Group Analysis

File: **afni24_GroupAna.pdf**

Gang Chen

SSCC/NIMH/NIH

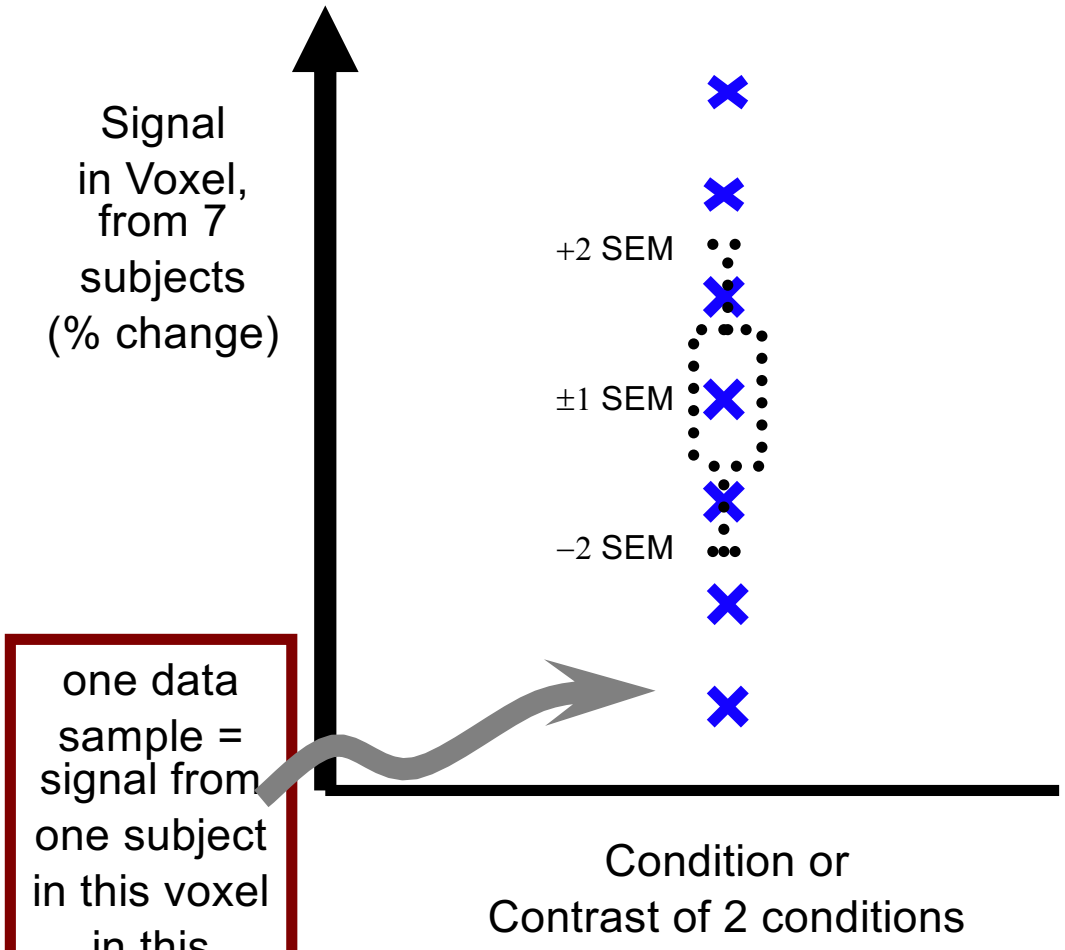
Preview

- Introduction: basic concepts and terminology
  - Background: why do we need to do group analysis?
  - Factor, quantitative covariates, main effect, interaction, …
- Group analysis approaches
  - *t*-test: 3dttest++ (3dttest), 3dMEMA
  - Regression: 3dttest++, 3dMEMA, 3RegAna
  - ANOVA: 3dANOVAx, 3dMVM, GroupAna
  - ANCOVA or GLM: 3dttest++, 3dMEMA, 3dMVM, 3dLME
  - Impact & consequence of FSM, ASM, and ESM
- Miscellaneous
  - Issues with covariates
  - Intra-Class Correlation (ICC)
  - Inter-Subject Correlation (ISC)
- Efficient modeling through information pooling

Why Group Analysis?

- Evolution of FMRI studies
  - Early days [1992-1994]: no need for group analysis
    - Seed-based correlation for one subject was revolutionary
  - Now: torture brain/data enough, and hope nature will confess!
    - Many ways to manipulate the brain (and data)
- Reproducibility and generalization
  - **Science strives for generality**: summarizing subject results
  - Typically 10 or more subjects per group
  - Exceptions: pre-surgical planning, lie detection, …
- Why not one analysis with a giant model for all subjects?
  - Computationally unmanageable and very hard to set up
  - Heterogeneity in data or experiment design across subjects
  - Model and data quality check at individual subject level

Simplest Group Analysis: One-Sample *t*-Test



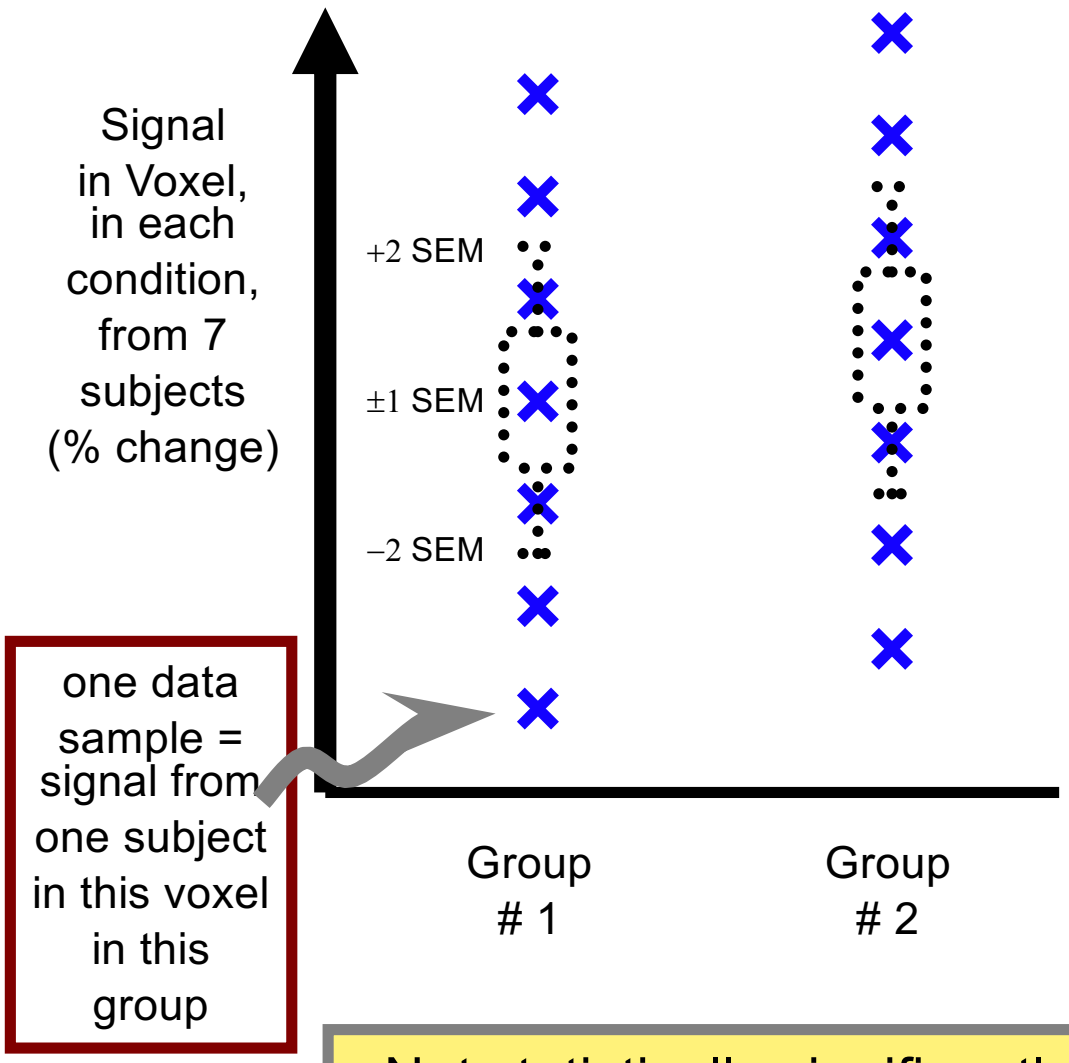
one data sample = signal from one subject in this voxel in this condition

- SEM = Standard Error of the Mean
  = standard deviation of sample, divided by square root of number of samples
  = estimate of uncertainty in sample mean

- One-sample *t*-test determines if sample mean is large enough relative to SEM

- statistically significantly different from 0!

Simplest Group Analysis: Two-Sample *t*-Test



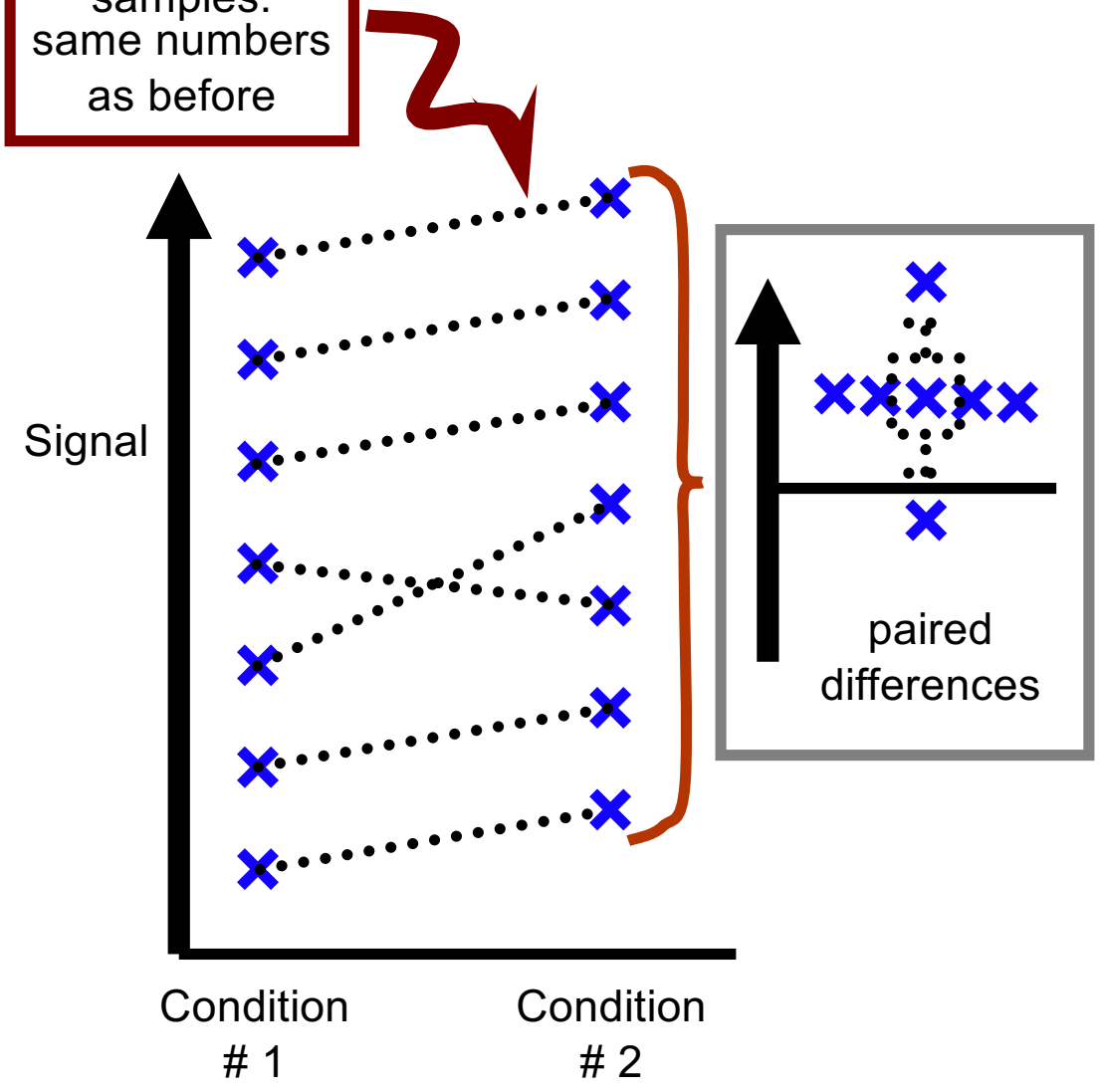
one data sample = signal from one subject in this voxel in this group

- Group = some way to categorize subjects (*e.g.*, sex, drug treatment, disease, …)
- SEM = Standard Error of the Mean = standard deviation of sample divided by square root of number of samples = estimate of uncertainty in sample mean
- Two-sample *t*-test determines if sample means are "far apart" compared to size of SEM

- Not statistically significantly different!

Simplest Group Analysis: Paired (~1-sample) *t*-Test

paired data samples: same numbers as before



- Significantly different!
- Condition #2 > #1, per subject

- Paired means that samples in different conditions should be linked together (*e.g.*, from same subjects)
- Test determines if differences between conditions in each pair are "large" compared to SEM of the differences
- Paired test can detect systematic *intra*-subject differences that can be hidden in *inter*-subject variations
- Lesson: properly separating ***inter***-subject and ***intra***-subject signal variations can be very important!
- **Essentially equivalent to one-sample *t*-test**

Toy example: one group with 7 subjects

- Responses from a group of subjects under one condition
  - data: (β_1, β_2, …, β_7)=(1.13, 0.87, …, 0.72) [% signal change]
- Centroid: average ($\beta_1+\beta_2+\ldots+\beta_7$)/7 = 0.92 is not enough
  - Variation/reliability measure: diversity, spread, deviation
  - How different is 0.92 from 0 compared to its deviation?
- Model building
  - Subject *i*'s response = group average + deviation of subject *i*: simple model GLM (one-sample *t*-test)

$$\hat{\beta}_i = b + \epsilon_i, \epsilon_i \sim N(0, \sigma^2)$$

  - If individual responses are consistent, ϵ_i should be small
  - How small (*p*-value)?
    - *t*-test: significance measure = $\hat{b}/(\hat{\sigma}/n)$
- 2 measures: ***b*** (dimensional) and ***t*** (dimensionless)

Caveats

- Results: two components (in afni GUI: OLay + Thr)
  - Effect estimates: have unit and physical meaning
  - Statistical significance (response to house **statistically** > face)
    - Popular but misleading: *p*-values solely focused in FMRI!
- Statistical significance (*p*-value) becomes obsession
  - Published papers: Big and tall parents (violent men, engineers) have more sons, beautiful parents (nurses) have more daughters
  - Statistical significance is not the same as practical importance
- **Fallacy**: **binarized thinking**
  - **Brain regions do NOT behave in a discrete fashion!!!**
  - Statistically insignificant effect might be real
    - Sample size, suboptimal model, poor alignment across subjects
  - Statistically significant effect might be false
  - Difference between the two is not necessarily significant

Caveats

- Conventional: voxel-wise (brain) or node-wise (surface)
  - Prerequisite: reasonable alignment to some template
  - Limitations: alignment could be suboptimal or even poor
    - Different folding patterns across subjects: better alignment could help (perhaps to 5 mm accuracy?)
    - Different cytoarchitectonic (or functional) locations across subjects: structural alignment of images won't help!
    - Impact on conjunction vs. selectivity
- Alternative: **ROI-based approach**
  - Predefine a list of regions
  - Easier: brain reduced to less than hundreds of values per subject
  - Model building and tuning
  - **No more multiple comparisons and *p*-value**
  - More robust and valid results reporting

Why big models?

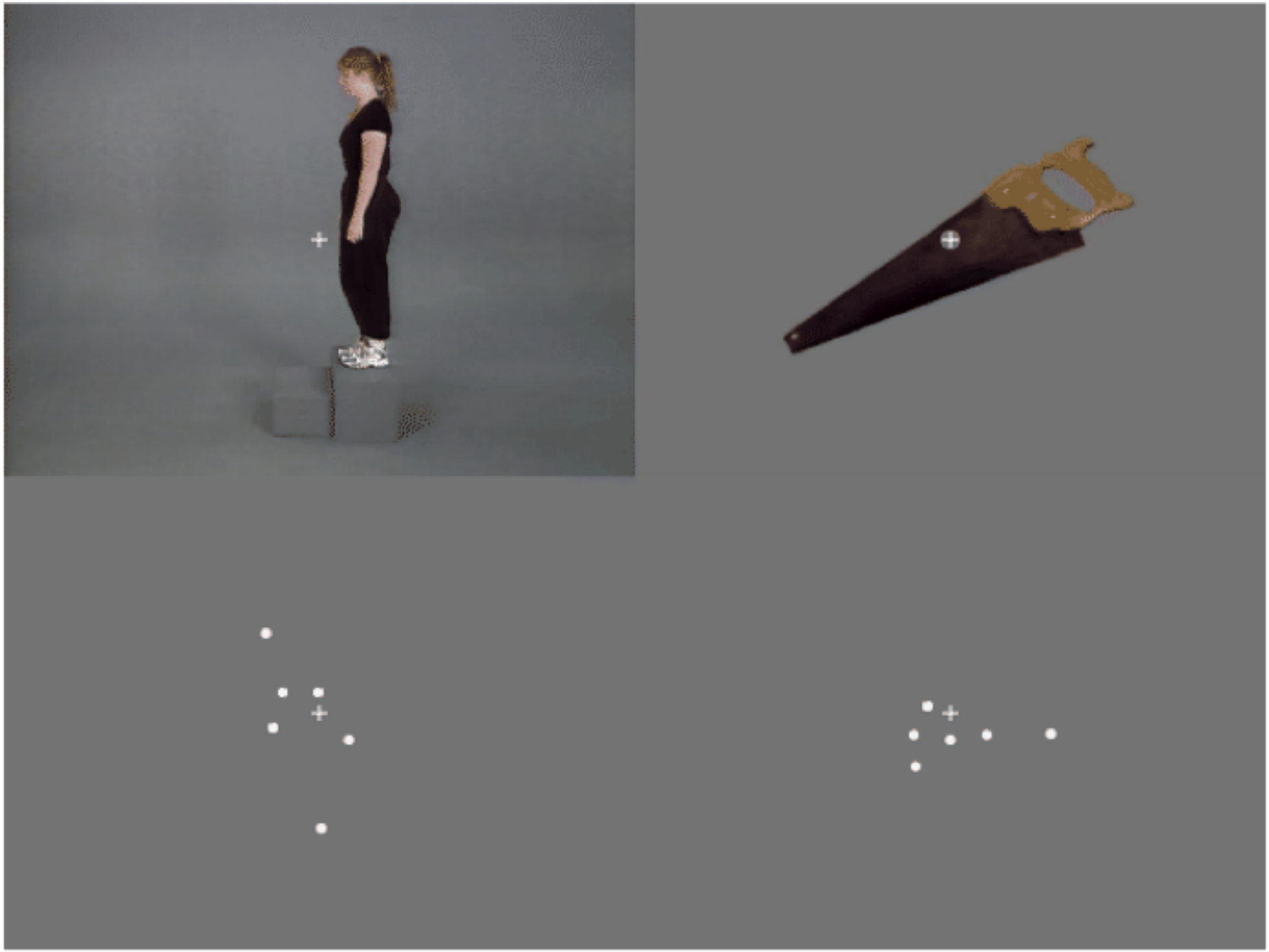
- Various group analysis approaches
  - Student's ***t*-test**: one-, two-sample, and paired
  - **ANOVA**: one or more categorical explanatory variables (factors)
  - **GLM**: AN(C)OVA
  - **LME**: linear mixed-effects modeling
- Easy to understand: *t*-tests not always practical or feasible
  - Tedious when layout (structure of data) is too complex
  - Main effects and interactions: desirable
  - Controlling for quantitative covariates
- Advantages of big models: AN(C)OVA, GLM, LME
  - All tests in one analysis (vs. piecemeal *t*-tests): omnibus *F*
  - Controlling for covariate effects
  - Power gain: combining subjects across groups for estimates of signal *and* noise parameters (*i.e.,* variances and correlations)

Terminology: Explanatory variables

- **Response/Outcome variable** (HDR): regression β coefficients
- **Factor**: categorical, qualitative, descriptive, nominal, or discrete
  - Categorization of conditions/tasks
    - Within-subject (repeated-measures) factor
  - Subject-grouping: group of subjects
    - Between-subjects factor (nesting)
    - Gender, patients/controls, genotypes, handedness, …
  - Subject: random factor measuring deviations
    - Of no interest, but served as random samples from a population
- **Quantitative** (numeric or continuous) **covariate**
  - Three usages of 'covariate'
    - Quantitative value (rather than strict separation into groups)
    - Variable of no interest: qualitative (scanner, sex, handedness) or quantitative
    - Explanatory variable (regressor, independent variable, or predictor)
  - Examples: age, IQ, reaction time, brain volume, …

Terminology: Fixed effects

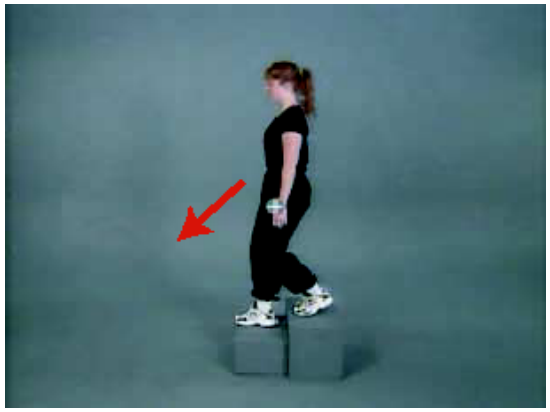
- Fixed-effects factor: categorical (qualitative or discrete) variable
  - Treated as a fixed variable (constant to be estimated) in the model
    - Categorization of conditions/tasks (modality: visual/auditory)
      - Within-subject (repeated-measures) factor: 3 emotions
    - Subject-grouping: Group of subjects (gender, controls/patients)
      - Between-subject factor
  - All factor levels are of interest: **not interchangeable/replaceable**
    - main effect, contrasts among levels
  - Fixed in the sense of statistical inferences
    - Apply only to the specific levels of the factor: : **replacement test**
      - Categories: human, tool
    - Don't extend to other potential levels that might have been included (but were not)
      - Inferences from viewing human and tool categories can't be generated to animals or clouds or Martians
- Fixed-effects variable: quantitative covariate



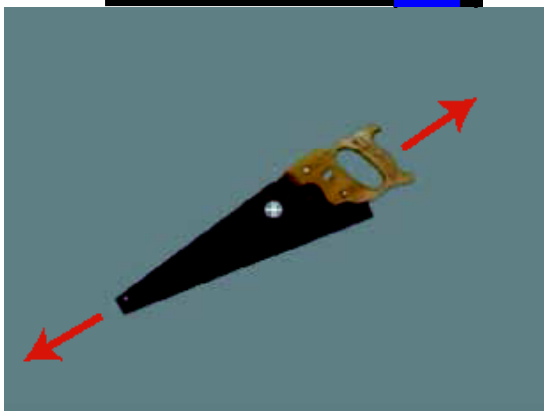
Each video is only shown once (2 seconds)

Remember This Study?

Human whole-body motion (HM)



Tool motion (TM)



Human point motion (HP)

From Figure 1
Beauchamp et al. 2003

Tool point motion (TP)

2 Factors, each with 2 levels

- **Factor A = type of object being viewed**
  - **Levels = Human or Tool**
- **Factor B = type of display seen by subject**
  - **Levels = Whole or Points**
- **This is repeated measures (4 βs per subject), 2 × 2 factorial**

Terminology: Random effects

- Random factor/effect
  - Random variable in the model: exclusively used for subject in FMRI
    - average + effects attributable to each subject: *e.g.* $N(\mu, \tau^2)$
    - Requires enough subjects to estimate properly
  - Each individual subject effect is of NO interest: **replacement test**
    - Group response = 0.92%, subject 7 = 1.13%, random effect = 0.21%
  - Random in the sense
    - Subjects as random samples (representations) from a population
    - Inferences can be generalized to a hypothetical population
- A generic group model: decomposing each subject's response
  - Fixed (population) effects: universal constants (immutable): $\boldsymbol{\beta}$

$$\boldsymbol{y}_i = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$$

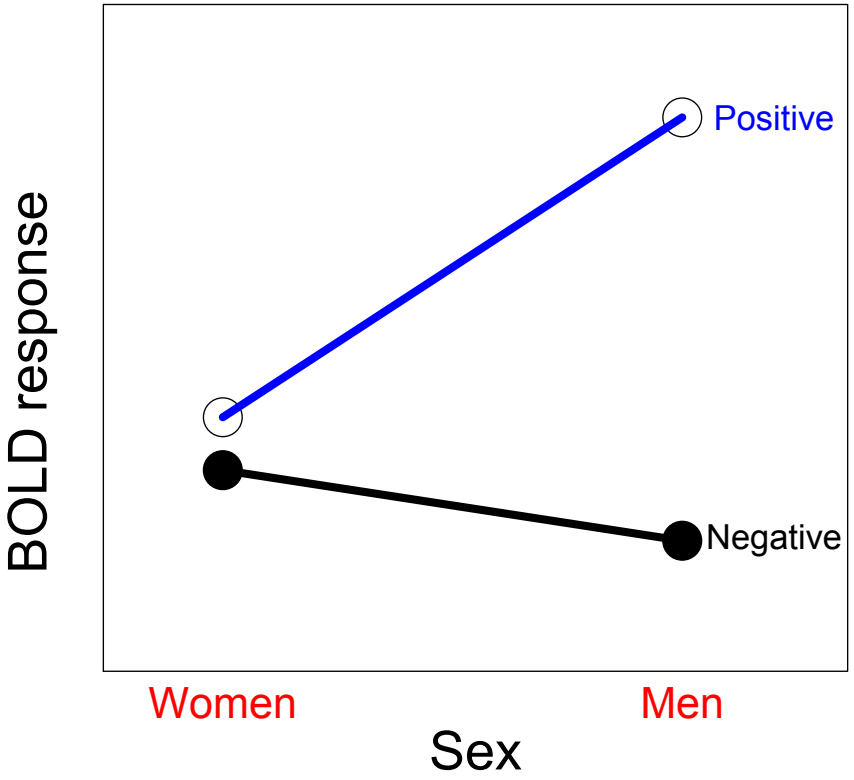
  - Random effects: individual subject's deviation from the population (personality: durable for subject i): $\boldsymbol{b}_i$
  - Residuals: noise (evanescent): $\boldsymbol{\varepsilon}_i$

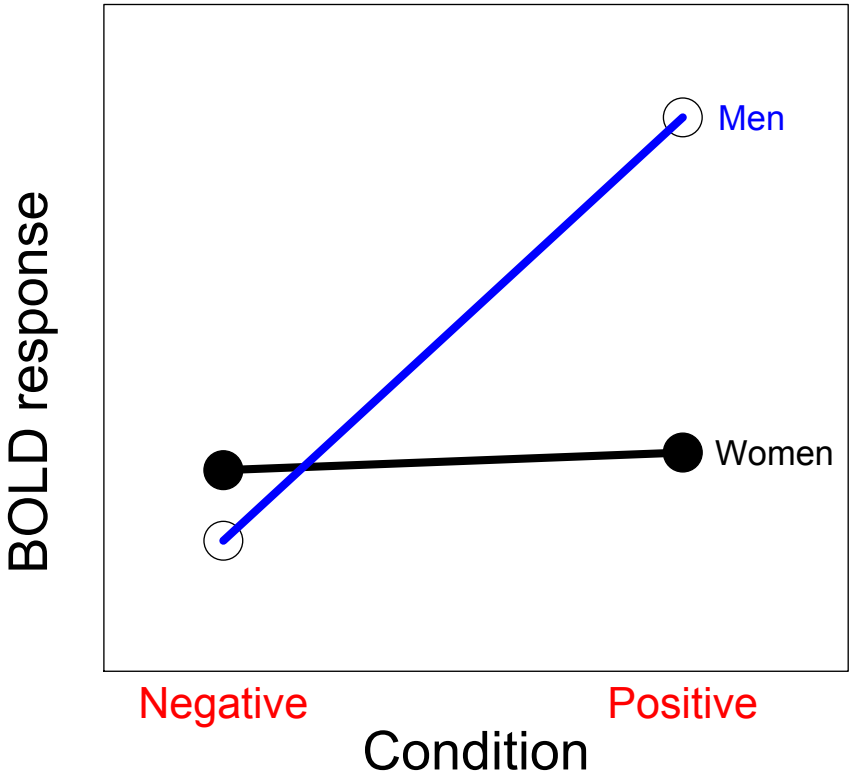
Fixed vs. Random effects

- Fixed effects
  - Effects treated as **constants**: happy, sad, neutral; house, face
  - Effects of interest: tasks, groups, age (or confounding effect)
  - Not exchangeable: order matters
  - Not replaceable
  - Not extendable to other effects: house, tool => face?
- Random effects
  - Representatives: subjects, families
  - Exchangeable: order does not matter
  - Replaceable
  - Inferences can be generalized to a hypothetical population
  - Associated with a probability distribution

Terminology: Omnibus tests - main effect and interaction

- Main effect: any difference across levels of a factor?
- Interactions: with ≥ 2 factors, interaction may exist
  - 2 × 2 design: *F*-test for interaction between A and B = *t*-test of (A1B1 - A1B2) - (A2B1 - A2B2) or (A1B1 - A2B1) - (A1B2 - A2B2)
    - *t* statistic is better than *F* : a positive *t* shows A1B1 - A1B2 > A2B1 - A2B2 and A1B1 - A2B1 > A1B2 - A2B2



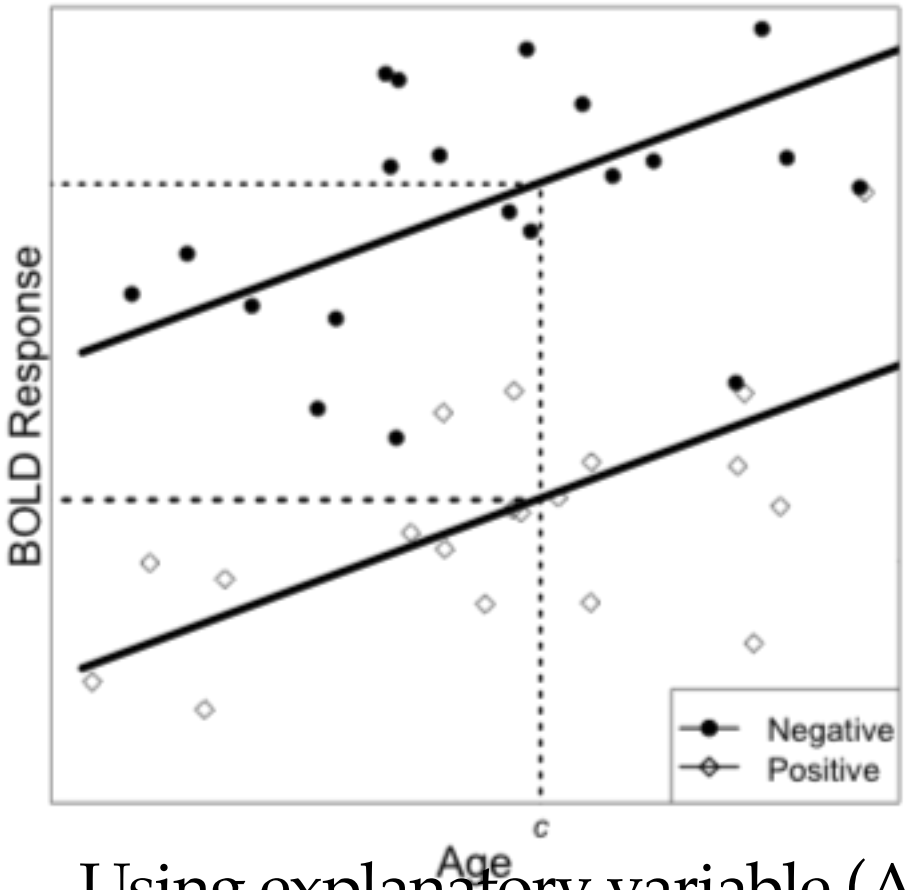


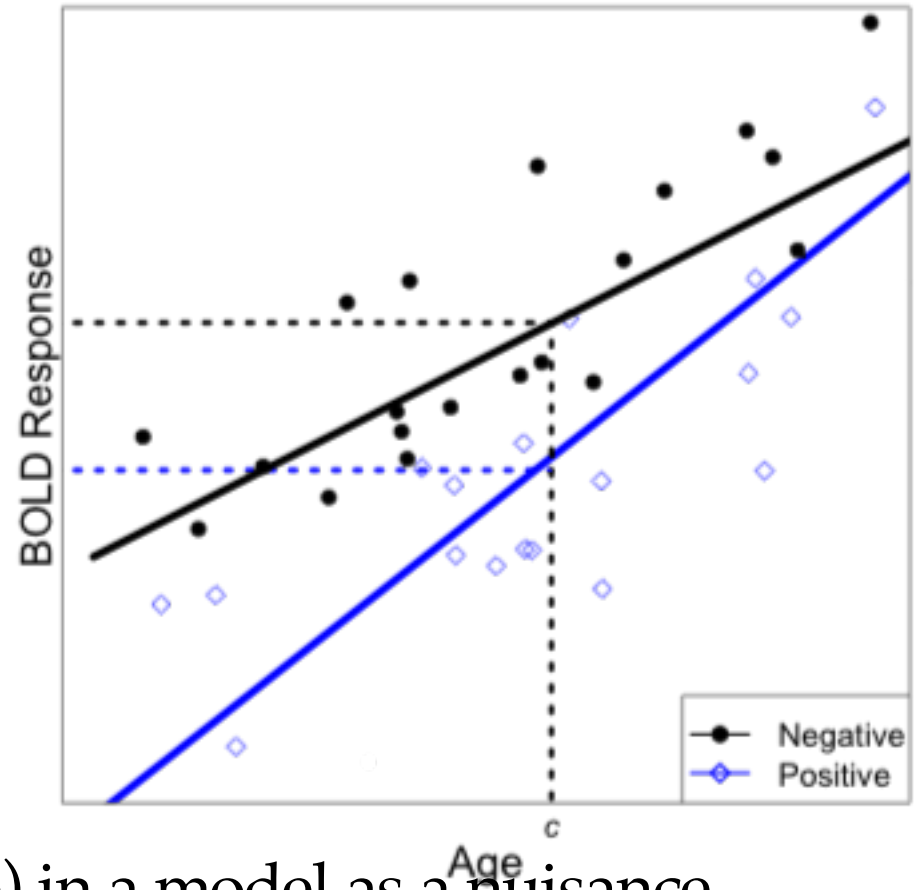
Terminology: Interaction

- Interactions: ≥ 2 factors
  - May become very tedious to sort out or understand!
    - ≥ 3 levels in a factor
    - ≥ 3 factors
  - Solutions: reduction (in complexity)
    - Pairwise comparison
    - Plotting: ROI averages
  - Requires sophisticated modeling
    - AN(C)OVA: 3dANOVAx, 3dMVM, 3dLME
- Interactions: quantitative covariates
  - In addition to linear effects, may have nonlinearity: y might depend on products of covariates: $x_1 * x_2$, or x^2

Terminology: Interaction

- Interaction: between a factor and a quantitative covariate





  - Using explanatory variable (Age) in a model as a nuisance regressor (additive effect) may not be enough
    - Model building/tuning: Potential interactions with other explanatory variables? (as in graph on the right)
    - Of scientific interest (*e.g.*, gender differences)

Models at Group Level

- Conventional approach: taking β (or linear combination of multiple βs) only for group analysis
  - Assumption: all subjects have same precision (reliability, standard error, confidence interval) about β
  - All subjects are treated equally
  - Student *t*-test: paired, 1- and 2-sample
  - AN(C)OVA, GLM, LME
- More precise method: taking both effect estimates and *t*-stats
  - *t*-statistic contains precision information about effect estimates
  - Each subject's β is weighted based on precision of effect estimate (more precise βs get more weight)
  - Currently only available for *t*-test types
- Best approach: combining all subjects in one big model

Piecemeal *t*-tests: 2 × 3 Mixed ANCOVA example

- A relatively simple model, but **challenging** for neuroimaging
  - 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
  - Quantitative covariate: Age
- Using Multiple *t*-tests for this study
  - Group comparison + age effect
  - Pairwise comparisons among three conditions
    - Cannot control for age effect
  - Effects that cannot be analyzed as *t*-tests
    - Main effect of Condition (3 levels is beyond *t*-test method)
    - Interaction between Group and Condition (6 levels total)
    - Age effect across three conditions (just too complicated)

Classical ANOVA: 2 × 3 Mixed ANOVA

- 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): **cannot** be modeled; **no** correction for sphericity violation

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

Different denominator

where

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

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

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

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

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

Univariate GLM: 2 x 3 mixed ANOVA

- Group: 2 levels (patient and control)
- Condition: 3 levels (pos, neg, neu)
- Subject: 3 ASD children and 3 healthy controls

Difficult to incorporate covariates

- Broken orthogonality of matrix

No correction for sphericity violation

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

Univariate GLM: popular in neuroimaging

- Advantages: more *flexible* than the method of sums of squares
  - No limit on the the number of explanatory variables (in principle)
  - Easy to handle unbalanced designs
  - Covariates easily modeled *when* no within-subject factors present
- Disadvantages: costs paid for the flexibility
  - Intricate dummy coding (to allow for different factors and levels)
  - Tedious pairing for numerator and denominator of *F*-stat
    - Choosing proper denominator SS is not obvious (errors in some software)
    - Can't generalize (in practice) to any number of explanatory variables
    - Susceptible to invalid formulations and problematic post hoc tests
  - **Cannot** handle covariates in the presence of within-subject factors
  - **No** direct approach to correcting for sphericity violation
    - **Unrealistic** assumption: **same** variance-covariance structure
- **Problems**: When overall residual SS is adopted for all tests
  - *F*-stat: valid only for **highest** order interaction of within-subject factors
  - Most post hoc tests are inappropriate with this denominator

Univariate GLM: **problematic implementations**

- Between-subjects Factor A (Group): 2 levels (patient, control)
- Within-subject Factor B (Condition): 3 levels (pos, neg, neu)
- A) Omnibus tests

Correct

$$F_A = \frac{MSA}{MSA(C)},$$

$$F_B = \frac{MSB}{MSE},$$

$$F_{AB} = \frac{MSAB}{MSE}$$

Incorrect

$$F_A = \frac{MSA}{MSE},$$

$$F_B = \frac{MSB}{MSE},$$

$$F_{AB} = \frac{MSAB}{MSE}$$

- B) Post hoc tests (contrasts)
- (1) **Incorrect** *t*-tests for factor A due to incorrect denominator
- (2) **Incorrect** *t*-tests for factor B or interaction effect AB when weights do not add up to 0
- C) How to handle multiple *β*s per effect (e.g., multiple runs)?
- -- **Artificially inflated DOF and assumption violation** when multiple *β*s are fed into program

Univariate GLM: **problematic implementations**

- Within-subjects Factor A (Object): 2 levels (house, face)
- Within-subject Factor B (Condition): 3 levels (pos, neg, neu)
- A) Omnibus tests

Correct

$$F_A = \frac{MSA}{MSAC},$$

$$F_B = \frac{MSB}{MSBC},$$

$$F_{AB} = \frac{MSAB}{MSE}$$

Incorrect

$$F_A = \frac{MSA}{MSE},$$

$$F_B = \frac{MSB}{MSE},$$

$$F_{AB} = \frac{MSAB}{MSE}$$

- B) Post hoc tests (contrasts)
- (1) **Incorrect** *t*-tests for both factors A and B due to incorrect denominator
- (2) **Incorrect** *t*-tests for interaction effect AB if weights don't add up to 0
- C) How to handle multiple *β*s per effect (e.g., multiple runs)?
- -- **Artificially inflated DOF and assumption violation** when multiple *β*s are fed into program

Better Approach: Multivariate GLM

- Why 3 ANOVA programs in AFNI: 3dANOVA, 3dANOVA2, 3dANOVA4?
- Group: 2 levels (patient and control)
- Condition: 3 levels (pos, neg, neu)
- Subject: 3 ASD children and 3 healthy controls
- Age: quantitative covariate

$$\boldsymbol{B}_{n\times m} = \boldsymbol{X}_{n\times q}\,\boldsymbol{A}_{q\times m} + \boldsymbol{D}_{n\times m}$$

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

Why use β , not t, values for group analysis?

- Why not use individual level statistics (t, F)?
  - Dimensionless, no physical meaning
  - Sensitive to sample size (number of trials) and to signal-to-noise ratio: may vary across subjects
    - Are t-values of 4 and 100 (or p-values of 0.05 and 10^{-8}) really informative? The HDR of the latter is not necessarily 25 times larger than the former
  - Distributional considerations – not Gaussian
- β values
  - Have physical meaning: measure HDR magnitude = % signal change (*i.e.,* how much BOLD effect); **importance of calibration**
- Using β values ***and*** their t-statistics at the group level
  - More accurate approach: 3dMEMA
  - **Mostly** about the same as the conventional (β only) approach
  - Identifying regions with substantial cross-subject variability

Road Map: Choosing a program for Group Analysis?

- Starting with HDR estimated via shape-fixed method (SFM)
  - One β per condition per subject
  - It might be significantly underpowered (more later)
- Two perspectives
  - Data structure
  - Ultimate goal: list **all** the tests you want to perform
    - Possible to avoid a big model this way
    - Use a piecemeal approach with 3dttest++ or 3dMEMA
      - Perform each test on your list separately
      - Difficulty: there can be ***many*** tests you *might* want
- Most analyses can be done with **3dMVM** and **3dLME**
  - Computationally inefficient
  - Last resort: not recommended if simpler alternatives (*e.g.*, *t*-tests) are available

Road Map: Student's *t*-tests

- **3dttest++** (new version of 3dttest) and **3dMEMA**
- Not for *F*-tests except for ones with 1 DoF for numerator
  - All factors are of two levels (at most), e.g., 2 x 2, or 2 x 2 x 2
- Scenarios
  - One-, two-sample, paired
  - Univariate GLM
    - Multiple regression: 1 group + 1 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:

      https://afni.nimh.nih.gov/sscc/gangc/MEMA.html
    - One group against a constant: 3dttest/3dttest++ –singletonA
      - The "constant" can depend on voxel, or be fixed

Road Map: between-subjects ANOVA

- One-way between-subjects ANOVA
  - **3dANOVA**
  - 2 groups of subjects: 3dttest++, 3dMEMA (OK with > 2 groups too)
- Two-way between-subjects ANOVA
  - Equal #subjects across groups: **3dANOVA2** –type 1
  - Unequal #subjects across groups: **3dMVM**
  - 2 x 2 design: **3dttest++**, **3dMEMA** (OK with > 2 groups too)
- Three-way between-subjects ANOVA
  - **3dANOVA3** –type 1
  - Unequal #subjects across groups: **3dMVM**
  - 2 x 2 design: **3dttest++**, **3dMEMA** (OK with > 2 groups too)
- *N*-way between-subjects ANOVA
  - **3dMVM**

Road Map: within-subject ANOVA

- ✧ Only one group of subjects
- ✧ One-way within-subject ANOVA
  - **3dANOVA2** –type 3
  - Two conditions: **3dttest++**, **3dMEMA**
- ✧ Two-way within-subject ANOVA
  - **3dANOVA3** –type 4
    - (2 or more factors, 2 or more levels each)
  - 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
  - Equal #subjects across groups: **3dANOVA3** –type 5
  - Unequal #subjects across groups: **3dMVM**
  - 2 x 2 design: **3dttest++**, **3dMEMA**
- More complicated scenarios
  - Multi-way ANOVA: **3dMVM**
  - Multi-way ANCOVA (between-subjects covariates only): **3dMVM**
  - HDR estimated with multiple bases: **3dANOVA3, 3dLME**, **3dMVM**
  - Missing data: **3dLME**
  - Within-subject covariates: **3dLME**
  - Subjects genetically related: **3dLME**
  - Trend analysis: **3dLME**

One-Sample Case

- One group of subjects ($n \geq 10$)
  - One condition (visual or auditory) effect
  - Linear combination of multiple effects (visual vs. auditory)
- Null hypothesis H_0: average effect = 0
  - Rejecting H_0 is of interest!
- Results
  - Average effect at group level (OLay)
  - Significance: ***t***-statistic (Thr - Two-tailed by default **in AFNI**)
- Approaches
  - **uber_ttest.py** (gen_group_command.py) – graphical interface
  - **3dttest++**
  - **3dMEMA**

One-Sample Case: Example

- **3dttest++**: taking β only for group analysis

```
3dttest++ -prefix VisGroup -mask mask+tlrc -zskip  \
 -setA 'FP+tlrc[Vrel#0_Coef]'                      \
       'FR+tlrc[Vrel#0_Coef]'                      \
       ......
       'GM+tlrc[Vrel#0_Coef]'
```

Voxel value = 0 → treated it as missing

- **3dMEMA**: taking β and t-statistic for group analysis

```
3dMEMA -prefix VisGroupMEMA -mask mask+tlrc -setA Vis \
 FP 'FP+tlrc[Vrel#0_Coef]' 'FP+tlrc[Vrel#0_Tstat]'    \
 FR 'FR+tlrc[Vrel#0_Coef]' 'FR+tlrc[Vrel#0_Tstat]'    \
       ......
 GM 'GM+tlrc[Vrel#0_Coef]' 'GM+tlrc[Vrel#0_Tstat]'    \
 -missing_data 0
```

Voxel value = 0 → treated it as missing

Two-Sample Case

- Two groups of subjects ($n \geq 10$ each): males and females
  - One condition (*e.g.*, visual or auditory) effect
  - Linear combination of multiple effects (*e.g.*, visual minus auditory)
  - Example: Gender difference in emotional effect of stimulus?
- Null hypothesis H_0: Group1 = Group2
  - Results
    - Group difference in average effect
    - Significance: *t*-statistic - **Two-tailed by default in AFNI**
- Approaches
  - **uber_ttest.py**, **3dttest++**, **3dMEMA**
  - One-way between-subjects ANOVA
    - **3dANOVA**: can also obtain individual group *t*-tests

Paired Case

- One groups of subjects ($n \geq 10$)
  - 2 conditions (visual or auditory): no missing data allowed (**3dLME**)
- Null hypothesis H_0: Condition1 = Condition2
  - Results
    - Average difference at group level
    - Significance: *t*-statistic (**two-tailed by default**)
- Approaches
  - **uber_ttest.py**, **gen_group_command.py**, **3dttest++**, **3dMEMA**
  - One-way within-subject (repeated-measures) **ANOVA**
    - **3dANOVA2 –type 3**: can also get individual condition test
  - Missing data (**3dLME**): only 10 of 20 subjects have both βs
- Essentially same as one-sample case using contrast as input

Paired Case: Example

- **3dttest++**: comparing two conditions

```
3dttest++ -prefix Vis_Aud          \
  -mask mask+tlrc -paired -zskip          \
  -setA 'FP+tlrc[Vrel#0_Coef]'    \
        'FR+tlrc[Vrel#0_Coef]'    \
        ……
        'GM+tlrc[Vrel#0_Coef]'    \
  -setB 'FP+tlrc[Arel#0_Coef]'    \
        'FR+tlrc[Arel#0_Coef]'    \
        ……
        'GM+tlrc[Arel#0_Coef]'
```

Paired Case: Example

- 3dMEMA: comparing two conditions using subject-level response magnitudes and estimates of error levels
  - Contrast should come from each subject
    - Instead of doing contrast inside 3dMEMA itself

```
3dMEMA -prefix Vis_Aud_MEMA                                    \
-mask mask+tlrc -missing_data 0                                \
-setA Vis-Aud                                                  \
FP 'FP+tlrc[Vrel-Arel#0_Coef]' 'FP+tlrc[Vrel-Arel#0_Tstat]' \
FR 'FR+tlrc[Vrel-Arel#0_Coef]' 'FR+tlrc[Vrel-Arel#0_Tstat]' \
  ……
GM 'GM+tlrc[Vrel-Arel#0_Coef]''GM+tlrc[Vrel-Arel#0_Tstat]'
```

One-Way Between-Subjects ANOVA

- Two **or more** groups of subjects ($n \geq 10$)
  - One condition or linear combination of multiple conditions
  - Example: visual, auditory, or visual vs. auditory
- Null hypothesis H_0: Group1 = Group2
  - Results
    - Average group difference
    - Significance: ***t***- and ***F***-statistic (two-tailed by default)
- Approaches
  - **3dANOVA** (for more than 2 groups)
  - **>** 2 groups: pair-group contrasts: **3dttest++**, **3dMEMA**
  - Dummy coding: **3dttest++**, **3dMEMA** (hard work)
  - **3dMVM**

Multiple-Way Between-Subjects ANOVA

- Two or more subject-grouping factors: factorial designs
  - One condition or linear combination of multiple conditions
  - Examples: gender, control/patient, genotype, handedness
- Testing main effects, interactions, single group, group comparisons
  - Significance: ***t***- (two-tailed by default) and *F*-statistic
- Approaches
  - Factorial design (imbalance not allowed): two-way (**3dANOVA2 –type 1**), three-way (**3dANOVA3 –type 1**)
  - **3dMVM**: no limit on number of factors (imbalance OK)
  - All factors have two levels: **3dttest++**, **3dMEMA**
  - Using group coding (via covariates) with **3dttest++**, **3dMEMA**: imbalance possible

One-Way Within-Subject ANOVA

- Also called **one-way repeated-measures**: one group of subjects ($n \geq 10$)
  - Two or more conditions: extension to paired *t*-test
  - Example: happy, sad, neutral conditions
- Main effect, simple effects, contrasts, general linear tests,
  - Significance: *t*- (two-tailed by default) and F-statistic
- Approaches
  - **3dANOVA2 -type 3** (2-way ANOVA w/ 1 random factor)
  - With two conditions, **equivalent** to paired case with **3dttest++**, **3dMEMA**
  - With more than two conditions, can break into pairwise comparisons with **3dttest++**, **3dMEMA**
  - **Univariate GLM**: testing one condition is **invalid**

One-Way Within-Subject ANOVA

- Example: visual vs. auditory condition

```
3dANOVA2 -type 3 -alevels 2 -blevels 10     \
 -prefix Vis_Aud -mask mask+tlrc            \
 -amean 1 Vis -amean 2 Aud -adiff 1 2 V-A   \
  -dset 1 1  'FP+tlrc[Vrel#0_Coef]'    \
  -dset 1 2  'FR+tlrc[Vrel#0_Coef]'    \
        ……
  -dset 1 10 'GM+tlrc[Vrel#0_Coef]'    \
  -dset 2 1  'FP+tlrc[Arel#0_Coef]'    \
  -dset 2 2  'FR+tlrc[Arel#0_Coef]'    \
        ……
  -dset 2 10 'GM+tlrc[Arel#0_Coef]'
```

Two-Way Within-Subject ANOVA

- Factorial design; also known as **two-way repeated-measures**
  - 2 within-subject factors
  - Example: emotion (happy/sad) and category (visual/auditory)
- Testing main effects, interactions, simple effects, contrasts
  - Significance: *t*- (two-tailed by default) and F-statistic
- Approaches
  - **3dANOVA3 –type 4** (three-way ANOVA with one random factor)
  - **3dMVM**
  - All factors have 2 levels (2x2): **3dttest++**, **3dMEMA**
  - Missing data?
    - Break into *t*-tests: **3dttest++**, **3dMEMA**
    - **3dLME**

Two-Way Mixed ANOVA

- Factorial design
  - One between-subjects and one within-subject factor
  - Example: between-subject factor = gender (male and female) and within-subject factor = emotion (happy, sad, neutral)
- Testing main effects, interactions, simple effects, contrasts
  - Significance: *t*- (two-tailed by default) and *F*-statistic
- Approaches
  - **3dANOVA3 –type 5** (three-way ANOVA with one random factor)
  - If all factors have 2 levels (2x2): **3dttest++**, **3dMEMA**
  - Missing data?
    - Unequal number of subjects across groups: **3dMVM**, GroupAna
    - Break into *t*-tests: **uber_ttest.py**, **3dttest++**, **3dMEMA**
    - **3dLME**

MVM Implementation in AFNI

✧ Program 3dMVM [written in R programming language]

- No tedious and error-prone dummy coding needed!
- Symbolic coding for variables and post hoc testing

Variable types

Post hoc tests

```
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 :'      \
        -dataTable                                                                           \
        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
```

Data layout

Group analysis with multiple basis functions

- Fixed-Shape method (**FSM**)
- Estimead-Shape method (**ESM**) via basis functions: TENTzero, TENT, CSPLINzero, CSPLIN
  - Area under the curve (AUC) approach
    - Ignore **shape** differences between groups or conditions
    - Focus on the response **magnitude** measured by AUC
    - Potential issues: Shape information lost; Undershoot may cause trouble (canceling out some of the positive signal)
  - Better approach: maintaining shape information
    - Take individual β values to group analysis (MVM)
- Adjusted-Shape method (**ASM**) via SPMG2/3
  - Only take the major component β to group level
  - ***or***, Reconstruct HRF, and take the effect estimates (*e.g.*, AUC)

Group analysis with multiple basis functions

- Analysis with effect estimates at consecutive time grids (from TENT or CSPLIN or reconstructed HRF)
  - Used to be considered very hard to set up (in GLM)
  - Extra variable in analysis: **Time** = $t_0, t_1, \ldots, t_k$
  - One group of subjects under one condition
    - **Accurate** null hypothesis is

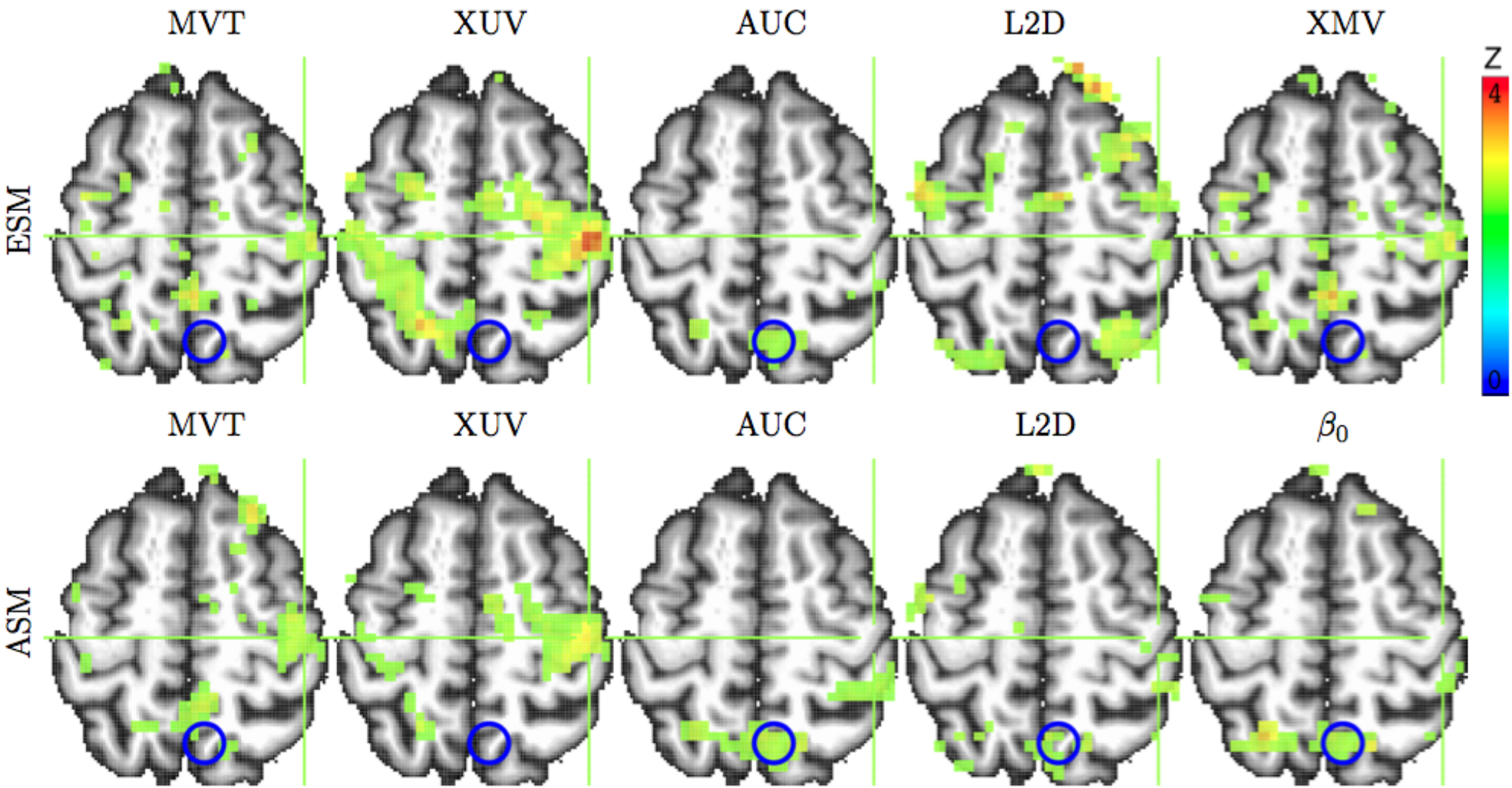
      H_0: $\beta_1=0, \beta_2=0, \ldots, \beta_k=0$ (NOT $\beta_1=\beta_2=\ldots=\beta_k$)
      - Testing the **centroid** (multivariate testing)
      - **3dLME**
    - **Approximate** hypothesis H_0: $\beta_1=\beta_2=\ldots=\beta_k$ (**main effect**)
      - **3dMVM**
    - Result: *F*-statistic for H_0 and *t*-statistic for each Time point

Group analysis with multiple basis functions

- Multiple groups (or conditions) under one condition (or group)
  - Accurate hypothesis: $\beta_1^{(1)} - \beta_1^{(2)} = 0, \beta_2^{(1)} - \beta_2^{(2)} = 0, \ldots, \beta_k^{(1)} - \beta_k^{(2)} = 0$
    - 2 conditions: **3dLME**
  - Approximate hypothesis: $\beta_1^{(1)} = \beta_1^{(2)}, \beta_2^{(1)} = \beta_2^{(2)}, \ldots, \beta_k^{(1)} = \beta_k^{(2)}$
    - Interaction
    - Multiple groups: **3dANOVA3 –type 5** (two-way mixed ANOVA: equal #subjects), or **3dMVM**
    - Multiple conditions: **3dANOVA3 –type 4**
  - Focus: do these groups/conditions have different response shape?
    - *F*-statistic for the interaction between Time and Group/Condition
    - *F*-statistic for main effect of Group: group/condition difference of AUC
    - *F*-statistic for main effect of Time: HDR effect across groups/conditions
- Other scenarios: factor, quantitative variables
  - **3dMVM**

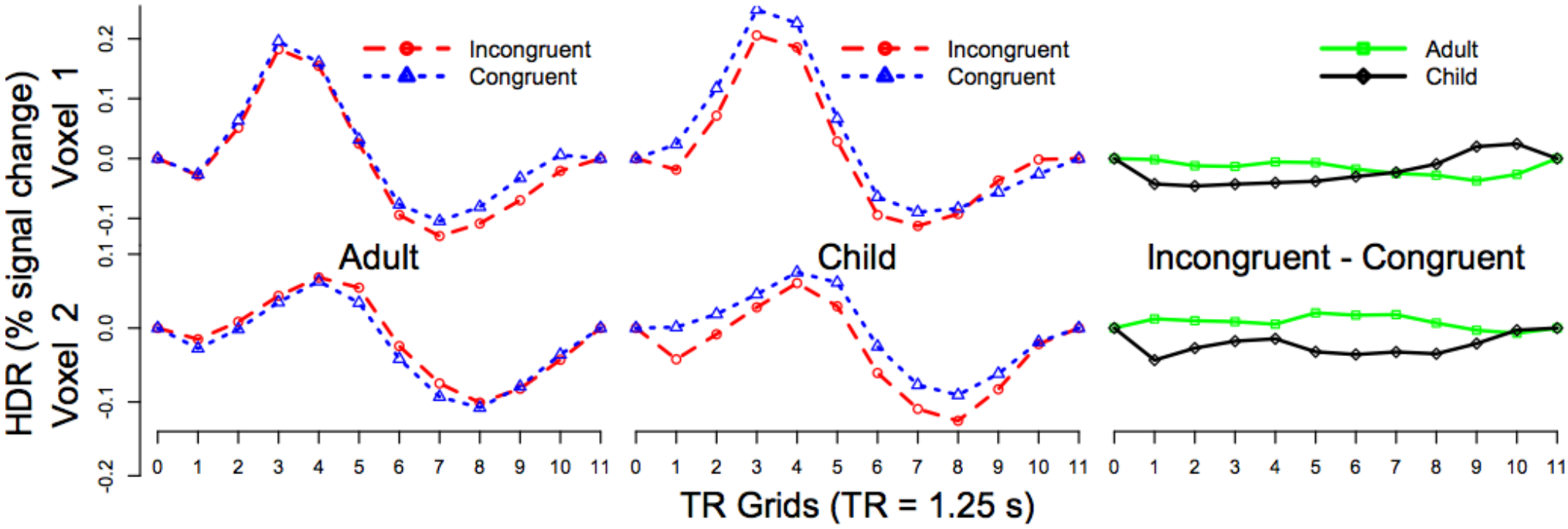
Group analysis with multiple basis functions

- 2 groups (children, adults), 2 conditions (congruent, incongruent), 1 quantitative covariate (age)
- 2 methods: HRF modeled by 10 (tents) and 3 (SPMG3) bases



Group analysis with multiple basis functions

- Advantages of ESM over FSM
  - More likely to detect HDR shape subtleties
  - Visual verification of HDR signature shape (vs. relying significance testing: *p*-values)
- Study: Adults/Children with Congruent/Incongruent stimuli (2 × 2)



Correlation analysis

- Correlation between brain response and behavioral measures

$$\hat{\beta}_i = \alpha_0 + \alpha_1 * x_i + \epsilon_i$$

  - Difference between correlation and regression?
    - Essentially the same
    - When explanatory (x_i) and response variable (β_i) are standardized (variance=1), then regression coefficient = correlation coefficient
  - Two approaches to get correlation from statistics software
    - Standardization
    - Convert *t*-statistic to *r* (or determination coefficient)

$$R^2 = t^2/(t^2 + DF)$$

    - Programs: **3dttest++**, **3dMEMA**, **3dMVM**, **3dRegAna**

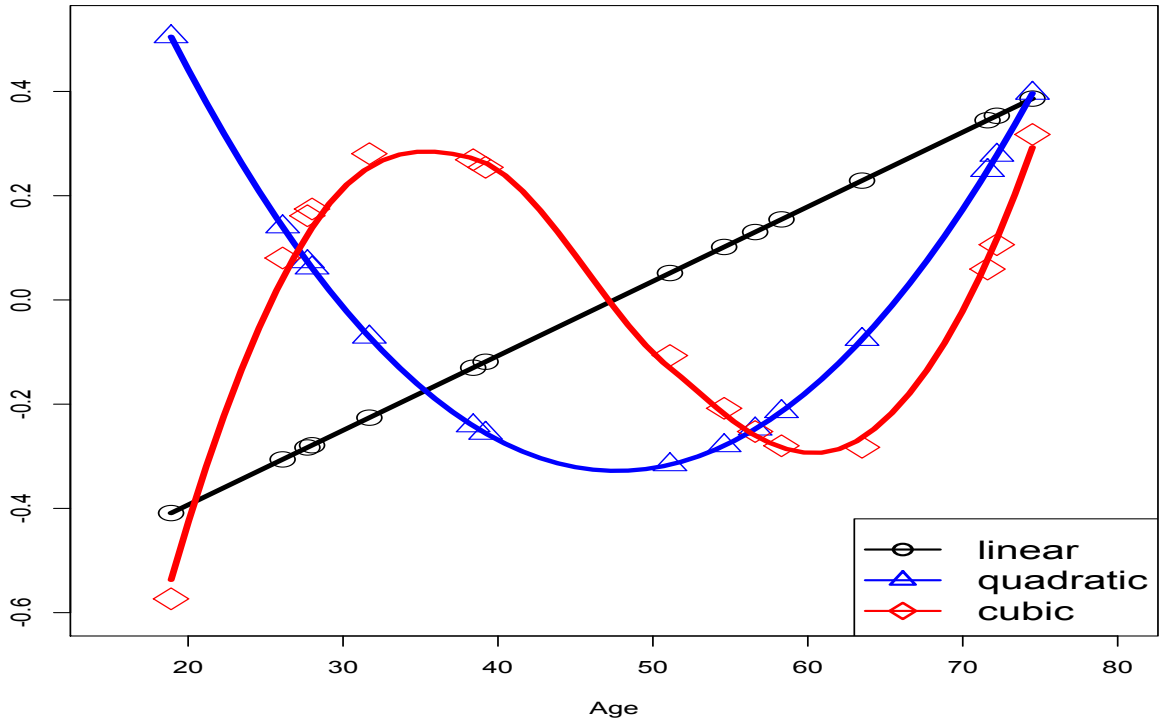
- Seed-based correlation for resting-state data
  - Fisher transform z has a variance of 1/(DoF – 2)
  - May consider further standardization by sqrt(DoF – 2)

Trend analysis

- Correlation between brain response and some gradation
  - **Linear**, **quadratic**, or **higher-order** effects
    - Habituation or attenuation effect across time (trials)
    - Between-subjects: Age, IQ
      - Fixed effect
    - Within-subject measures (covariates): morphed images
      - Random effects (trends in different subjects) : **3dLME**
  - Modeling: weights based on gradation
    - Equally-spaced: coefficients from **orthogonal polynomials**
    - With 6 equally-spaced levels, *e.g.*, 0, 20, 40, 60, 80, 100%,
      - Linear: -5 -3 -1 1 3 5
      - Quadratic: 5 -1 -4 -4 -1 5
      - Cubic: -5 7 4 -4 -7 5

Trend analysis

- Correlation between brain response and some gradation
  - Modeling: weights based on gradation
    - Not equally-spaced: constructed from, *e.g.*, poly() in R
    - Ages of 15 subjects: 31.7 38.4 51.1 72.2 27.7 71.6 74.5 56.6 54.6 18.9 28.0 26.1 58.3 39.2 63.5
    - https://afni.nimh.nih.gov/sscc/gangc/Trend.html



Trend analysis: summary

- **Cross-trials** trend: AM2 single subject analysis with weights
- Modeling with within-subject trend: **3** approaches
  - Set up GLT weights among factor levels at group level (not directly using covariate values) **3dANOVA2/3**, **3dMVM**, **3dLME**: best with equally-spaced with **even number of levels**
  - Set up the covariates as the values of a variable
    - Needs to account for deviation of each subject (random trends)
    - **3dLME**
  - Run trend analysis at individual level (*i.e.*, -gltsym), and then take the trend effect coefficient estimates to group level
    - Simpler than the other two approaches of doing trend analysis at the group level

Group analysis with quantitative variables

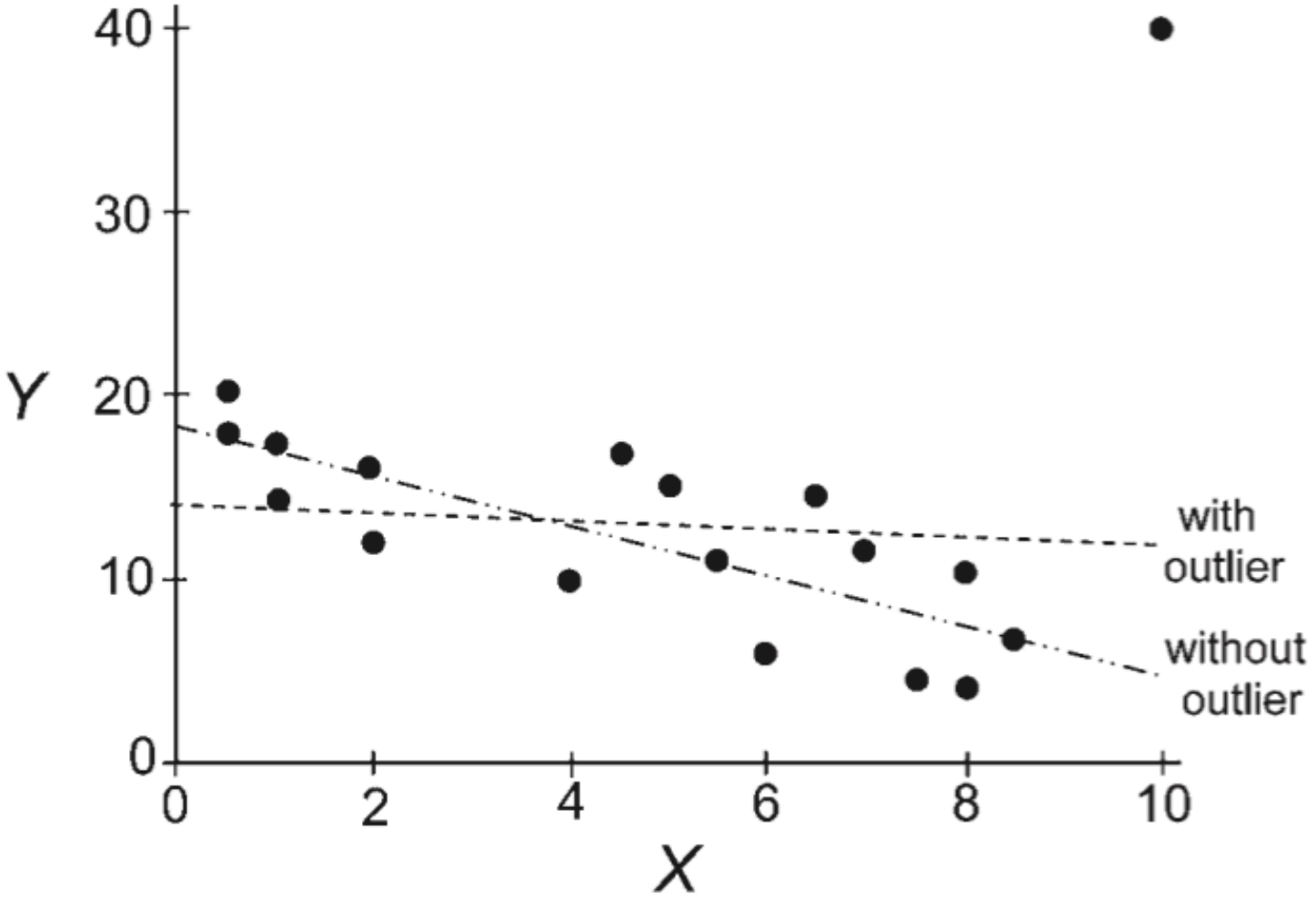
- Covariate: 3 usages
  - Quantitative (vs. categorical) variable of interest
    - Age, IQ, behavioral measures, …
  - Of no interest to the investigator (trying to remove variance)
    - Age, IQ, sex, handedness, scanner,…
  - Any explanatory variables in a model
- Variable selection
  - Infinite candidates for covariates: relying on prior information
  - Typical choices: age, IQ, RT (reaction time), …
  - RT: individual vs. group level
    - Amplitude Modulation regression: cross-trial variability at **individual** level (*cf*. Advanced Regression talk)
    - Group level: variability across subjects

Group analysis with quantitative variables

- Conventional framework
  - ANCOVA: one between-subjects factor (e.g., sex) + one quantitative variable (e.g., age)
    - Extension to ANOVA: GLM
    - Homogeneity of slopes
- Broader framework
  - Any modeling approaches involving quantitative variables
    - Regression, GLM, MVM, LME
    - Trend analysis, correlation analysis
- Interpretations
  - Regress/covariate out x?
  - "Controlling x at …", "holding x constant": centering

Caveats with covariate modeling

- Regression with few data points: sensitive to outliers
- Option **-robust** in **3dMVM**

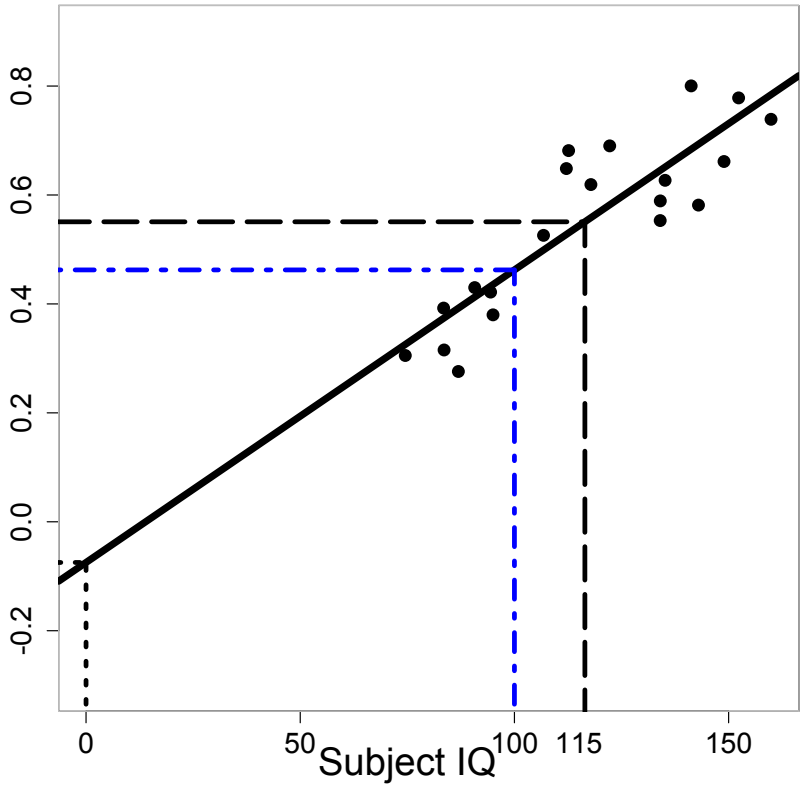


Caveats with covariate modeling

- **Specification error**: excluding a crucial explanatory variable may lead to incorrect or distorted interpretations (**spuriousness**)
  - Toddler's vocabulary ~ α * shoe size: α = .50
  - Toddler's vocabulary ~ α * shoe size + β * age: α = .04, β = .6
    - Explanatory variables (shoe size, age) are highly correlated: r = 0.8!
    - Excluding one may lead to overestimated effect for the other, but not *always* the case
- **Suppression**:
  - y (# suicide attempts) ~ 0.49 * x_1 (depression)
  - y ~ 0.19 * x_2 (amount of psychotherapy)
  - y ~ 0.70 * x_1 - 0.30 * x_2 (r_{12} = 0.7)
  - Imagine that x_1 is **head motion** in FMRI!

Quantitative variables: subtleties

- Regression: one group of subjects + quantitative variables $\hat{\beta}_i = \alpha_0 + \alpha_1 * x_{1i} + \alpha_2 * x_{2i} + \epsilon_i$
  - Interpretation of effects (results of regression)
    - α_1 - slope (change rate, marginal effect): effect per unit of x
    - α_0 – intercept: group effect when ***x*=0**
      - Not necessarily meaningful
      - Linearity may not hold
      - Solution: centering crucial for interpretability
      - Mean centering? or Median centering?

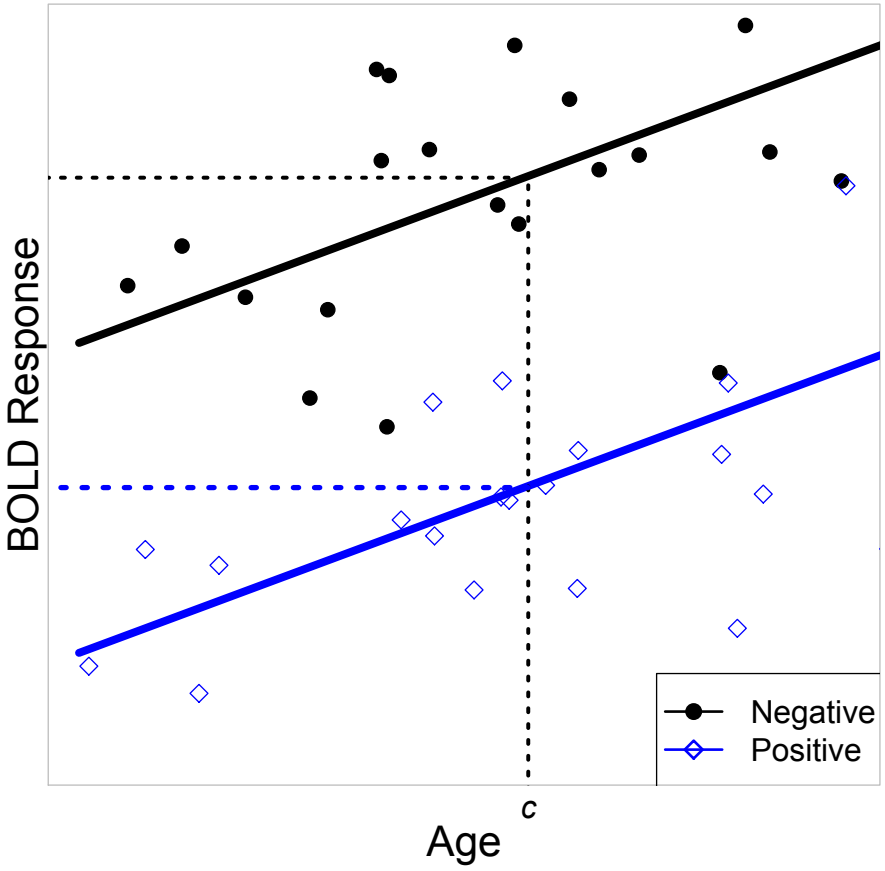


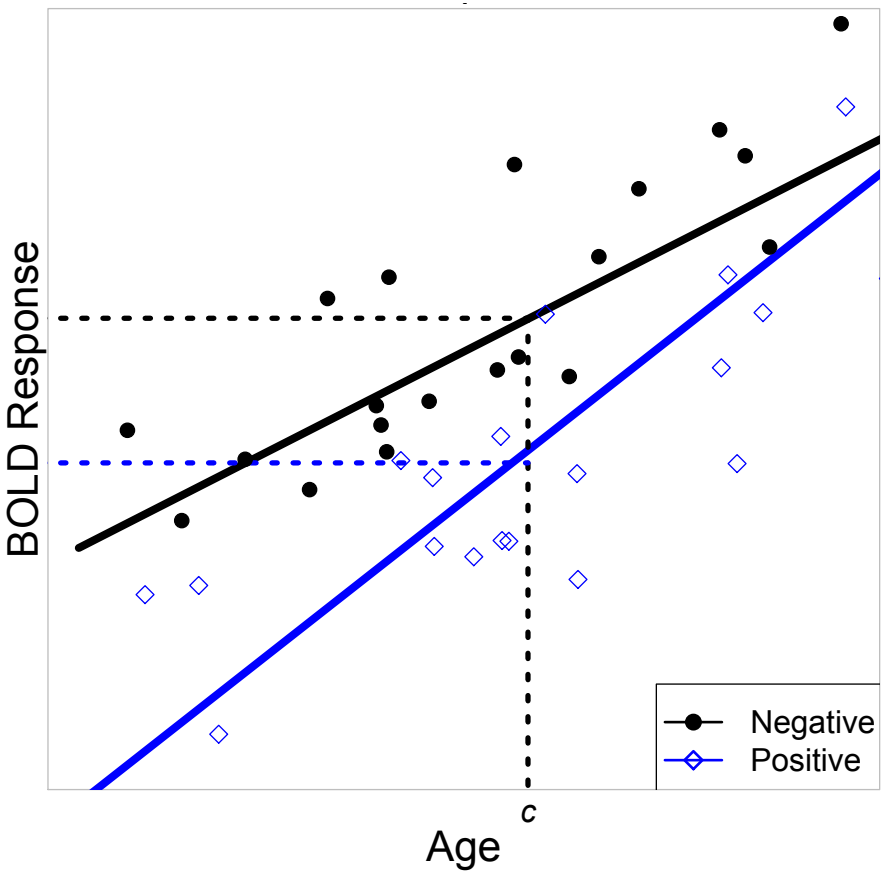
Quantitative variables: subtleties + confusion

- Trickier scenarios with two or more groups

$$\hat{\beta}_i = \alpha_0 + \alpha_1 * x_{1i} + \alpha_2 * x_{2i} + \alpha_3 * x_{3i} + \epsilon_{ij}$$

  - Interpretation of effects
    - Slope: Interaction! Same or different slope?
    - α_0 (intercept) – same or different center?

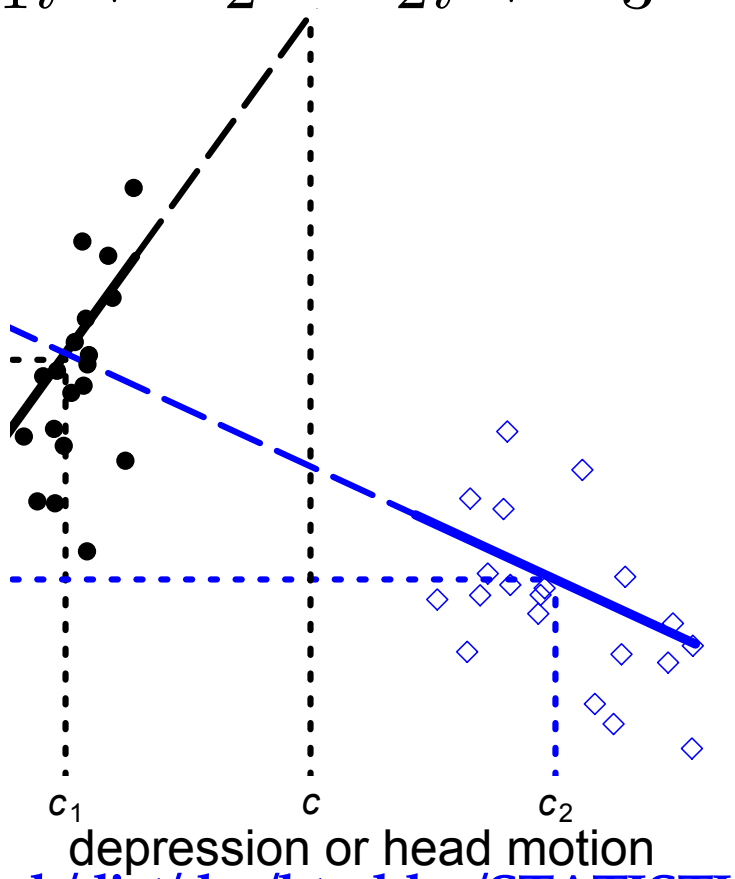




Quantitative variables: subtleties

- Trickiest scenario with two or more groups in addition to interaction

$$\hat{\beta}_i = \alpha_0 + \alpha_1 * x_{1i} + \alpha_2 * x_{2i} + \alpha_3 * x_{3i} + \epsilon_{ij}$$



- https://afni.nimh.nih.gov/pub/dist/doc/htmldoc/STATISTICS/center.html

IntraClass Correlation (ICC)

- Reliability (consistency, agreement/reproducibility) across two or more measurements of the same condition/task (sessions, scanners, sites, studies, twins -- monozygous or dizygous): extent to which the levels of a factor are related to each other
  - Example: 20 subjects scanned in two scanners (effect estimate of a condition/task, contrast between 2 conditions/tasks, correlation, etc.)
  - Classic example (Shrout and Fleiss, 1979): n targets are rated by k raters/judges
  - Relationship with Pearson correlation
    - Pearson correlation can be for any two different types of measure: e.g., BOLD response vs. RT
    - ICC is for the **same** measurement with the **same** assumption $G(\mu, \sigma^2)$

ICC(1,1)

- Traditional definition: one-way random-effects ANOVA

$$y_{ij} = b_0 + \lambda_j + \epsilon_{ij}$$

  - Assumptions: family $\boldsymbol{\lambda_j} \sim \boldsymbol{G}(0, \sigma_\lambda^2)$, $\varepsilon_{ij} \sim \boldsymbol{G}(0, \sigma_\varepsilon^2)$
  - Order cannot be assigned across multiple measurements
    - *e.g.*, twins: fixed or random effect of twins (index *j*) not considered
  - **ICC(1,1)** estimated via ANOVA (Shrout & Fleis, 1979)

$$\hat{\rho}_1 = \frac{MS_\lambda - MS_\epsilon}{MS_\lambda + (k-1)MS_\epsilon}$$

  - Conceptualized as an LME model

$$\rho_1 = \frac{\sigma_\lambda^2}{\sigma_\lambda^2 + \sigma_\epsilon^2}$$

ICC(2,1)

- Traditional definition: one-way random-effects ANOVA

$$y_{ij} = b_0 + \pi_i + \lambda_j + \epsilon_{ij}$$

  - Assumptions: $\pi_i \sim G(0, \sigma_\pi^2)$, subject $\lambda_j \sim G(0, \sigma_\lambda^2)$, $\varepsilon_{ij} \sim G(0, \sigma_\varepsilon^2)$
  - Order cannot be assigned across multiple measurements
    - *e.g.*, twins: fixed or random effect of twins (index *j*) not considered
  - **ICC(2,1)** estimated via ANOVA (Shrout & Fleis, 1979)

$$\hat{\rho}_2 = \frac{MS_\lambda - MS_\epsilon}{\frac{k}{n}(MS_\pi - MS_\epsilon) + MS_\lambda + (k-1)MS_\epsilon}$$

  - Conceptualized as an LME model

$$\rho_2 = \frac{\sigma_\lambda^2}{\sigma_\pi^2 + \sigma_\lambda^2 + \sigma_\epsilon^2}$$

ICC(3,1)

- Traditional definition: one-way random-effects ANOVA

$$y_{ij} = b_0 + b_i + \lambda_j + \epsilon_{ij}$$

  - Assumptions: subject $\boldsymbol{\lambda}_j \sim \boldsymbol{G}(0, \sigma_{\boldsymbol{\lambda}}{}^2)$, $\varepsilon_{ij} \sim \boldsymbol{G}(0, \sigma_{\varepsilon}{}^2)$
  - Order cannot be assigned across multiple measurements
    - *e.g.*, twins: fixed or random effect of twins (index *j*) not considered
  - **ICC(3,1)** estimated via ANOVA (Shrout & Fleis, 1979)

$$\hat{\rho}_3 = \frac{MS_\lambda - MS_\epsilon}{MS_\lambda + (k-1)MS_\epsilon}$$

  - Conceptualized as an ANOVA model

$$\rho_3 = \frac{\sigma_\lambda^2}{\sigma_\lambda^2 + \sigma_\epsilon^2}$$

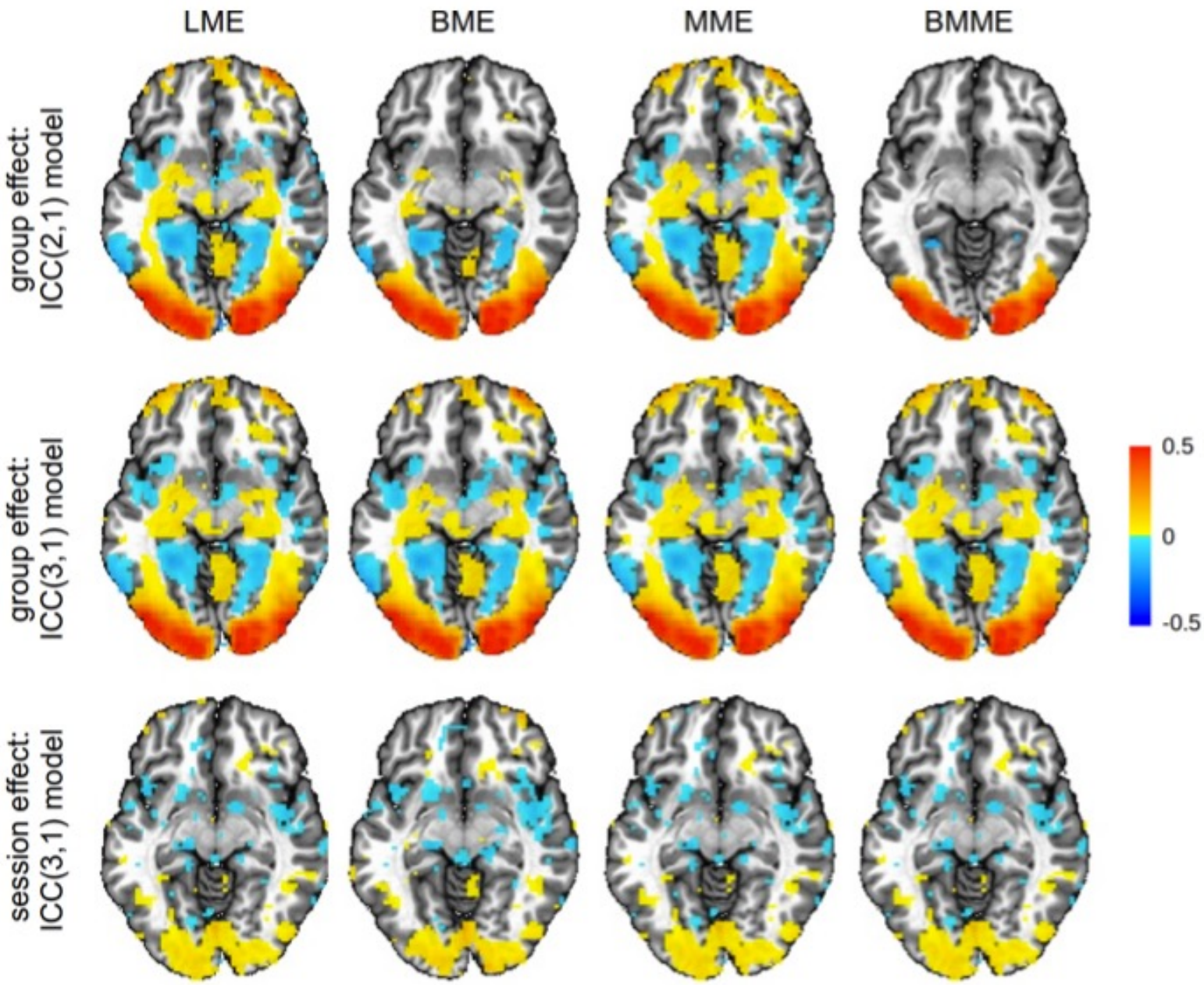
LME application 4: ICC under LME framework

- 3dLME –ICC or –ICCb: ICC(2,1)
- 3dICC

Issues	ANOVA	LME	BME	MME	BMME
negative ICC	✗	✓	✓	✓	✓
zero ICC	✗	✗	✓	✗	✓
missing data	✗	✓	✓	✓	✓
confounding effects	✗	✓	✓	✓	✓
sampling error	✗	✗	✗	✓	✓
type selection	✗	✓	✓	✓	✓
discrete responses	✗	✓	✓	✓	✓

LME application 4: ICC under LME framework

- 3dICC



Inter-Subject Correlation: Overview

- Naturalistic FMRI
  - A middle point between task-related and resting-state scanning
  - A special case of task-related FMRI: task from beginning to end
  - Resting-state data: an asymptotic case of naturalistic canning
- Challenges of analyzing naturalistic scanning data
- Survey of previous approaches
- Exploration of new nonparametric methods
- Flexibility of linear mixed-effects (LME) modeling (**program publicly available**)
- Potential application to resting-state data
  - Focus on whole brain instead of one seed

Two popular types of FMRI scanning

- Task-related experiments
  - Meticulously designed, well controlled
  - Trial duration: a few TRs or less
  - Effect of interest: regional responses to a task or a contrast
  - Models: BOLD responses estimated through time series regression
  - Potential issues: sensitivity (underpowered)

- Resting state
  - No explicit tasks
  - Spontaneous fluctuations
  - Effect of interest: regional correlation, networks
  - Models: seed-based correlation, data-driven methods, etc.
  - Caveats: difficult to separate physiological confounds

Naturalistic scanning

- Subjects view a natural scene during scanning
  - Visuoauditory movie clip (e.g., http://studyforrest.org/)
  - Neural responses shared across languages
  - Music, speech, games, …

- Duration: lasting for a few minutes or more
- Close to naturalistic settings: minimally manipulated; naturalistically, continuously, and dynamically evolving
- Effect of interest
  - Extent of synchronization/entrainment, similarity, or shared processing at the same brain regions across subjects in shared memory, communication and understanding through a common ground

- Hasson et al., 2004. Intersubject synchronization of cortical activity during natural vision. Science 303:1634-1640.

Inter-Subject Correlation (ISC)

- Modeling with task-related regressors won't work
  - One regressor for the whole task: BOLD can't be separated from baseline and drift effects
  - Feature extractions: too rich or complicated to be practical

- Inter-subject correlation (ISC)
  - Proper preprocessing
    - Nonlinear alignment to template space
    - Removing physiological confounds (e.g., regressing out signal in the white matter and principal components from the CSF signal)
    - Censoring out time points when significant motion occurred

Inter-Subject Correlation (ISC)

- Inter-subject correlation (ISC)
  - Correlation of time series between two subjects at the same voxel
  - No presumption of HDR
  - Measuring synchronization/similarity/entrainment
  - Avoiding the arbitrariness of seed selection

- Voxel-wise ISC between any subject pair
  - $n = 3$ subjects (A, B, C): 3 ISC values (AB, AC, BC)
  - $n = 4$ subjects: 6 ISCs
  - $n = 5$ subjects: 10 ISCs
  - n subjects: $n(n\text{-}1)/2$ ISCs

- ISC group analysis
  - Summarization at the group level
  - Investigate differences across groups in synchronization (ISC)
  - Difficulty: some of ISC values are correlated
    n independent samples correspond to $n(n\text{-}1)/2$ ISCs

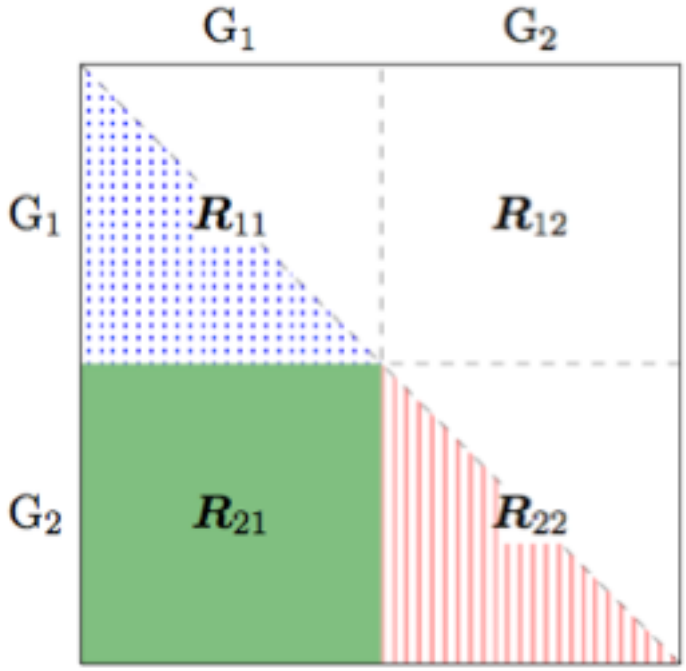
ISC group analysis

- Voxel-wise ISC matrix (usually Fisher-transformed)
  - One group

$$\boldsymbol{R}^{(n)} = \begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \\ \vdots \\ S_n \end{array} \begin{array}{c} \begin{array}{ccccc} S_1 & S_2 & S_3 & \cdots & S_n \end{array} \\ \begin{pmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1n} \\ r_{21} & 1 & r_{23} & \cdots & r_{2n} \\ r_{31} & r_{32} & 1 & \cdots & r_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & r_{n3} & \cdots & 1 \end{pmatrix} \end{array}$$

$$\boldsymbol{Z}^{(n)} = \begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \\ \vdots \\ S_n \end{array} \begin{array}{c} \begin{array}{ccccc} S_1 & S_2 & S_3 & \cdots & S_n \end{array} \\ \begin{pmatrix} - & z_{12} & z_{13} & \cdots & z_{1n} \\ z_{21} & - & z_{23} & \cdots & z_{2n} \\ z_{31} & z_{32} & - & \cdots & z_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & z_{n3} & \cdots & - \end{pmatrix} \end{array}$$

- Two groups
  - Within-group ISC: R11, R22
  - Inter-group ISC: R21
  - 3 group comparisons: R11 vs R22, R11 vs R21, R22 vs R21



Correlation pattern of ISC values

- 2 ISC values associated with a common subject are correlated with each other: 5 subjects, 10 ISC values
- $\rho \neq 0$ characterizes non-independent relationship

	Z_{21}	Z_{31}	Z_{41}	Z_{51}	Z_{32}	Z_{42}	Z_{52}	Z_{43}	Z_{53}	Z_{54}
Z_{21}	1	ρ	ρ	ρ	ρ	ρ	ρ	0	0	0
Z_{31}	ρ	1	ρ	ρ	ρ	0	0	ρ	ρ	0
Z_{41}	ρ	ρ	1	ρ	0	ρ	0	ρ	0	ρ
Z_{51}	ρ	ρ	ρ	1	0	0	ρ	0	ρ	ρ
Z_{32}	ρ	ρ	0	0	1	ρ	ρ	ρ	ρ	0
Z_{42}	ρ	0	ρ	0	ρ	1	ρ	ρ	0	ρ
Z_{52}	ρ	0	0	ρ	ρ	ρ	1	0	ρ	ρ
Z_{43}	0	ρ	ρ	0	ρ	ρ	0	1	ρ	ρ
Z_{53}	0	ρ	0	ρ	ρ	0	ρ	ρ	1	ρ
Z_{54}	0	0	ρ	ρ	0	ρ	ρ	ρ	ρ	1

- **Challenge**: how to handle this irregular correlation matrix?

ISC group analysis: previous methods

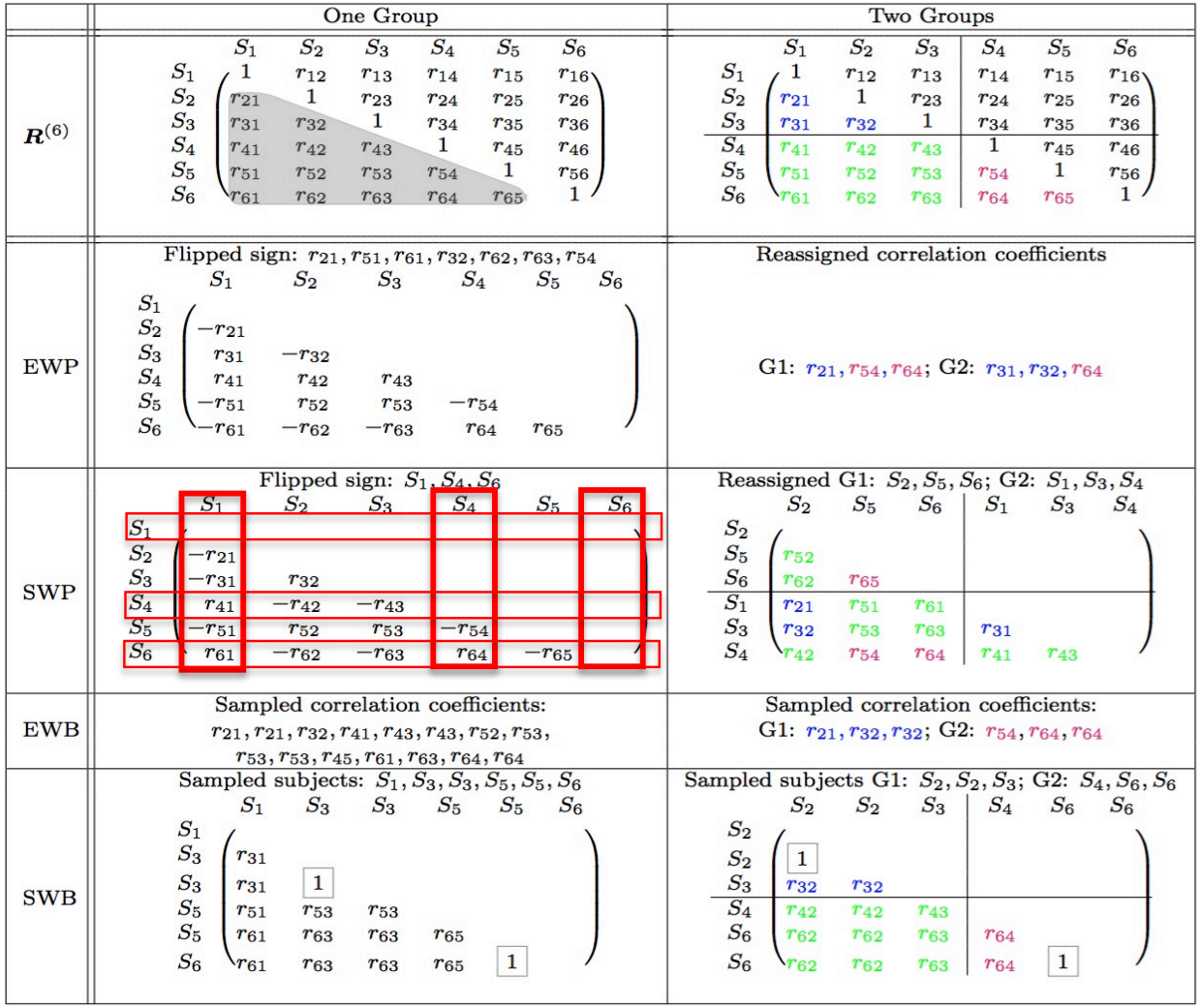
- Student's *t*-test
  - Independence violation acknowledged but not accounted for
  - Justification via observations that "null data" (generated by ISC values with randomly shifted time series) followed *t*(*N*-1)

- Various nonparametric methods
  - Permutations: null distribution via randomization across space (voxels) and time (e.g., circularly shifting each subject's time series by a random lag)
    - Matlab package: **ISC Toolbox** (Kauppi et al, 2014)
  - Leave one out (LOO): Kauppi et al, 2010
    - Compute ISC of a subject between a voxel's BOLD time course in the subject and the average of that voxel's time course in the remaining subjects
    - Perform Student *t*-test on the LOO ISC values

- **All these methods have poor FPR controllability**

ISC group analysis: exploration with new nonparametric approaches

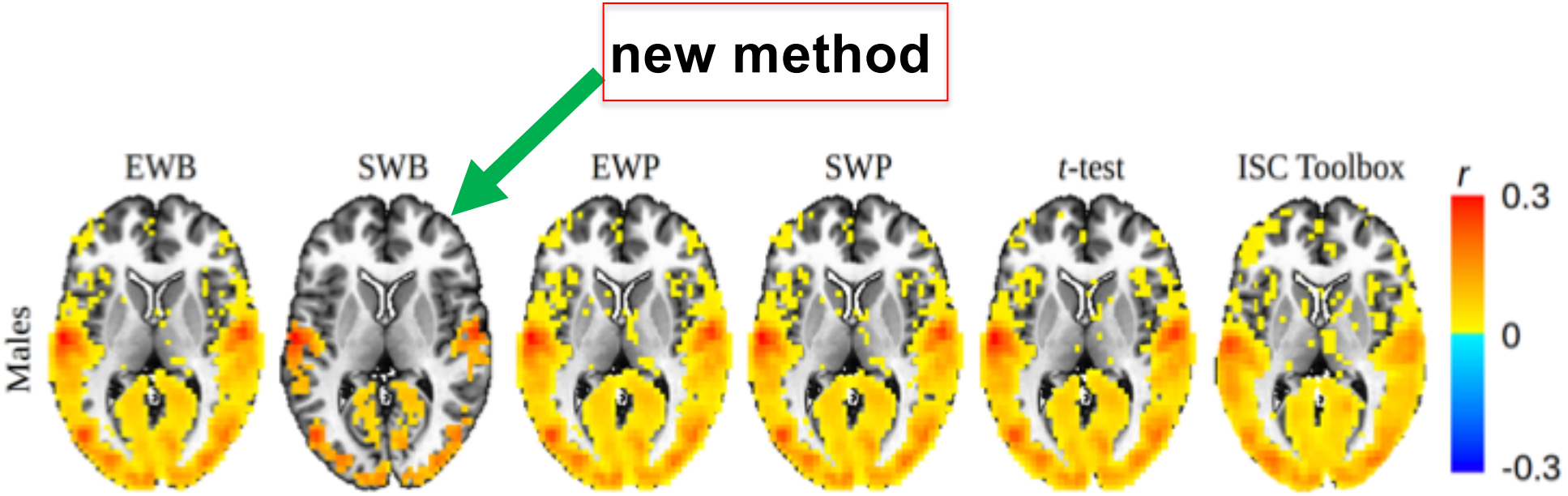
- Schematic demo of how different methods work



Chen et al, 2016a. Untangling the relatedness among correlations, part I: Nonparametric approaches to inter-subject correlation analysis at the group level. Neuroimage (in press).

New nonparametric approaches: real data

- One group: 24 male subjects
- 6 movie clips, 406 time points



- Similar results for group comparisons with **SWP**

- Results with real data are consistent with simulation results

Chen et al, 2016a. Untangling the relatedness among correlations, part I: Nonparametric approaches to inter-subject correlation analysis at the group level. Neuroimage (in press).

Linear mixed-effects modeling (LME)

- Modeling via effect partitioning: crossed random-effects LME

$$z_{ij} = b_0 + \theta_i + \theta_j + \epsilon_{ij}, \quad i \neq j$$

$$\theta_i, \theta_j \overset{iid}{\sim} G(0, \zeta^2) \text{ and } \epsilon_{ij} \overset{iid}{\sim} G(0, \eta^2)$$

cross-subject **within-subject**

- Characterizing the relatedness among ISCs via LME

$$\rho = Corr(z_{ij}, z_{jl}) = \frac{Cov(z_{ij}, z_{jl})}{\sqrt{Var(z_{ij})Var(z_{jl})}} = \frac{\zeta^2}{2\zeta^2 + \eta^2}$$

$$0 \leq \rho = \frac{\zeta^2}{2\zeta^2 + \eta^2} = \frac{\zeta^2}{\sigma^2} \leq 0.5$$

Chen et al, 2016b. Untangling the Relatedness among Correlations, Part II: Inter-Subject Correlation Group Analysis through Linear Mixed-Effects Modeling. Neuroimage (in press).

Linear mixed-effects modeling (LME)

- Formulation: crossed random-effects LME

$$z_{ij} = b_0 + \theta_i + \theta_j + \epsilon_{ij}, \quad i \neq j$$

$$\theta_i, \theta_j \overset{iid}{\sim} G(0, \zeta^2) \text{ and } \epsilon_{ij} \overset{iid}{\sim} G(0, \eta^2)$$

- Extendibility/flexibility of LME
  - Easy to incorporate explanatory variables: between- and within-subject factors (or quantitative covariates), similar to extension of *t*-test to GLM

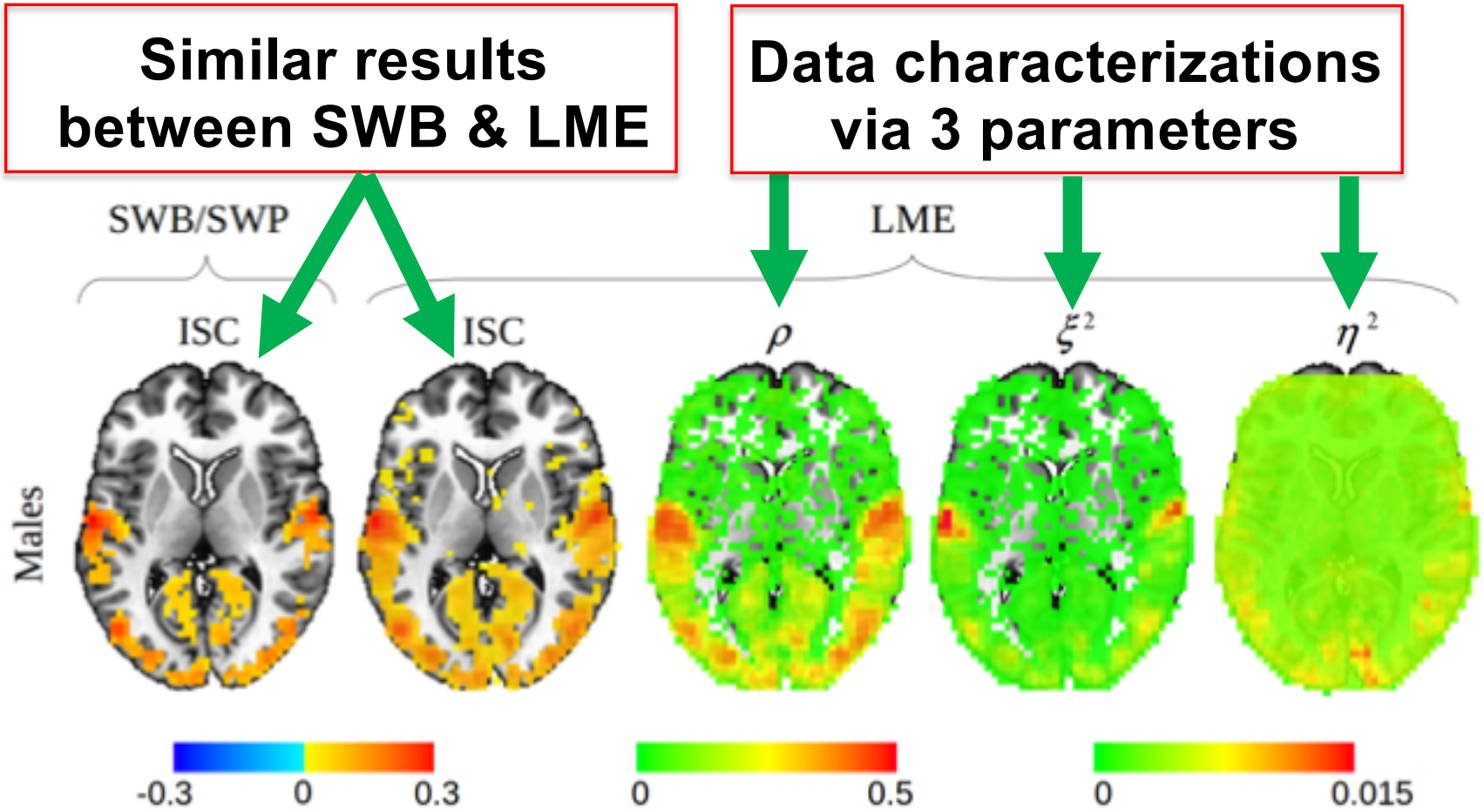
- Data characterization and model quality: unavailable for nonparametric approaches
  - Cross-subject variance ζ^2
  - Within-subject variance η^2
  - Relatedness of ISCs ρ

$$0 \leq \rho = \frac{\zeta^2}{2\zeta^2 + \eta^2} = \frac{\zeta^2}{\sigma^2} \leq 0.5$$

Chen et al, 2016b. Untangling the Relatedness among Correlations, Part II: Inter-Subject Correlation Group Analysis through Linear Mixed-Effects Modeling. Neuroimage (in press).

LME: real experiment data

- 48 subjects (24 males, 24 females)
- 6 movie clips, 406 time points



Chen et al, 2016b. Untangling the Relatedness among Correlations, Part II: Inter-Subject Correlation Group Analysis through Linear Mixed-Effects Modeling. Neuroimage (in press).

Benefits of naturalistic paradigm

- Similar to resting-state FMRI
- Extendable to other modalities
  - EEG, MEG, ECoG, fNIRS…
- No presumption about HDR function
- More controlled and engaging (especially for children)
- Practical benefit: subject less likely to fall asleep
- Analysis benefits
  - Less vulnerable to head motion effects
  - Statistically more powerful
  - Not dependent on seed selection (in seed-based approach)
  - Not dependent on dimension reduction and component selection
  - Well-fit by powerful LME with crossed random effects paradigm

Efficient Modeling through Information Pooling

Gang Chen

SSCC/NIMH

Preview

- Background and motivations
  - Typical whole brain group analysis
  - Cross-regional data analysis at the group level
    - Correlation matrix among ROIs
    - DTI data
- How can we do better?
  - Handling data analysis at ROI level
  - Pooling information across ROIs

- Applications
  - Group analysis at some ROIs instead of whole brain
  - Cross-regional analysis
  - DTI tractography: WM network
  - Naturalistic data analysis

Background and Motivation

- **Cluster failure (2016)**
  - Parametric modeling approaches: not stringent enough in cluster thresholding
    - Monte Carlo simulations: 3dClustSim (alphasim)
    - Random field theory (RFT)
  - Permutation: immune to the problem?
- Most people are feeling the stringency impact
  - Parametric methods
    - Uncorrected *p*: 0.001
  - Permutation: golden method?
- Cross-regional data analysis
  - Arbitrary thresholding: Garden of forking path
  - Arbitrary parameters

Something lost in the analogy

- **World is not always discrete**
  - Guilt vs. Innocent
  - Science: black or white?
  - Brain region: activated vs. inactivated?
- Does thresholding make sense?
  - To some extent: when signal is STRONG and noise low!
  - Policy making, drunk driving, emission test
  - Science: modeling results
    - *p*-value as a gatekeeper?
    - How to incorporate prior knowledge]?
      - Anatomical structure
      - Previous results

Problems with *p*-value

- **Straw man: null hypothesis – witch hunt**
  - H_0: scientifically uninteresting; unrealistic characterization
  - "False positive": misnomer
- Interpretation: conditional probability p(evidence | H_0)
  - Strong tendency to equating it to p(H_0 | data): **p(evidence | H_0) ≠ p(H0 | data)**!
- Abusive interpretation
  - Statistically insignificant = non-existing effect?
  - Set threshold (e.g., 0.001) and done
  - **Disillusion**: higher significance → more confidence!
- Threshold: dichotomized decision
  - *p*-value of 0.05 vs 0.051, or cluster size of 54 vs 53 voxels
  - Sidedness: one- or two-sided?
  - Difference btw a "significant" result and an "insignificant" result
  - Selection bias in effect estimates
  - Power analysis: not useful
  - A source of reproducibility problem
  - How reliable is meta analysis?

Problems with p-value

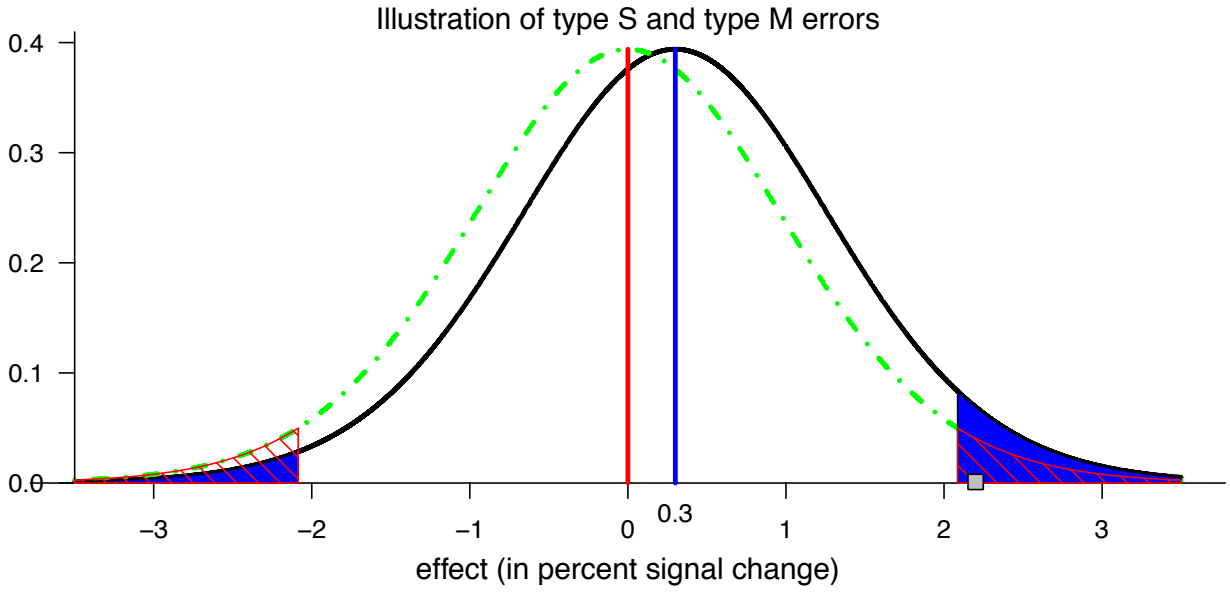
- Low hanging fruits almost gone
  - Large effects
  - Difficult to winnow out smaller effects from noise
- Cluster threshold – iceberg approach
  - Using spatial extent as a leverage to counter false positives
  - Penalizing anatomically small regions: **discrimination!**
    - 2 regions with same signal strength: 1 large and 1 small size
    - 2 regions with same signal strength: 1 case (distant) and 1 case (contiguous)
- Objective? Easy for journals/reviewers to make decisions?
  - *p*-hacking (even encouraged): one-tailed; small volume correction
- Is false positive rate the only issue we should care about?
  - False negatives
  - False sign (type S)
  - False magnitude (type M)

Problems with p-value

- Type S and Type M errors: $t(20)$

Table 2: Power, type S and type M errors estimated from simulations[a]

se / *ef*	0.1			0.3			0.5			0.7			1.0		
	pwr	*S*	*M*	*pwr*	*S*	*M*	*pwr*	*S*	*M*	*pwr*	*S*	*M*	*pwr*	*S*	*M*
0.1	0.15	0.02	2.66	0.06	0.21	7.58	0.05	0.31	12.86	0.05	0.36	17.71	0.05	0.40	25.55
0.3	0.81	0.00	1.12	0.15	0.02	2.66	0.08	0.09	4.28	0.07	0.15	5.96	0.06	0.23	8.59
0.5	1.00	0.00	1.00	0.34	0.00	1.67	0.15	0.02	2.66	0.10	0.06	3.66	0.07	0.12	5.16
0.7	1.00	0.00	1.00	0.60	0.00	1.28	0.25	0.00	1.96	0.15	0.02	2.67	0.10	0.06	3.74
1.0	1.00	0.00	1.00	0.89	0.00	1.07	0.47	0.00	1.44	0.26	0.00	1.91	0.15	0.02	2.65



Alternative approaches?

- Abandoning NHST?
  - Do not care about H_0
  - No more *p*-value
- Totality of results instead of "significant" part only
  - Colored clusters with *t*-value bar?
- Soften harsh thresholding
  - Show unfortunate clusters with supporting information
- Different modeling approaches
  - Multilevel modeling: pooling information
  - Bayesian modeling: numerical perspective

Bayesian approaches

Table 7: Interpretation differences between NHST and Bayesian framework

	Probability p	**Effect Interval** $[L, U]$
NHST	If H_0 is true, the probability of having the current result or more extreme is p (based on what would have occurred under other possible datasets); e.g., $P(\|T(\boldsymbol{y})\| > t_c \| \text{happy} = \text{sad}) = p$, where $T(\boldsymbol{y})$ is a statistic (e.g., Student's t) based on data $\boldsymbol{y}$ and t_c is a threshold.	If the study is exactly repeated an infinite number of times, the percentage of those confidence intervals will cover the true effect is $1 - p$; e.g., $P(L \leq \text{happy} - \text{sad} \leq U) = 1 - p$, where "happy - sad" is treated as being fixed while L and U are random.
Bayesian	The probability of having the *current* result being different from zero is p (given the dataset); e.g., $P(\text{happy} - \text{sad} < L \text{ or happy} - \text{sad} > U \| \boldsymbol{y}) = p$, where L and U are lower and upper bounds of the $(1 - p)100\%$ quantile interval.	The probability that the effect falls in the predictive interval is $1 - p$ (given the data); e.g., $P(L \leq \text{happy} - \text{sad} \leq U \| \boldsymbol{y}) = 1 - p$, where "happy - sad" is considered random while L and U are known conditional on data $\boldsymbol{y}$.

ROI-based approach

- Major source for multiple comparisons: segmented modeling
  - Massively univariate model: analyze each voxel **separately**
    - FDR: too conservative
    - FWE (multiple comparisons - Monte Carlo simulations, random field theory, permutations): **biased** on the large regions and **unfair to or penalizing / discriminating** small regions; conservative
    - Prior information not considered: anatomy, previous results
- Analysis on a list of regions
  - Predefined ROIs: atlas, parcellation, data partitioning
  - Ideally independent
- A few possible applications
  - ROI-based group analysis
  - Cross-regional analysis: connectivity, path, edge, hub, …
  - DTI
  - ROI-based inter-subject correlation

First application: ROI-Based Group Analysis

- Conventional approach: univariate GLM
  - Vulnerable to multiple testing issue
  - Leading to very conservative inferences
- BML
  - Don't treat ROIs as unrelated entities
  - Commonality: similar scaling and range
- Dataset
  - Subjects: n = 124; Resting-state data
  - Seed-based correlation for each of 124 subjects
  - Group analysis: effect of behavioral measure (x) on seed-based correction (y: z-score)
  - $y = a + b\,x + e$

First application: ROI-Based Group Analysis

- Conventional approach: whole brain analysis
- Statisticians demand that **correction for FWE be warranted**!

Table 3: ROIs and FWE correction for their associated clusters[a]

voxel-wise p	cluster threshold	number of surviving ROIs	ROIs
0.001	28	2	R PCC, PCC/PrC
0.005	66	4	R PCC, PCC/PrC., L IPL, L TPJ
0.01	106	4	R PCC, PCC/PrC., L IPL, L TPJ
0.05	467	4	R PCC, PCC/PrC., L IPL, L TPJ
0.05*	467	(4)	(L aMTS/aMTG, R TPJp, vmPFC, dmPFC)

[a]Monte Carlo simulations were conducted using spatial autocorrelation function (Cox et al., 2017) instead of FWHM to determine the cluster threshold (voxel size: $3 \times 3 \times 3$ mm^3). The ROI abreviations are listed in Table 4.
*Special note for the last row (voxel-wise p-value of 0.05): four ROIs including L IPL, L TPJ, R PCC, PCC/PrC survived together with their clusters from the FWE correction, and the other four ROIs listed here (L aMTS/aMTG, R TPJp, vmPFC, and dmPFC) did not survive with their clusters but showed some evidence of effect when the cluster size requirement was dropped.

- **Inefficient modeling!**

First application: ROI-Based Group Analysis

- Conventional approach: univariate GLM – totally ***r* models**!

  $y_{i1} = a_1 + b_1 x_i + e_{i1}$, i = 1,2,…, *n*

  $y_{i2} = a_2 + b_2 x_i + e_{i2}$, i = 1,2,…, *n*

  …

  $y_{ir} = a_r + b_r x_i + e_{ir}$, i = 1,2,…, *n*

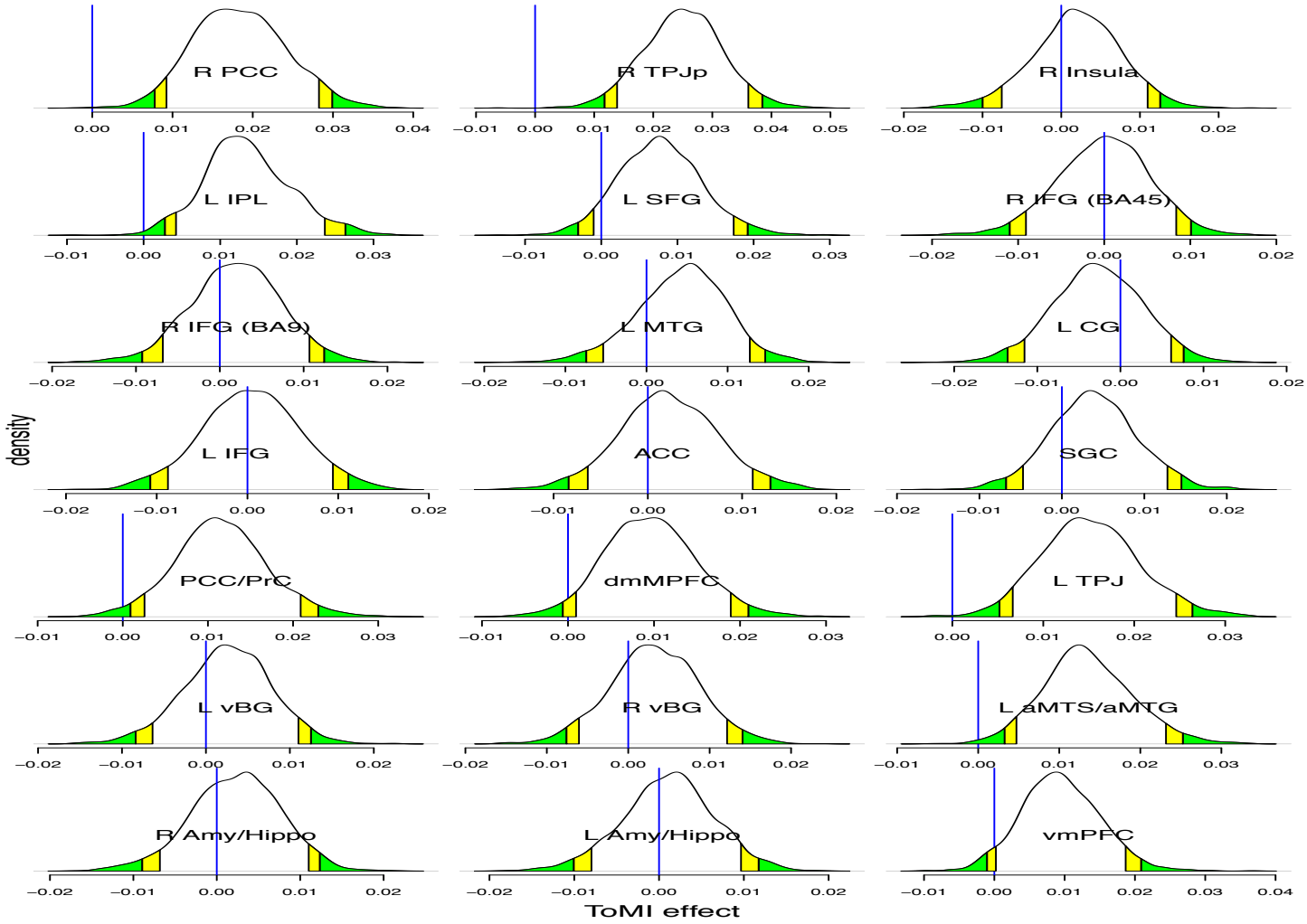
  - Statisticians demand that **correction for FWE be warranted**!
    - ↙ Bonferroni
    - ↙ Nobody conducts ROI-based analysis!

- Combining the data from *r* ROIs to one model (pooling variances)
  - $y_{ij} = a_j + b_j x_i + e_{ij}$, i = 1,2,…, *n*, j = 1, 2, …, *r*
  - a_j and b_j are constants, free to vary from $-\infty$ to $+\infty$, $e_{ij} \sim N(0, \sigma^2)$
  - Fully trust the data
  - **Inefficient modeling!**
  - Can we do better than this?

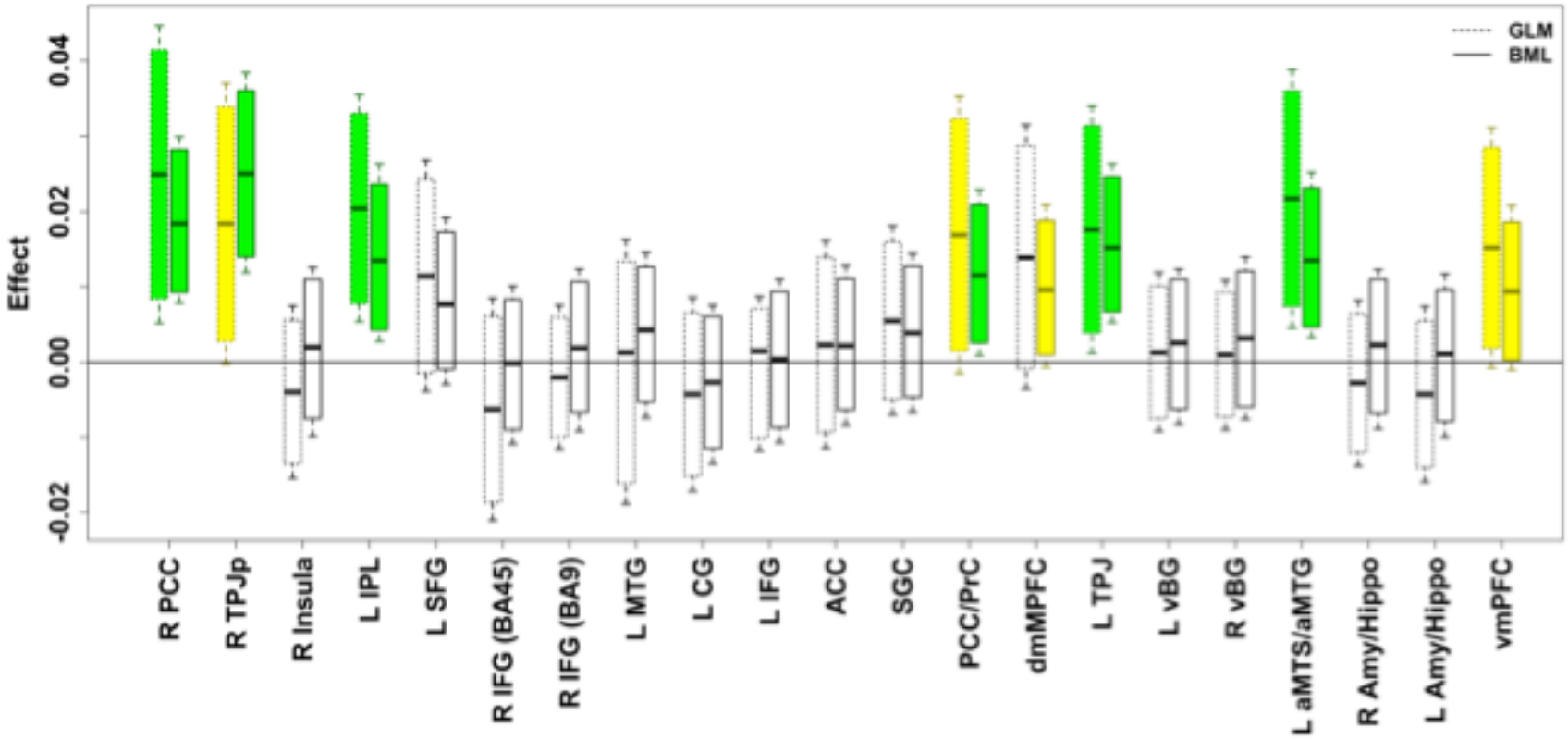
First application: ROI-Based Group Analysis

- Partial pooling
  - $y_{ij} = a_j + b_j x_i + \xi_i + e_{ij}$, i = 1,2,…, *n*, j = 1, 2, …, *r*
  - a_j and b_j are random: loosely bound with $a_j \sim N(a, \lambda_1^2)$, $b_j \sim N(b, \lambda_2^2)$,
  - No more far-fetched than assumption about subjects $\xi_i \sim N(0, \tau^2)$ and residuals $e_{ij} \sim N(0, \sigma^2)$
  - Loose harness: not fully trust the data
  - Further defense
    - Effects at brain regions share similar scale and range
- Why better?
  - One (instead of *r*) model!
  - No need to correct for multiple testing
  - Sharing information among the ROIs: partial pooling or shrinkage
  - Do not trust individual effects (shrinking the effects to the center)
  - Controlled type S and type M errors

First application: ROI-Based Group Analysis



First application: ROI-Based Group Analysis



First application: ROI-Based Group Analysis

Table 6: Comparison of results between the conventional approach with no pooling and BHM[a]

result / ROI	ToMI effect		standard error		2.5%		5%		95%		97.5%	
	GLM	BHM	GLM	BHM	GLM	BHM	GLM	BHM	GLM	BHM	GLM	BHM
R PCC	0.025	0.018	0.010	0.006	0.005	0.008	0.008	0.009	0.041	0.028	0.045	0.030
R TPJp	0.018	0.025	0.009	0.007	0.000	0.012	0.003	0.014	0.034	0.036	0.037	0.038
R Insula	-0.004	0.002	0.006	0.006	-0.015	-0.010	-0.014	-0.008	0.006	0.011	0.007	0.013
L IPL	0.020	0.014	0.008	0.006	0.005	0.003	0.008	0.004	0.033	0.024	0.035	0.026
L SFG	0.011	0.008	0.008	0.006	-0.004	-0.003	-0.001	-0.001	0.024	0.017	0.027	0.019
R IFG (BA45)	-0.006	0.000	0.007	0.005	-0.021	-0.011	-0.019	-0.009	0.006	0.008	0.008	0.010
R IFG (BA9)	-0.002	0.002	0.005	0.005	-0.012	-0.009	-0.010	-0.007	0.006	0.011	0.008	0.012
L MTG	-0.001	0.004	0.009	0.005	-0.019	-0.007	-0.016	-0.005	0.013	0.013	0.016	0.015
L CG	-0.004	-0.003	0.007	0.005	-0.017	-0.014	-0.015	-0.011	0.007	0.006	0.009	0.008
L IFG	-0.002	0.000	0.005	0.005	-0.012	-0.011	-0.010	-0.009	0.007	0.009	0.009	0.011
ACC	0.002	0.002	0.007	0.005	-0.012	-0.008	-0.009	-0.006	0.014	0.011	0.016	0.013
SGC	0.006	0.004	0.006	0.005	-0.007	-0.007	-0.005	-0.005	0.016	0.013	0.018	0.014
PCC/PrC	0.017	0.012	0.009	0.005	-0.001	0.001	0.002	0.003	0.032	0.021	0.035	0.023
dmMPFC	0.014	0.010	0.009	0.005	-0.004	-0.001	-0.001	0.001	0.029	0.019	0.032	0.021
L TPJ	0.018	0.015	0.008	0.005	0.001	0.005	0.004	0.007	0.031	0.025	0.034	0.026
L vBG	0.001	0.003	0.005	0.005	-0.009	-0.008	-0.007	-0.006	0.010	0.011	0.012	0.012
R vBG	0.001	0.003	0.005	0.005	-0.009	-0.008	-0.007	-0.006	0.009	0.012	0.011	0.014
L aMTS/aMTG	0.022	0.013	0.009	0.006	0.005	0.003	0.007	0.005	0.036	0.023	0.039	0.025
R Amy/Hippo	-0.003	0.002	0.006	0.005	-0.014	-0.009	-0.012	-0.007	0.006	0.011	0.008	0.012
L Amy/Hippo	-0.004	0.001	0.006	0.005	-0.016	-0.010	-0.014	-0.008	0.005	0.010	0.007	0.012
vmPFC	0.015	0.009	0.008	0.006	-0.001	-0.001	0.002	0.000	0.029	0.019	0.031	0.021

First application: ROI-Based Group Analysis

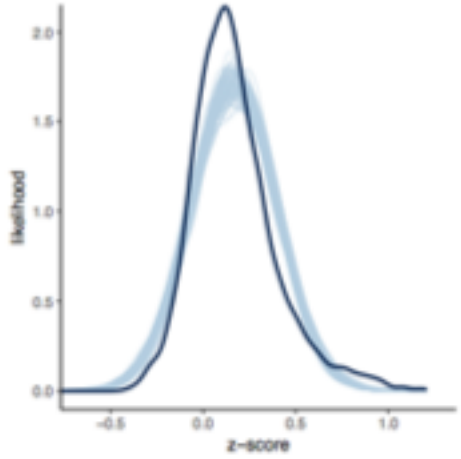
Unique: subject-level inferences

	Estimate	Est.Error	2.5%ile	5%ile	50%ile	95%ile	97.5%ile
HMN001	0.224	0.047	0.135	0.148	0.224	0.298	0.316
HMN002	0.218	0.046	0.128	0.144	0.219	0.293	0.311
HMN003	0.168	0.046	0.078	0.096	0.168	0.246	0.260
HMN004	0.080	0.047	-0.017	0.001	0.079	0.157	0.172
HMN005	0.200	0.048	0.104	0.121	0.203	0.275	0.287
HMN006	0.197	0.047	0.104	0.122	0.195	0.276	0.288
HMN007	0.155	0.047	0.060	0.074	0.156	0.232	0.246
HMN008	0.038	0.048	-0.057	-0.040	0.039	0.115	0.133
HMN011	0.169	0.046	0.078	0.094	0.169	0.246	0.259
HMN012	0.157	0.048	0.061	0.079	0.157	0.235	0.249
HMN013	0.197	0.048	0.101	0.119	0.197	0.274	0.290
HMN014	0.252	0.047	0.155	0.174	0.253	0.327	0.342
...							

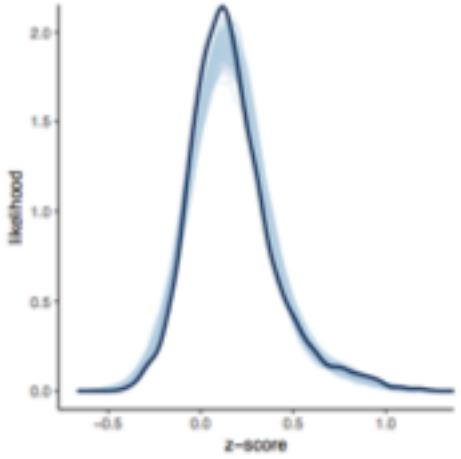
First application: ROI-Based Group Analysis

Cross-validation

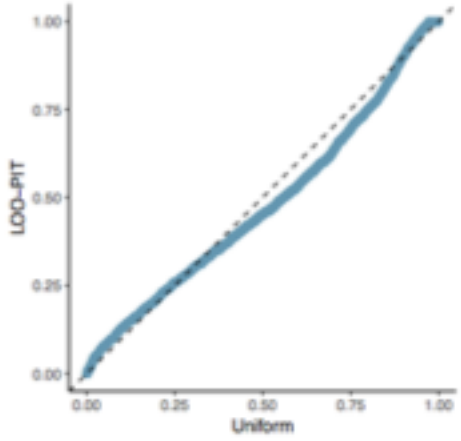
	LOOIC	SE
GLM	-300.39	98.25
BML	-2247.06	86.42
GLM - BML	1946.67	96.35



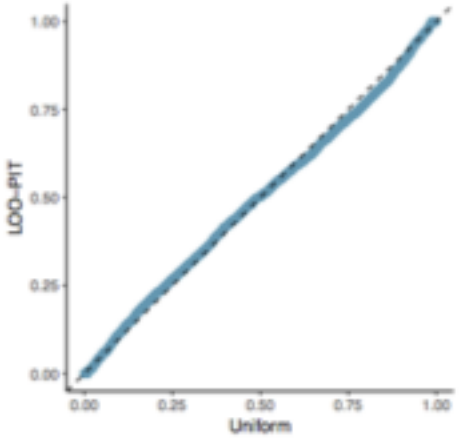
(a) GLM posterior predictive density



(b) BHM posterior predictive density



(c) GLM cross-validation: Q-Q plot (uniform)



(d) BHM cross-validation: Q-Q plot (uniform)

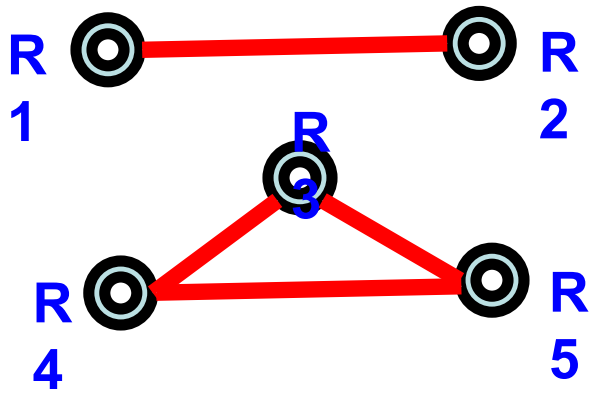
Second Application: cross-regional analysis

- Cross-regional analysis (e.g. graph theory)
  - Garden of forking path problem
  - Hard thresholding on correlation or edge density
  - What's the justification and underlying mechanism?
  - Does nature rigorously follow such thresholding?
  - How precise are those correlations and density measures?
  - How sensitive and consequential is the thresholding for such steps and final results?

Second Application: cross-regional analysis

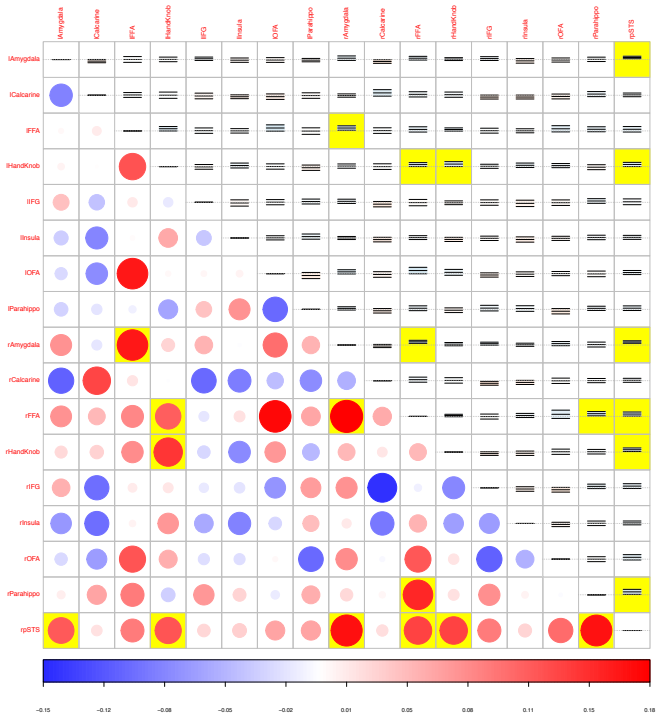
- Face recognition tasks
  - 17 subjects
  - 17 ROIs
  - 17×17 correlation matrix per subject: 136 values per subject

- Conventional approach
  - 136 Student's *t*-tests
  - Multiple comparisons
    - ↙ **Network Based Statistics**

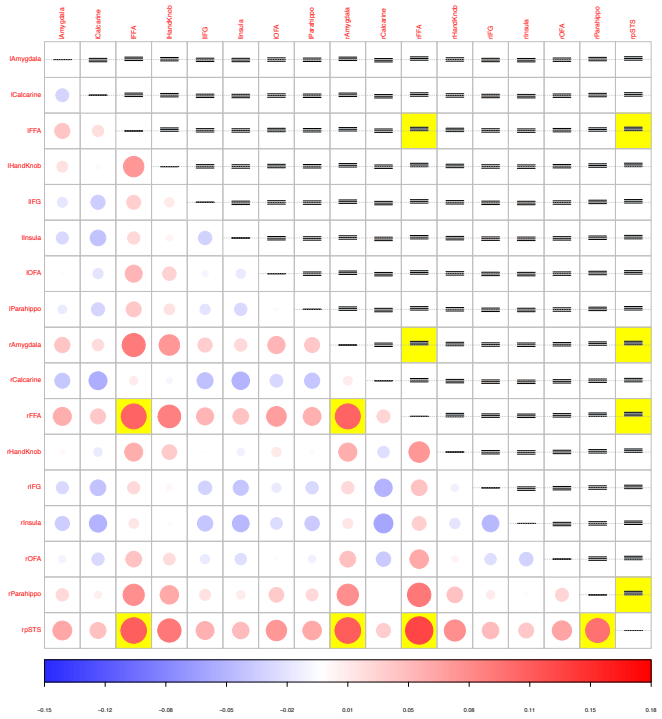


BML: cross-regional analysis

- T-tests vs. New



136 t-tests



New

BML: cross-regional analysis

Unique: region-level inferences

	Estimate	Est.Error	2.5%ile	5%ile	50%ile	95%ile	97.5%ile
lAmygdala	-0.053	0.049	-0.154	-0.133	-0.054	0.028	0.046
lFFA	-0.092	0.048	-0.191	-0.174	-0.090	-0.016	0.000
lOFA	-0.073	0.048	-0.174	-0.156	-0.074	0.005	0.024
rAmygdala	-0.092	0.048	-0.194	-0.171	-0.091	-0.014	0.002
rFFA	-0.099	0.048	-0.200	-0.179	-0.098	-0.020	-0.006
rHandKnob	-0.074	0.048	-0.174	-0.153	-0.074	0.004	0.019
rOFA	-0.068	0.048	-0.161	-0.148	-0.067	0.013	0.028
rpSTS	-0.098	0.048	-0.198	-0.180	-0.097	-0.019	-0.007
lHandKnob	-0.075	0.048	-0.171	-0.154	-0.075	0.004	0.021

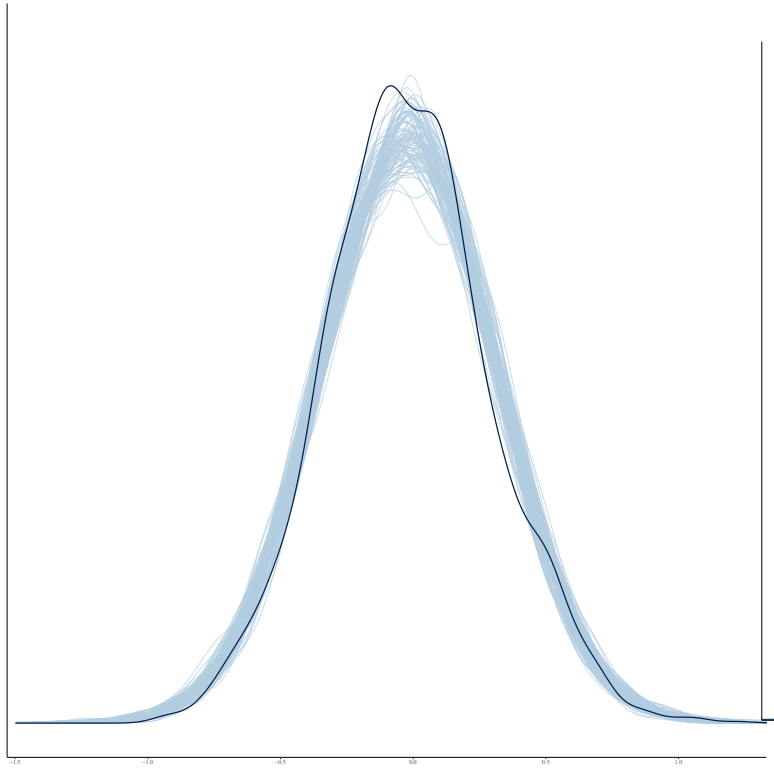
BML: cross-regional analysis

Unique: subject-level inferences

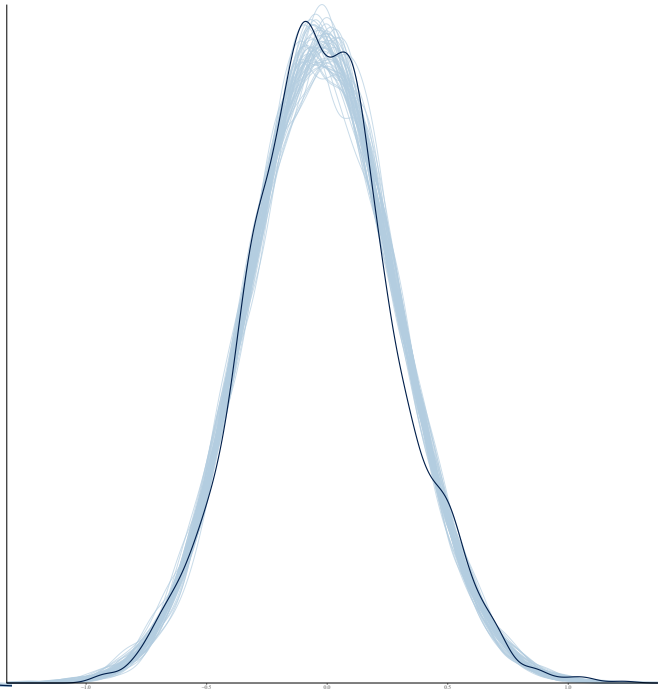
	Estimate	Est.Error	2.5%ile	5%ile	50%ile	95%ile	97.5%ile
SBJ01	-0.17	0.05	-0.27	-0.25	-0.18	-0.09	-0.07
SBJ02	-0.16	0.05	-0.26	-0.24	-0.16	-0.09	-0.07
SBJ03	0.23	0.05	0.13	0.15	0.23	0.31	0.32
SBJ04	0.09	0.05	-0.01	0.01	0.09	0.17	0.19
SBJ05	-0.14	0.05	-0.23	-0.22	-0.14	-0.06	-0.04
SBJ06	-0.01	0.05	-0.10	-0.09	-0.01	0.07	0.09
SBJ07	-0.22	0.05	-0.32	-0.30	-0.22	-0.14	-0.13
SBJ08	-0.08	0.05	-0.18	-0.16	-0.08	0.00	0.02
SBJ09	-0.06	0.05	-0.15	-0.13	-0.06	0.02	0.04
SBJ10	-0.14	0.05	-0.24	-0.22	-0.14	-0.06	-0.04
SBJ11	0.10	0.05	0.00	0.02	0.10	0.17	0.19
SBJ12	-0.21	0.05	-0.30	-0.29	-0.21	-0.12	-0.11
SBJ13	-0.07	0.05	-0.17	-0.15	-0.07	0.00	0.02
SBJ14	-0.20	0.05	-0.30	-0.28	-0.20	-0.12	-0.10
SBJ15	0.13	0.05	0.03	0.05	0.13	0.21	0.22
SBJ16	-0.16	0.05	-0.26	-0.24	-0.16	-0.08	-0.06
SBJ17	-0.31	0.05	-0.40	-0.39	-0.31	-0.23	-0.22

BML: cross-regional analysis

- Cross-validation: *t*-tests vs. BML



136 t-tests



New

Limitations

- Runtime: minutes – days
- ROI definition
- Whole brain analysis?
  - Voxel-wise
  - ROIs covering most or all regions
- Data complexity: t-test, ANOVA, GLM?
  - Models under NHST
  - Data structure
- Exchangeability violation
  - Theoretical convenience
  - Practical perspectives: validations
- Prior subjectivity
  - Priors
  - Hyperpriors
    - Regularization: Ridge regression, LASSO
    - Numerical convenience

A few words about modeling

- Geocentric or Ptolemaic model
  - Solar system and heavenly sphere
  - A refinement of previous models developed by Greek astronomers
  - Ptolemy's model could so accurately explain the motions of solar planets
  - A model for understanding the structure of the solar system

- Statistical models are like geocentricity
  - "All models are wrong, but some of them are useful" (G.E.P. Box)
  - Data generative process or constructs that approximates the reality
    - Linear models: linearity rarely holds!
  - Knowledge updating/evolving
    - Model checking, cross-validation

Overview

- Background and motivations
  - Typical whole brain group analysis
  - Cross-regional data analysis at the group level
    - Correlation matrix among ROIs
    - DTI data
- How can we do better?
  - Handling data analysis at ROI level
  - Pooling information across ROIs

- Applications
  - Group analysis at some ROIs instead of whole brain
  - Cross-regional analysis
  - DTI tractography: WM network
  - Naturalistic data analysis