (10), Met/Met (9)
 - 1 effect estimate from each subject
- What program?
 - Almost everybody immediately jumps to this question!
- Tests of interest?
 - Individual group effects: A, B, and C
 - Pairwise group comparisons: A-B, A-C, and B-C: Two-sample *t*-test
 - Any difference across all three groups? Omnibus *F*-test
- What program?
 - One- or two-sample *t*-test: 3dtttest++, 3dMEMA
 - One-way between-subjects ANOVA: 3dANOVA, 3dMVM

Case 2: three groups

- One-way between-subjects ANOVA
 - Each subject has only one response value!
 - GLM, not really a random-effects model:

$$\hat{\beta}_{i(j)} = \alpha_0 + \alpha_1 * x_{1i(j)} + \alpha_2 * x_{2i(j)} + \epsilon_{i(j)}$$

- Coding for subjects: with one group (A) as base (reference) for dummy coding (0s and 1s), $\alpha_0 = A$, $\alpha_1 = B - A$, and $\alpha_2 = C - A$.
- 3dANOVA
 - Don't directly solve GLM
 - Compute sums of squares: computationally efficient!
- Alternatives: 3dtttest++, 3dMEMA

Case 3: multi-way ANOVA

- Data information
 - 1 subject-grouping variable (Group): young (15) and older (14)
 - 3 within-subject factors:
 - task - 2 levels: Perception and Production
 - Syllable - 2 levels: Simple and Complex
 - Sequence - 2 levels: Simple and Complex
- Tests of interest?
 - Comparisons under various combinations
 - Interactions among the 4 factors
- What program?
 - 3dttest++, 3dMEMA, 3dMVM

Case 4: Within-subject covariate

- Data information
 - 1 within-subject variable: Condition (2 levels: house, face)
 - 1 quantitative (within-subjects) variable: RT (mean RT not significantly different across conditions)
- Tests of interest?
 - Main effects, interactions, various contrasts
- Model
- What program? 3dLME

$$\hat{\beta}_{ij} = \alpha_1 * x_{1j} + \dots + \alpha_k * x_{kj} + \delta_i + \epsilon_{ij}$$

Case 5: one group with multiple basis functions

- Data information
 - 15 subjects
 - One effect of interest modeled with 8 basis (TENT) functions
- Tests of interest?
 - Any overall response at a voxel (brain region)?
- Model $\hat{\beta}_{ij} = \alpha_1 * x_{1j} + \dots + \alpha_k * x_{kj} + \delta_i + \epsilon_{ij}$
 - No intercept
 - Test of interest: $\alpha_1 = \dots = \alpha_k = 0$
 - Residuals ϵ_{ij} are most likely serially correlated
- What program? 3dLME

Case 6: two groups with multiple basis functions

- Data information
 - 15 subjects
 - One effect of interest modeled with 8 basis (TENT) functions
- Tests of interest?
 - Any overall response at a voxel (brain region)?
- Model
 - No intercept
 - Test of interest:
 - Residuals ε_{ij} are most likely serially correlated
- What program? 3dANOVA3 –type 5, 3dMVM

Case 7: ANCOVA

- Data information
 - 2 subject-grouping variables
 - Group (2 levels): control () and ssd ()
 - Gender (2 levels): males () and females ()
 - 1 within-subject variable: Condition (4 levels: visWord, visPSW, visCStr, audWord, audPSW)
 - 1 quantitative (between-subjects) variable: Age (mean age not significantly different across groups)
- Tests of interest?
 - Main effects, interactions, various contrasts
- Model $\hat{\beta}_{ij} = \alpha_1 * x_{1j} + \dots + \alpha_k * x_{kj} + \delta_i + \epsilon_{ij}$
- What program? **3dMVM**, 3dLME

Overview: learning by 11 examples

- BOLD responses estimated with one basis function
 - 3 groups
 - 2 conditions
 - 2 conditions with missing data
 - 3 groups + 2 genders
 - 3 groups + 2 conditions
 - 3 groups + 2 genders + 1 numeric variable (between-subjects)
 - 3 groups + 2 conditions + 1 numeric variable (between-subjects)
 - 3 groups + 2 conditions + 2 numeric variables (1 within-subject and 1 between-subjects)
- BOLD responses estimated with multiple basis functions
 - 1 group
 - 2 groups
 - 2 groups + 2 conditions