Surface-based Group Analysis in FreeSurfer

Outline

- Objectives & Example
- GLM Theory & Linear Algebra Review
- Command-line Stream
  - Assemble Data
 - Design/Contrast
  - Analyze
  - Visualize

Group Analysis Objective

- To create a model that can describe patterns of interactions and associations

- The **parameters** of the model provide measures of the strength of associations

- A General Linear Model (GLM) focuses on *estimating the parameters of the model* such that they can be applied to new data sets to create reasonable inferences.

Types of Questions

- Does a specific variable have a significant association with an outcome?

- If we control for the effects of a second variable, is the association still significant?

- Is there a group difference in outcome?

- Does a specific variable affect individual outcome differently between groups of individuals?
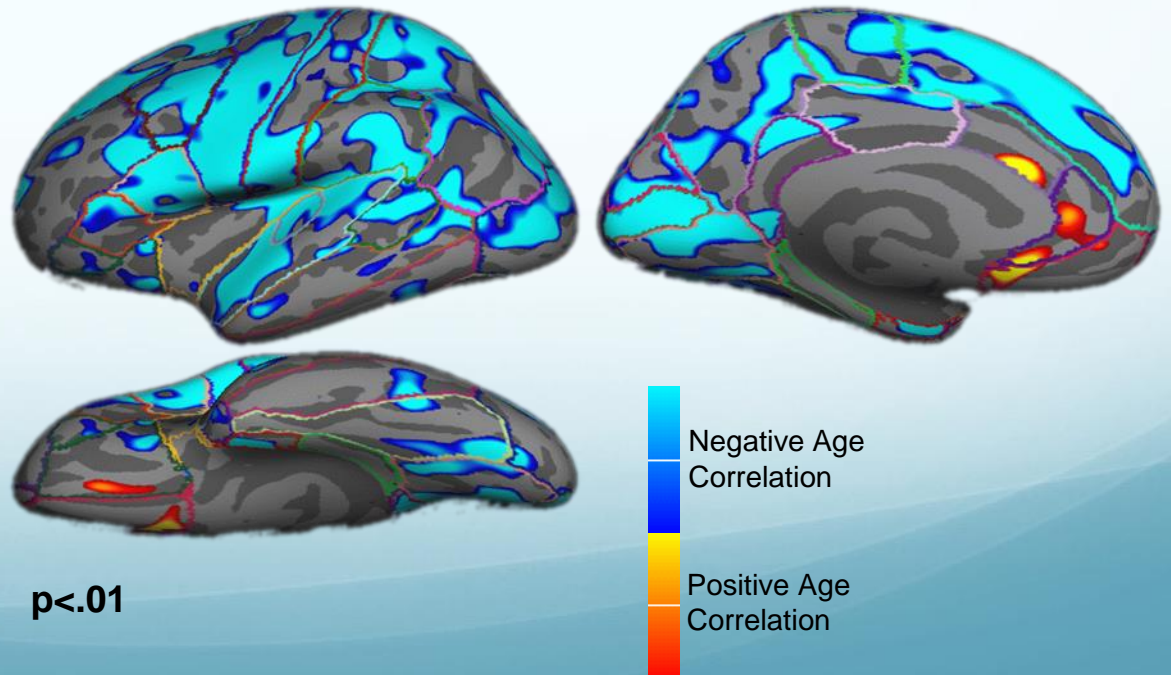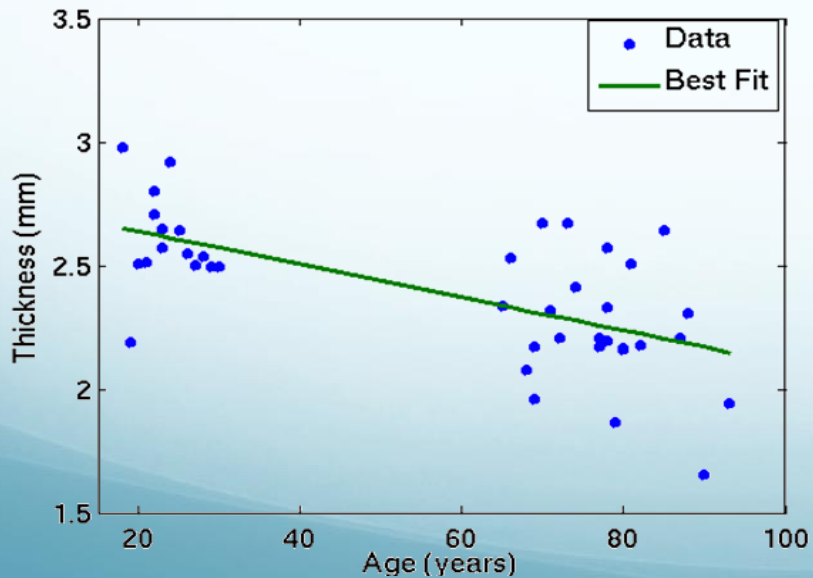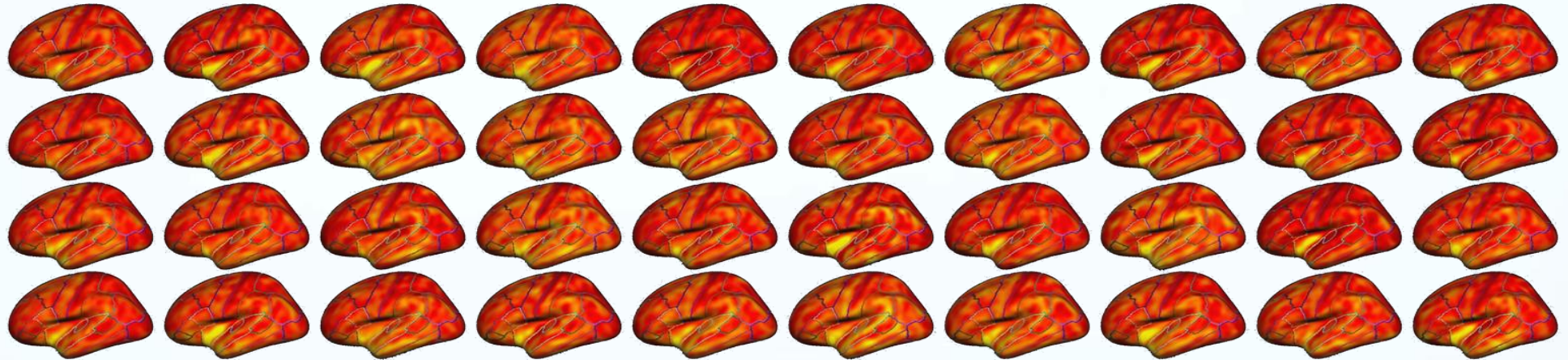
Aging Exploratory Analysis
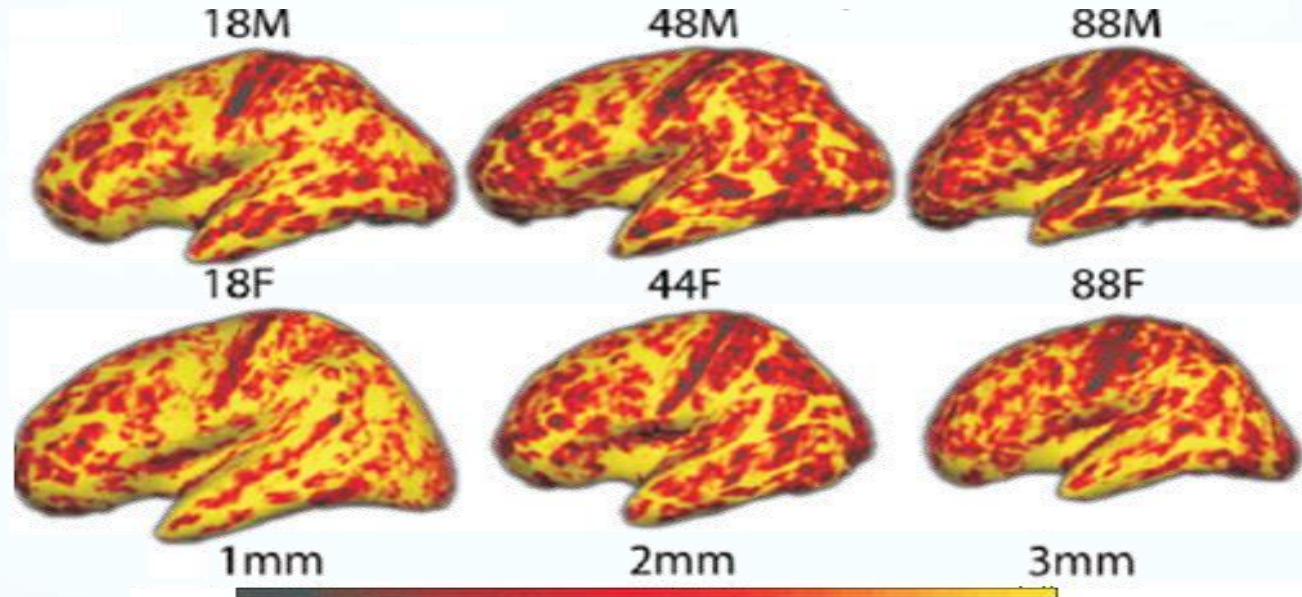
In which areas does thickness
Change with age?

Cortical Thickness vs Aging
Salat et al, 2004, Cerebral Cortex

Aging Thickness Study

N=40 (all in fsaverage space)

p<.01

Negative Age Correlation

Positive Age Correlation

Surface-based Measures

- Morphometric (e.g., thickness)

- Functional

- PET

- MEG/EEG

- Diffusion (?) sampled just under the surface
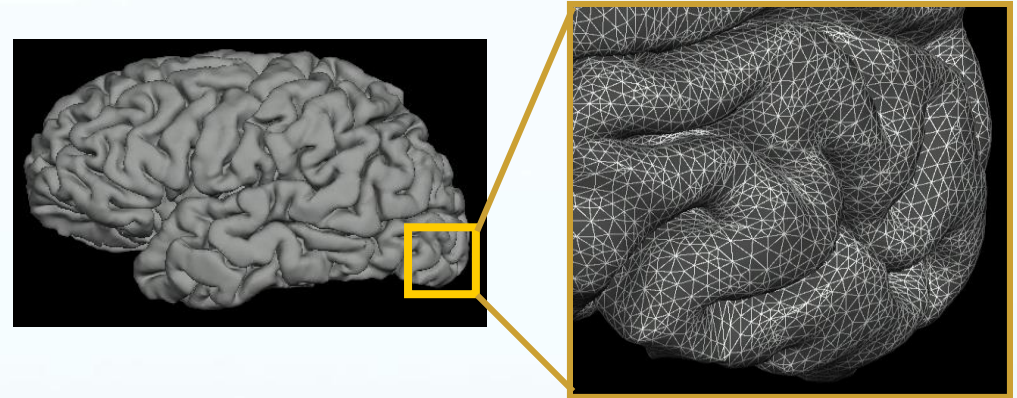
The General Linear Model (GLM)

GLM Theory

Is Thickness correlated with Age?

Dependent Variable, Measurement

Thickness
IQ, Height, Weight, etc.

Thickness

Subject 1

Subject 2

y1

y2

x1   x2

Age
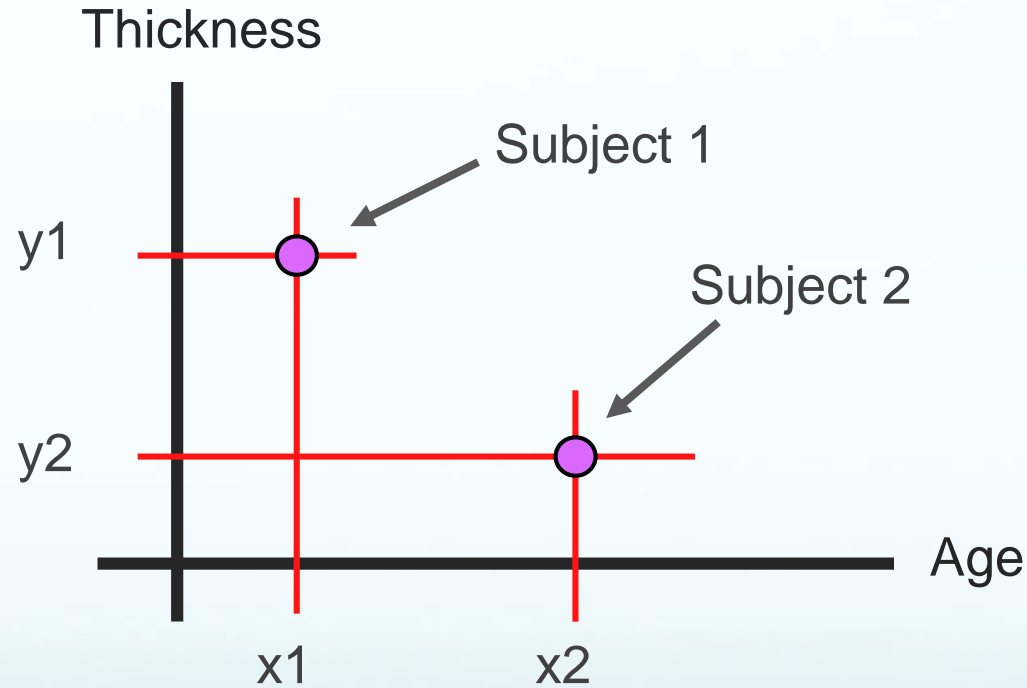
Of course, you would need more then two subjects …

Independent Variable

Linear Algebra Review (stay calm…)

Dependent variable (outcome) → $y = mx + b$ ← **Intercept**

Slope ↗ m — **Independent variable**
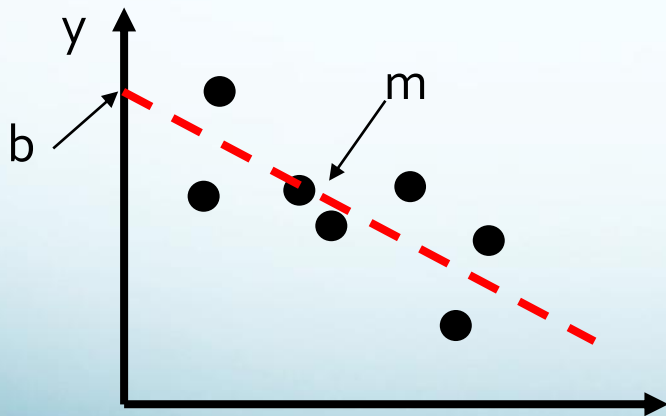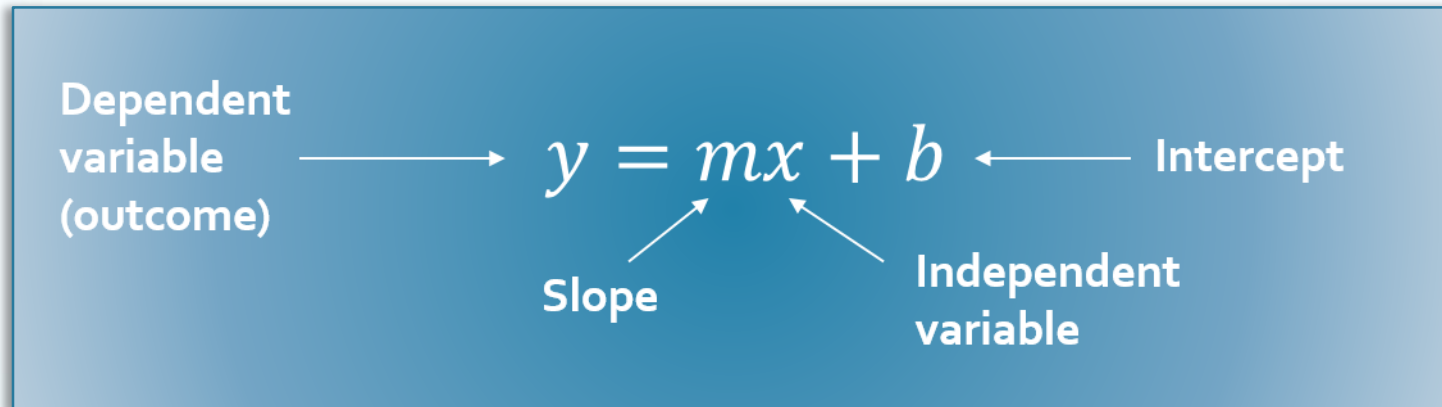
Linear Algebra Review (stay calm...)

We can put this in matrix format:

$$y = mx + b$$
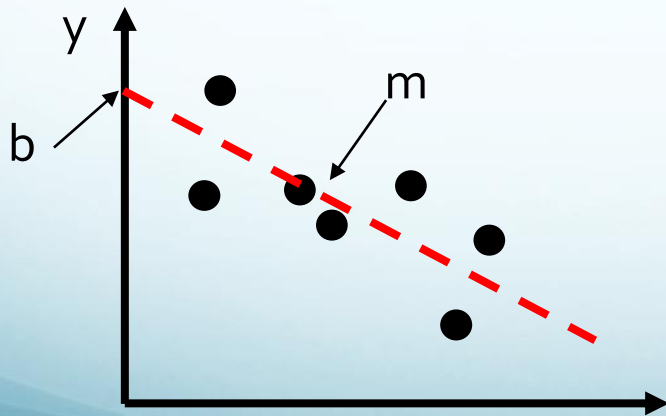
$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix} * \begin{bmatrix} b \\ m \end{bmatrix}$$

Design Matrix   **Regression Coefficients (parameters)**

-One row per data point
-Add column of 1's for the offset term (*b*)
-One set of parameters

Matrix Multiplication

$$\begin{bmatrix} y1 \\ y2 \\ y3 \\ y4 \end{bmatrix} = \begin{bmatrix} 1 & x1 \\ 1 & x2 \\ 1 & x3 \\ 1 & x4 \end{bmatrix} * \begin{bmatrix} b \\ m \end{bmatrix}$$

$y1 = 1*b + x1*m$

$y2 = 1*b + x2*m$

$y3 = 1*b + x3*m$

$y4 = 1*b + x4*m$

System of Linear Equations

Linear Model

Thickness

Intercept: b
(=Offset)

Slope: m

y1

y2

Age

x1   x2

X = Design Matrix
b = Regression Coefficients
  = Parameter estimates
  = "betas"
  = beta.mgh (mri_glmfit output)

System of Linear Equations

$$y1 = 1*b + x1*m$$
$$y2 = 1*b + x2*m$$

Matrix Formulation

$$\begin{bmatrix} y1 \\ y2 \end{bmatrix} = \begin{bmatrix} 1 \; x1 \\ 1 \; x2 \end{bmatrix} * \begin{bmatrix} b \\ m \end{bmatrix} \Bigg\}$$ Two parameters

-One row per subject
-x values are independent variable (age)
-Column of 1's is the 'offset' term (to multiply by b)

$$\boxed{Y = X*b}$$ $$b = \begin{bmatrix} b \\ m \end{bmatrix}$$

13

Error

BUT… if we have the same *m* and *b* for all data points, we will have errors:
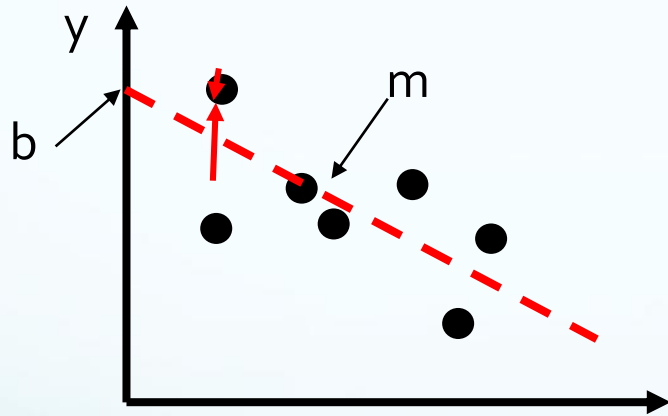
GOAL: minimize the sum of the square of error terms when estimating our *m* and *b* terms
There are lots of ways to do this! (Beyond the scope of this talk, but FreeSurfer does it for you!)

More than Two Data Points

Thickness

Intercept: b

Slope: m

Age

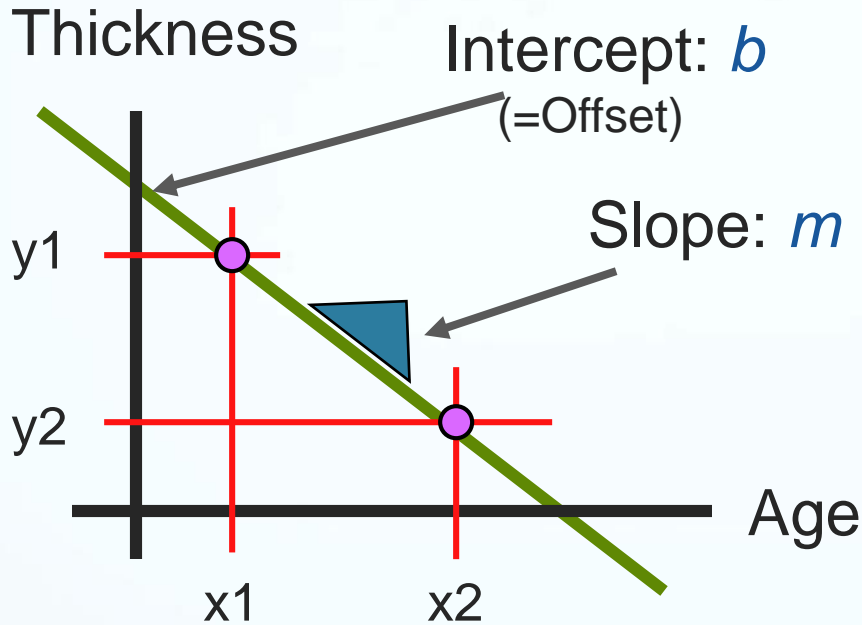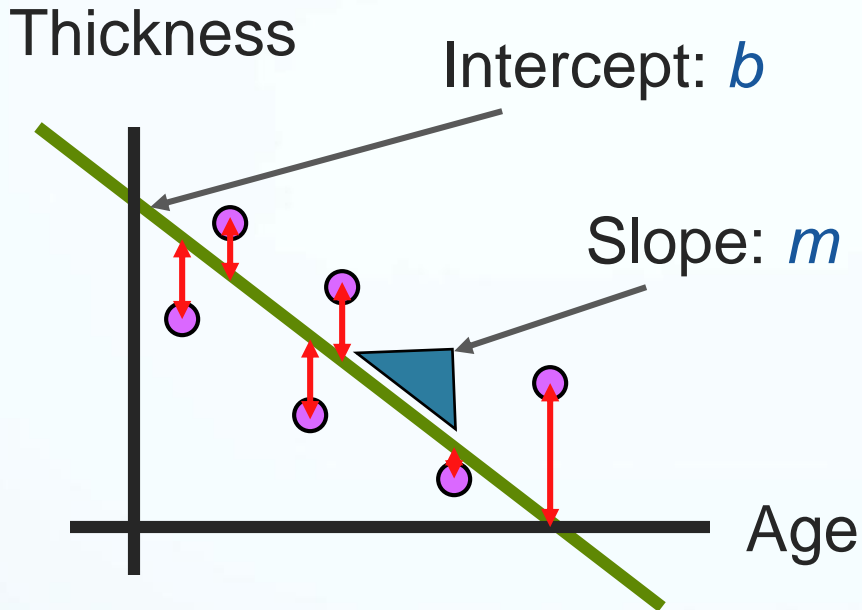$$\begin{bmatrix} y1 \\ y2 \\ y3 \\ y4 \end{bmatrix} = \begin{bmatrix} 1 & x1 \\ 1 & x2 \\ 1 & x3 \\ 1 & x4 \end{bmatrix} * \begin{bmatrix} b \\ m \end{bmatrix} + \begin{bmatrix} n1 \\ n2 \\ n3 \\ n4 \end{bmatrix}$$

$$Y = X*b+n$$

$y1 = 1*b + x1*m + n1$
$y2 = 1*b + x2*m + n2$
$y3 = 1*b + x3*m + n3$
$y4 = 1*b + x4*m + n4$

- Model Error
- Noise
- Residuals
- eres.mgh

Forming a Hypothesis

- Now, we can fit our parameters, but we need a hypothesis

- Our example: Is there a significant association between age and thickness?

- **Formal Hypothesis: The slope of age v. thickness (m) is significantly different from zero**
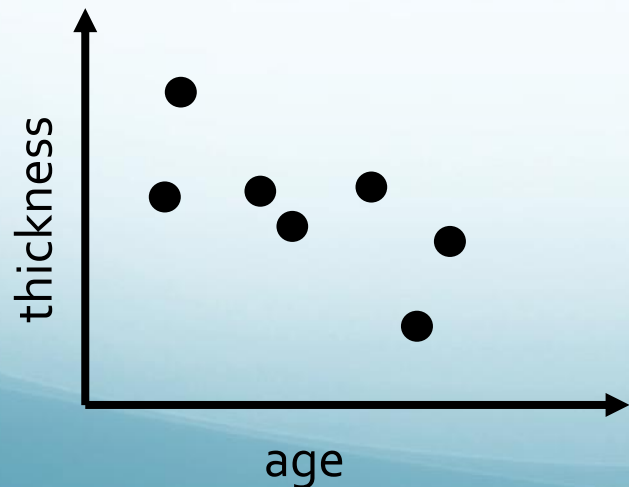
Null hypothesis: m = 0

Testing Our Hypothesis

- Once we fit our model for the optimal regression coefficients (m and b), we need to **test them for significance** as well as test the **direction of the effect**

- We do this by forming something called a **contrast matrix** that isolates our parameter of interest

- We can multiply our contrast matrix by our regression coefficient matrix to compute a variable **g,** which tells us the *direction of our effect*

- In this example, since our hypothesis is about the slope m we will design our contrast matrix to be **[0 1]**

$$\begin{pmatrix} \text{thickness1} \\ \text{thickness2} \\ \text{thickness3} \\ \text{thickness4} \end{pmatrix} = \begin{pmatrix} 1 & \text{age1} \\ 1 & \text{age2} \\ 1 & \text{age3} \\ 1 & \text{age4} \end{pmatrix} * \begin{pmatrix} b \\ m \end{pmatrix}$$

$$g = [0 \quad 1] * \begin{pmatrix} b \\ m \end{pmatrix}$$

If g is negative, then the direction of our effect (slope) is also negative

Testing our Hypothesis

- We still need to test for **_significance_**

- We'll use our **contrast matrix [0 1]** again here in a _t-test:_

Contrast matrix

Regression coefficients

$$t = \frac{C * \beta}{\sqrt{\sigma^2 C * (X^T X)^{-1} C^T}}$$

Design matrix

This t-value corresponds to a **p-value** that depends on your sample size. This **p-value** is between 0 and 1, values closer to 0 indicate a more significant result.

p-values

<u>p-value/significance</u>

• value between 0 and 1

• depends on your sample size

• closer to 0 means more significant

<u>FreeSurfer stores p-values as –log10(p):</u>

• $0.1 = 10^{-1}$→sig=1, $0.01 = 10^{-2}$→sig=2

• sig.mgh files

• Signed by sign of g

• p-value is for an unsigned test

Putting it all together

1. We used our empirical data to form a **design matrix: X**

2. We fit **regression coefficients (b and m)** to our x,y data

3. We created a **contrast matrix: C** to test our hypothesis for:
   1. Direction of effect: **g = C*β**
   2. Significance of effect: **t-test**

Two Groups

Thickness

Intercept: b1

Slope: m1

Slope: m2

Intercept: b2

Age
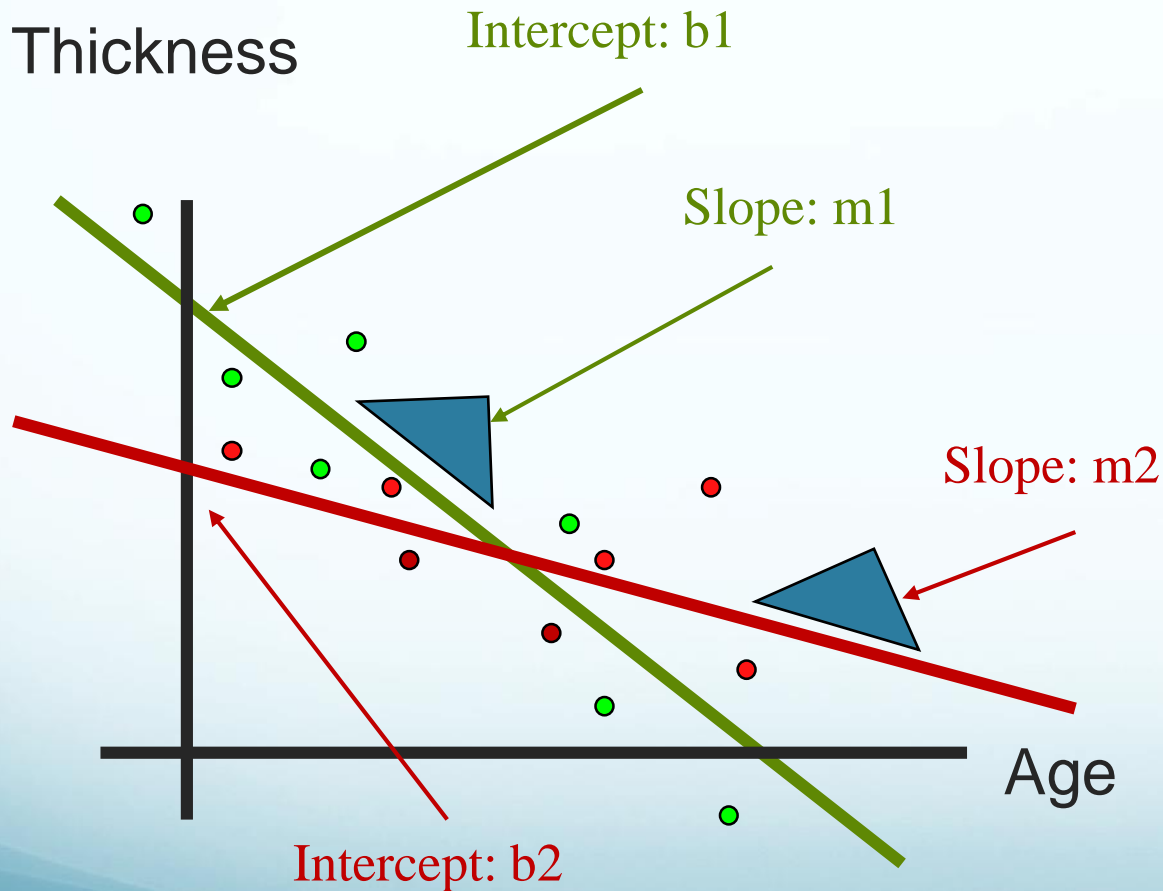
- Do groups differ in Intercept?
- Do groups differ in Slope?
- Is average slope different from 0?
- …

Two Groups

Thickness

Intercept: b1

Slope: m1

Slope: m2

Intercept: b2

Age
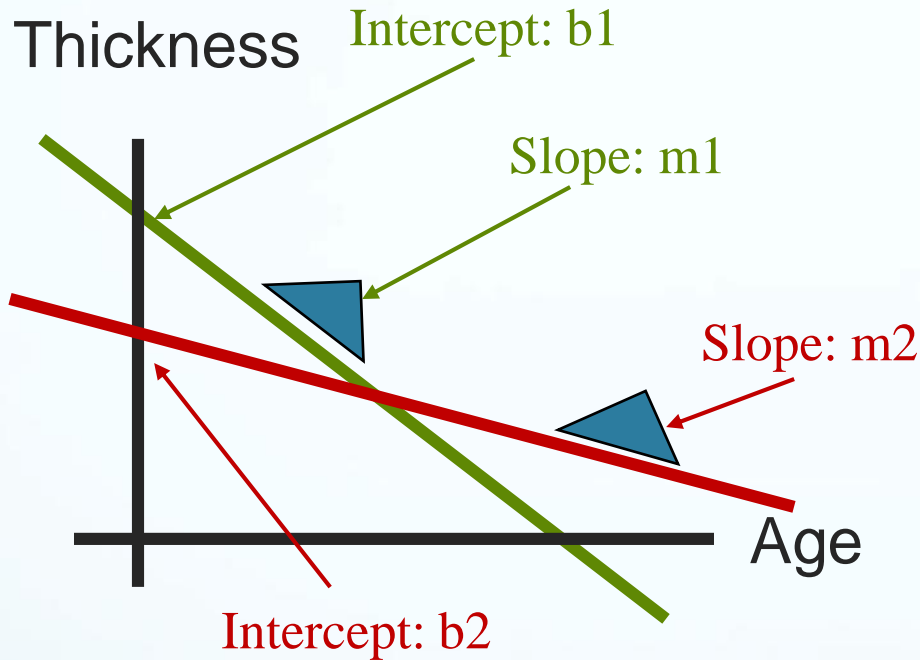
$$\begin{bmatrix} y11 \\ y12 \\ y21 \\ y22 \end{bmatrix} = \begin{bmatrix} 1 & 0 & x11 & 0 \\ 1 & 0 & x12 & 0 \\ 0 & 1 & 0 & x21 \\ 0 & 1 & 0 & x22 \end{bmatrix} * \begin{bmatrix} b1 \\ b2 \\ m1 \\ m2 \end{bmatrix} + n$$

number of columns = (number of groups)*(number of parameters)

$y11 = 1*b1 + 0*b2 + x11*m1 + \quad 0*m2 + n11$

$y12 = 1*b1 + 0*b2 + x12*m1 + \quad 0*m2 + n12$

$y21 = 0*b1 + 1*b2 + \quad 0*m1 + x21*m2 + n21$

$y22 = 0*b1 + 1*b2 + \quad 0*m1 + x22*m2 + n22$

Two Groups

Do groups differ in Intercept?
Does b1=b2?
Does b1-b2 = 0?
C = [+1 -1 0 0], g = $C*b$

Do groups differ in Slope?
Does m1=m2?
Does m1-m2=0?
C = [0 0 +1 -1], g = $C*b$

Is average slope different than 0?
Does (m1+m2)/2 = 0?
C = [0 0 0.5 0.5], g = $C*b$

$$Y = X*b+n$$

$$b= \begin{bmatrix} b1 \\ b2 \\ m1 \\ m2 \end{bmatrix}$$

Thickness

Intercept: b1

Slope: m1

Slope: m2

Age

Intercept: b2
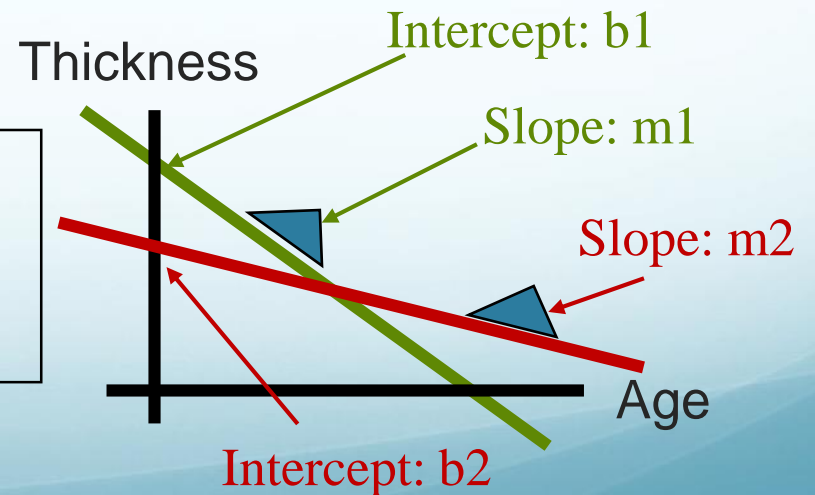
Surface-based Group Analysis in FreeSurfer

- Create your own design and contrast matrices

- Create an FSGD File
  - FreeSurfer creates design matrix
  - You still have to specify contrasts

- QDEC
  - Limited to 2 discrete variables, 2 levels max
  - Limited to 2 continuous variables

Processing Stages

- Specify Subjects and Surface measures
- Assemble Data:
  - Resample into Common Space
  - Smooth
  - Concatenate into one file
- Model and Contrasts (GLM)
- Fit Model (Estimate)
- Correct for multiple comparisons
- Visualize

Command-line Processing Stages

- Assemble Data (*mris_preproc*)
  - Resample into Common Space
  - Smooth
  - Concatenate into one file
- Model and Contrasts (GLM) (FSGD)
- Fit Model (Estimate) (*mri_glmfit*)
- Correct for multiple comparisons
- Visualize (*tksurfer*)

} *recon-all -qcache*

Run <u>after</u> all editing is done.

Specifying Subjects

Subject ID

$SUBJECTS_DIR

bert fred   jenny   margaret ...

SUBJECTS_DIR environment variable
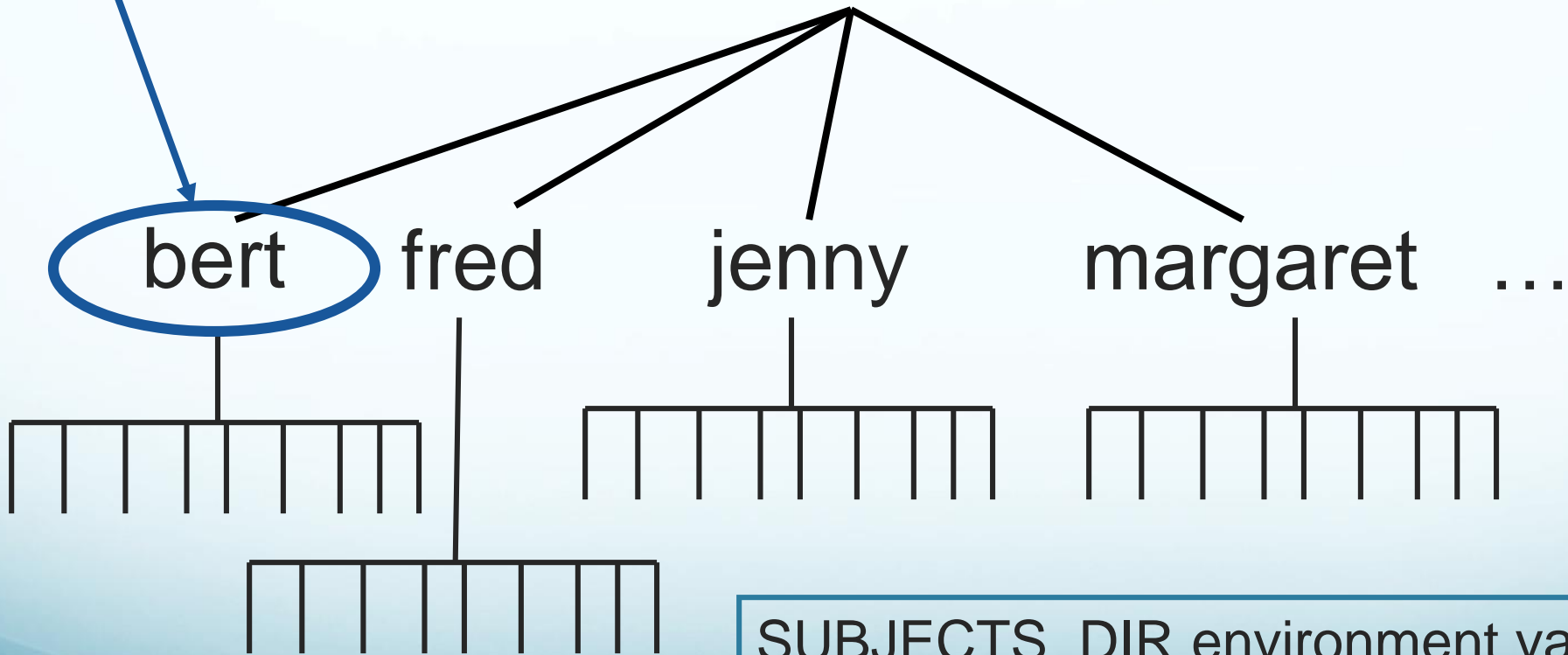
FreeSurfer Directory Tree
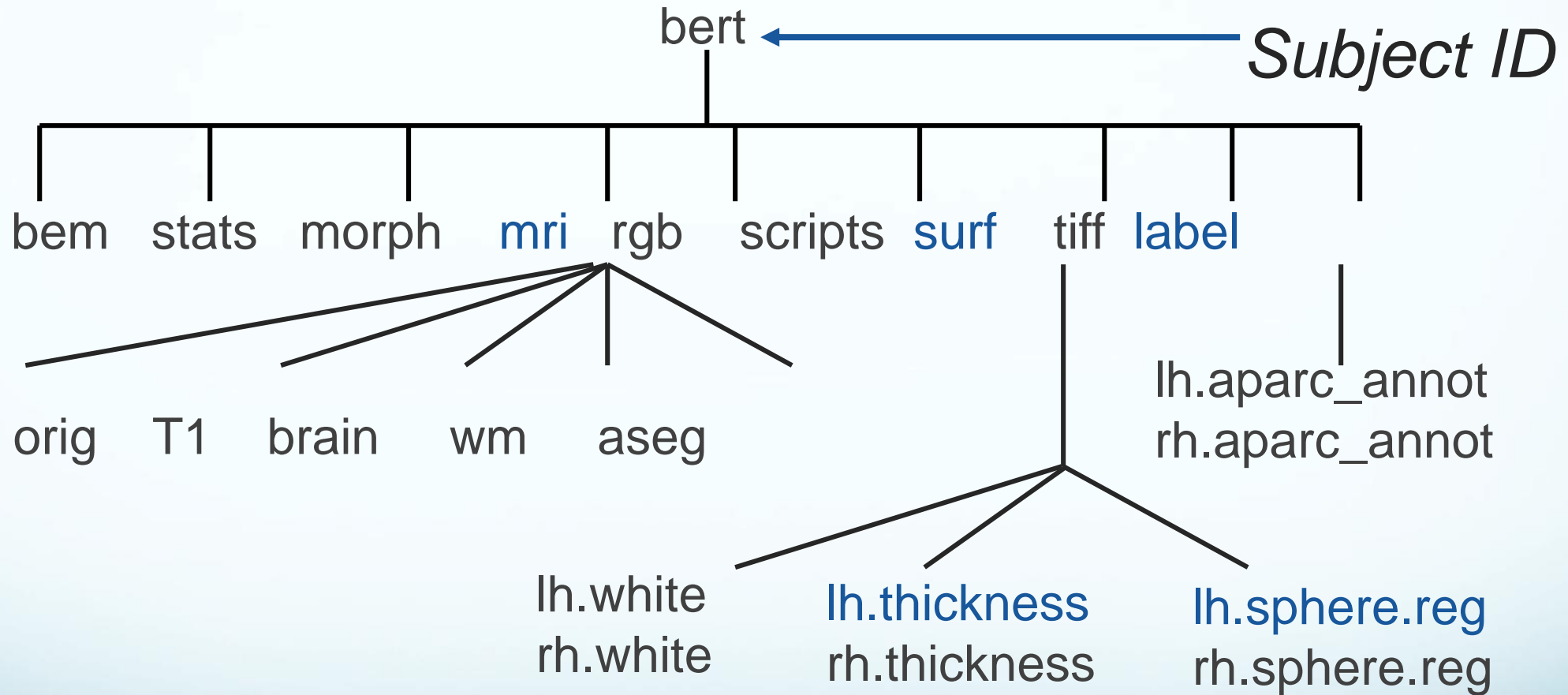
SUBJECTS_DIR environment variable

Example: Thickness Study

1. $SUBJECTS_DIR/bert/surf/lh.thickness

2. $SUBJECTS_DIR/fred/surf/lh.thickness

3. $SUBJECTS_DIR/jenny/surf/lh.thickness

4. $SUBJECTS_DIR/margaret/surf/lh.thickness

5. …

FreeSurfer Group Descriptor (FSGD) File

- Simple text file
- List of all subjects in the study
- Accompanying demographics
- Automatic design matrix creation
- You must still specify the contrast matrices

Note: *Can* specify design matrix explicitly with --design

FSGD Format

```
 GroupDescriptorFile 1
Class Male
Class Female
Variables                          Age     Weight        IQ
Input bert            Male         10      100           1000
Input fred            Male         15      150           1500
Input jenny           Female       20      200           2000
Input margaret        Female       25      250           2500
```

- One Discrete Factor (Gender) with Two Levels (M&F)
- Three Continuous Variables: Age, Weight, IQ

Class = Group

Note: Can *specify* design matrix explicitly with --design

FSGDF → X (Automatic)

Female Group

Male Group

Male Age

Female Age

$$X = \begin{bmatrix} 1 & 0 & 10 & 0 & 100 & 0 & 1000 & 0 \\ 1 & 0 & 15 & 0 & 150 & 0 & 1500 & 0 \\ 0 & 1 & 0 & 20 & 0 & 200 & 0 & 2000 \\ 0 & 1 & 0 & 25 & 0 & 250 & 0 & 2500 \end{bmatrix}$$

Age   Weight   IQ

C = [-1 1 0 0 0 0 0 0]

Tests for the difference in <u>intercept/offset</u> between groups

C = [0 0 -1 1 0 0 0 0]

Tests for the difference in <u>age slope</u> between groups

DODS – Different Offset, Different Slope

Another FSGD Example

- Two Discrete Factors
  - Gender: Two Levels (M&F)
  - Handedness: Two Levels (L&R)
- One Continuous Variable: Age

```
GroupDescriptorFile 1
Class MaleRight
Class MaleLeft
Class FemaleRight
Class FemaleLeft
Variables                              Age
Input bert          MaleLeft           10
Input fred          MaleRight          15
Input jenny         FemaleRight        20
Input margaret      FemaleLeft         25
```

Class = Group

Interaction Contrast

- Two Discrete Factors (no continuous, for now)
  - Gender: Two Levels (M&F)
  - Handedness: Two Levels (L&R)
- Four Regressors (Offsets)
  - MR (b_1), ML (b_2), FR (b_3), FL (b_4)

GroupDescriptorFile 1
Class MaleRight
Class MaleLeft
Class FemaleRight
Class FemaleLeft
Input bert   MaleLeft
Input fred   MaleRight
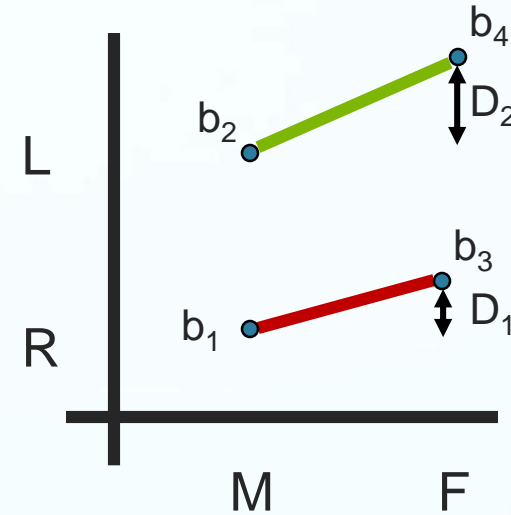Input jenny   FemaleLeft
Input margaret   FemaleRight

$$g = D_1 - D_2 \overset{?}{=} 0$$
$$g = (b_3 - b_1) - (b_4 - b_2)$$
$$= -b_1 + b_2 + b_3 - b_4$$
$$C = [-1\ +1\ +1\ -1]$$

Factors, Levels, Groups

Usually each Group/Class:

- Has its own Intercept

- Has its own Slope (for each continuous variable)

- *NRegressors = NClasses * (NVariables+1)*   DODS

- *Nregressors = NClasses + Nvariables*   DOSS

Why is this important? Because you will need to create contrast matrices, and the contrast matrix must have Nregressor elements.

Factors, Levels, Groups, Classes

Continuous Variables/Factors: Age, IQ, Volume, etc.

Discrete Variables/Factors:
    Gender, Handedness, Diagnosis

Levels of Discrete Variables:
- Handedness: *Left and Right*
- Gender: *Male and Female*
- Diagnosis: *Normal, MCI, AD*

Group or Class: Specification of All Discrete Factors
- Left-handed Male MCI
- Right-handed Female Normal

Assemble Data: mris_preproc

mris_preproc --help

--fsgd FSGDFile	: Specify subjects thru FSGD File
--hemi lh	: Process left hemisphere
--meas thickness	: subjectid/surf/hemi.thickness
--target fsaverage	: common space is subject fsaverage
--o lh.thickness.mgh	: output "volume-encoded surface file"

Lots of other options!

→ Output: lh.thickness.mgh – file with stacked
   thickness maps for all subjects
→ Input to Smoother or GLM

Surface Smoothing

- mri_surf2surf --help
- Loads stacked lh.thickness.mgh
- 2D surface-based smoothing
- Specify FWHM (eg, fwhm = 10 mm)
- Saves stacked lh.thickness.sm10.mgh
- Can be slow (~10-60min)
- recon-all –qcache (computes for each subject, run after you are finished editing subject)

mri_glmfit

- Reads in FSGD File and constructs **X**
- Reads in your contrasts (**C**1, **C**2, etc.)
- Loads data (lh.thickness.sm10.mgh)
- Fits GLM (ie, computes **b**)
- Computes contrasts (g=**C*****b**)
- t or F ratios, significances
- Significance -log10(p) (.01 → 2, .001 → 3)

mri_glmfit

mri_glmfit
  --y lh.thickness.sm10.mgh
  --fsgd gender_age.txt
  --C age.mtx –C gender.mtx
  --surf fsaverage lh
  --cortex
  --glmdir lh.gender_age.glmdir

  mri_glmfit --help

mri_glmfit

mri_glmfit

--y lh.thickness.sm10.mgh
--fsgd gender_age.txt
--C age.mtx –C gender.mtx
--surf fsaverage lh
--cortex
--glmdir lh.gender_age.glmdir

- Input file (output from smoothing).
- Stack of subjects, one frame per subject.

mri_glmfit

mri_glmfit
  --y lh.thickness.sm10.mgh
  --fsgd gender_age.txt
  --C age.mtx –C gender.mtx
  --surf fsaverage lh
  --cortex
  --glmdir lh.gender_age.glmdir

- FreeSurfer Group Descriptor File (FSGD)
- Group membership
- Covariates

mri_glmfit

mri_glmfit
  --y lh.thickness.sm10.mgh
  --fsgd gender_age.txt
  --C age.mtx –C gender.mtx
  --surf fsaverage lh
  --cortex
  --glmdir lh.gender_age.glmdir

- Contrast Matrices
- Simple text/ASCII files
- Test hypotheses

mri_glmfit

mri_glmfit
 --y lh.thickness.sm10.mgh
 --fsgd gender_age.txt
 --C age.mtx –C gender.mtx
 --surf fsaverage lh
 --cortex
 --glmdir lh.gender_age.glmdir

- Perform analysis on left hemisphere of fsaverage subject
- Masks by fsaverage cortex.label
- Computes FWHM in 2D

mri_glmfit

mri_glmfit
  --y lh.thickness.sm10.mgh
  --fsgd gender_age.txt
  --C age.mtx –C gender.mtx
  --surf fsaverage lh
  --cortex
  --glmdir lh.gender_age.glmdir

Output directory:
lh.gender_age.glmdir/
    beta.mgh – parameter estimates
    rvar.mgh – residual error variance
    etc …
    age/
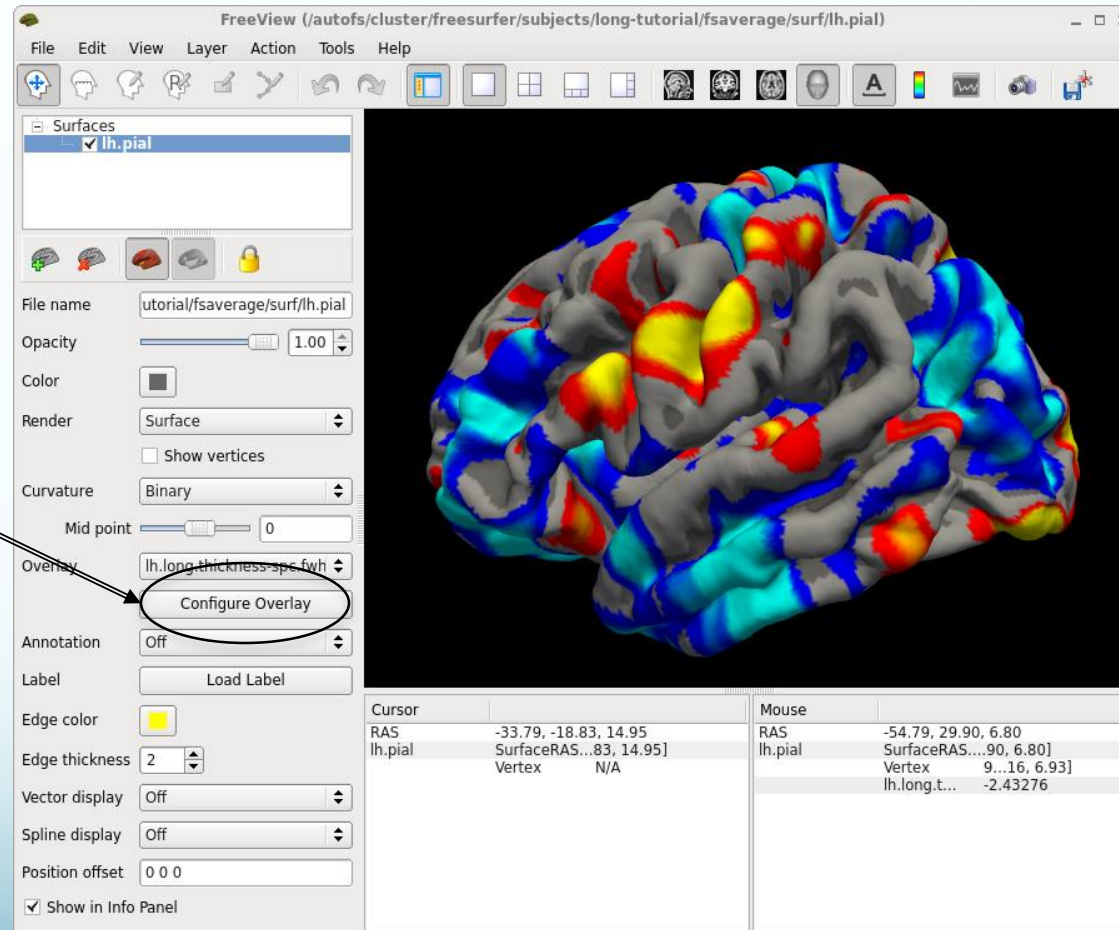   sig.mgh – -log10(p), uncorrected
      gamma.mgh, F.mgh
    gender/
      sig.mgh – -log10(p), uncorrected
      gamma.mgh, F.mgh

Visualization with freeview

freeview –f $FREESURFER_HOME/subjects/fsaverage/surf/lh.pial:overlay=sig.mgh

Use "Configure Overlay" tool to change thresholds for visualization (recall: lower threshold of 1.3 will only display regions where p<0.05)

GLM Analysis Using Aseg/Aparc Stats Files

mri_glmfit
--table aparc_lh_vol_stats.txt
--fsgd gender_age.txt
--C age.mtx --C gender.mtx
--glmdir roi.gender_age.glmdir

- Use "--table table.txt" instead of "--y" to specify input

- The rest of the command-line is the same as you would use for a group study (eg, FSGD file and contrasts).

- Output is text file sig.table.dat that lists the significances (-log10(p)) for each ROI and contrast.

Tutorial

Command-line Stream

- Create an FSGD File for a thickness study
- Age and Gender
- Run
  - mris_preproc
  - mri_surf2surf
  - mri_glmfit