MGH/HST Athinoula A. Martinos Center for Biomedical Imaging

MASSACHUSETTS GENERAL HOSPITAL

Harvard-MIT Health Sciences & Technology

Surface-based Group Analysis in FreeSurfer

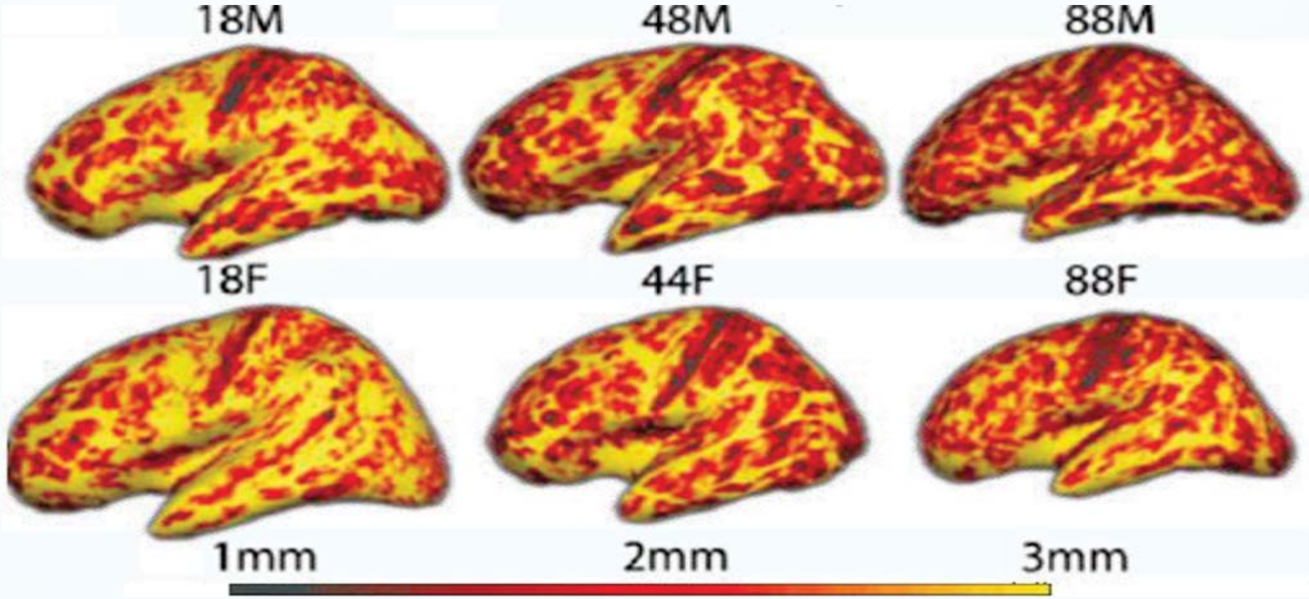
Group Analysis Objective

- 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

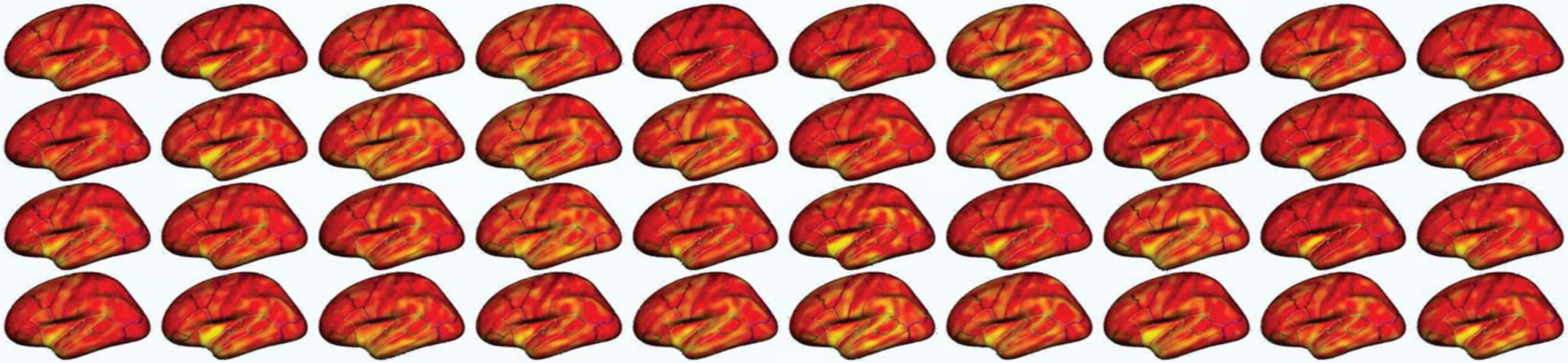


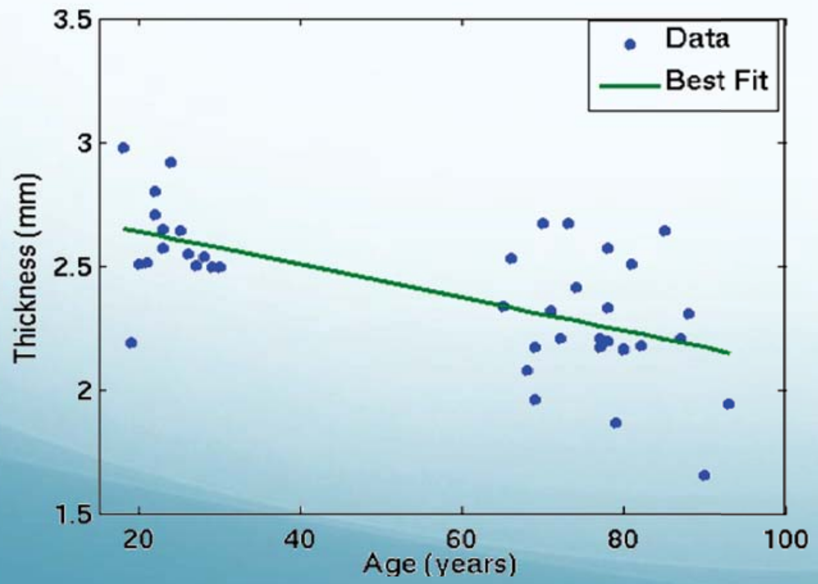
In which areas does thickness
Change with age?

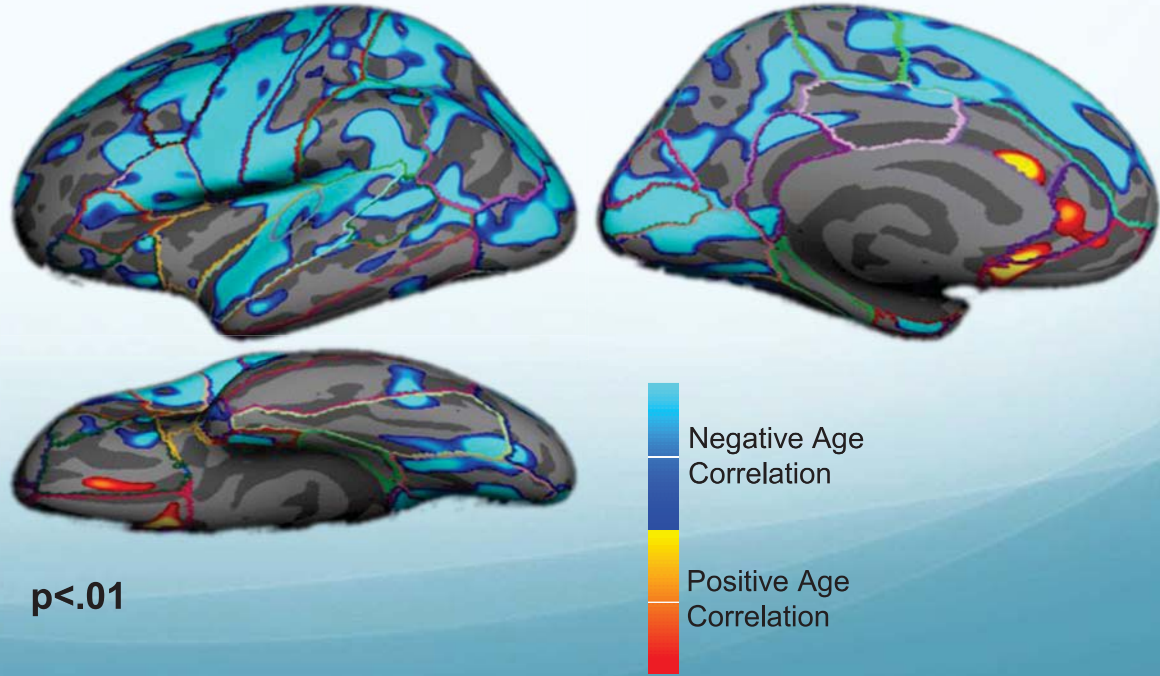
Cortical Thickness vs Aging
Salat et al, 2004, Cerebral Cortex

Aging Thickness Study

N=40 (all in fsaverage space)







p<.01

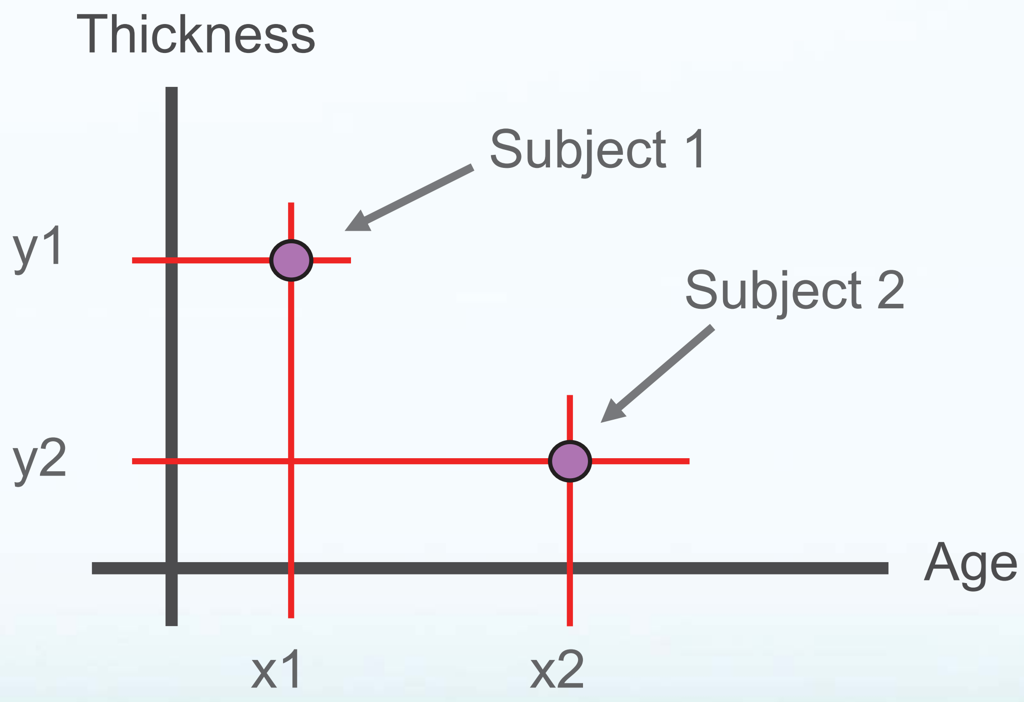
The General Linear Model (GLM)

GLM Theory

Is Thickness correlated with Age?

Dependent Variable, Measurement

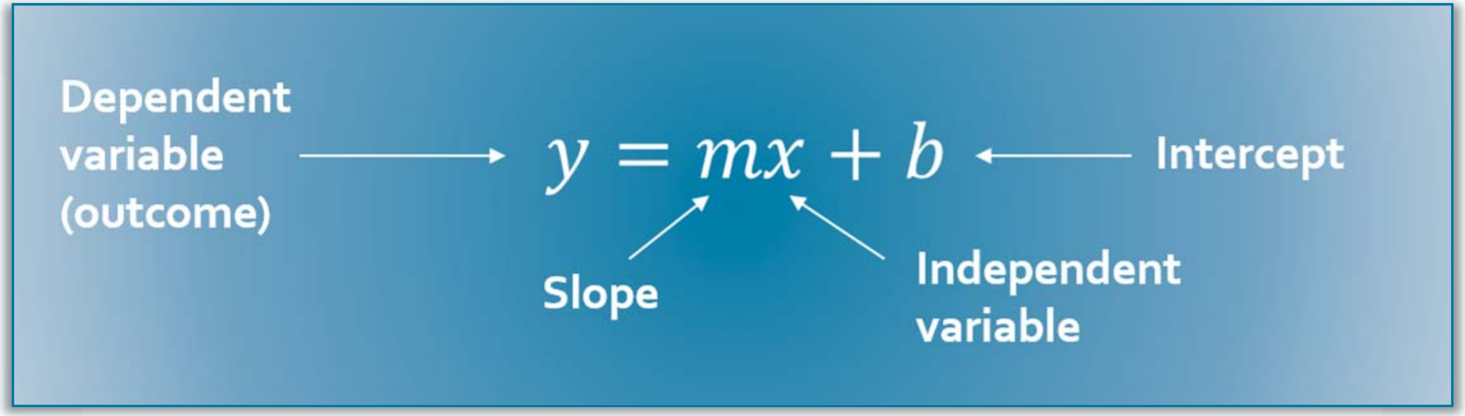
Thickness
IQ, Height, Weight, etc.

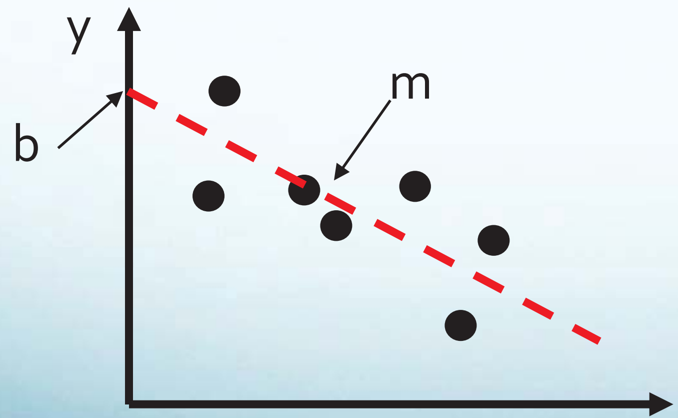


Independent Variable

Of course, you would need more then two subjects …

Linear Algebra Review (stay calm…)

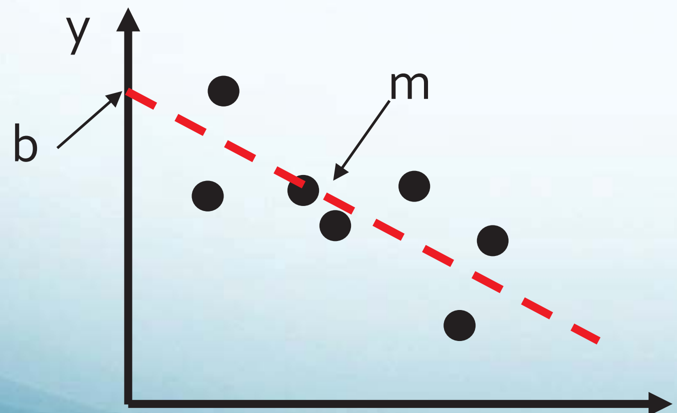




Linear Algebra Review (stay calm…)

We can put this in matrix format:

$$y = mx + b$$



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

Design Matrix

Regression Coefficients (parameters)

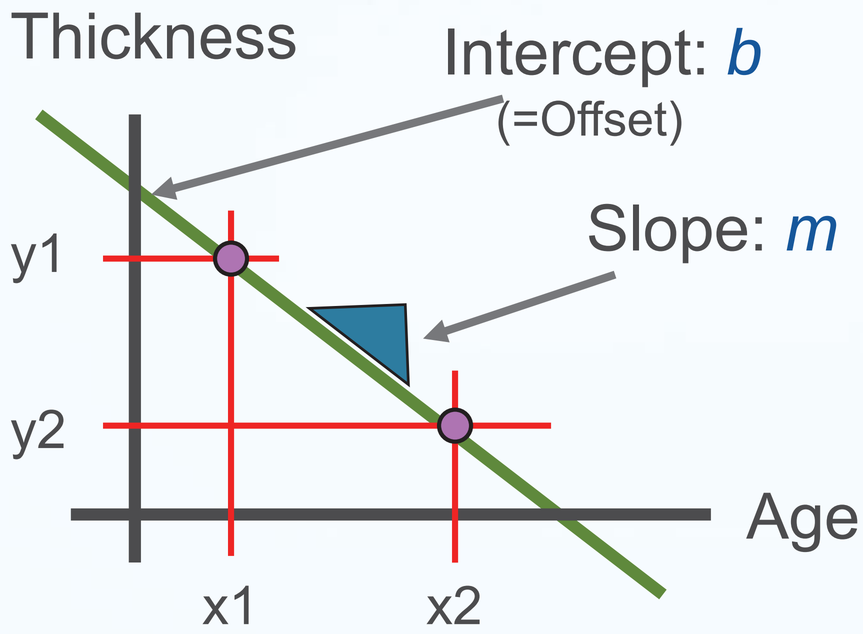
-One row per data point
-Add column of 1's for the offset term (*b*)
-One set of parameters

Matrix Multiplication

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

$$\left.\begin{aligned} y1 &= 1*b + x1*m \\ y2 &= 1*b + x2*m \\ y3 &= 1*b + x3*m \\ y4 &= 1*b + x4*m \end{aligned}\right\} \text{System of Linear Equations}$$

Linear Model



System of Linear Equations

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

Matrix Formulation

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

Two parameters

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

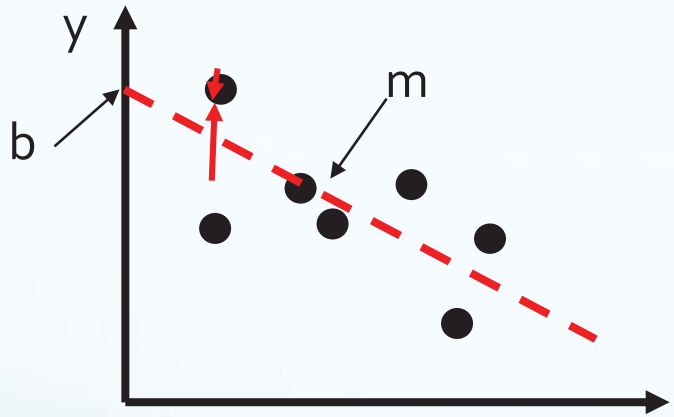
X = Design Matrix
b = Regression Coefficients
= Parameter estimates
= "betas"
= beta.mgh (mri_glmfit output)

$$\mathbf{Y} = \mathbf{X}*\mathbf{b}$$

$$\mathbf{b} = \begin{bmatrix} b \\ m \end{bmatrix}$$

Error

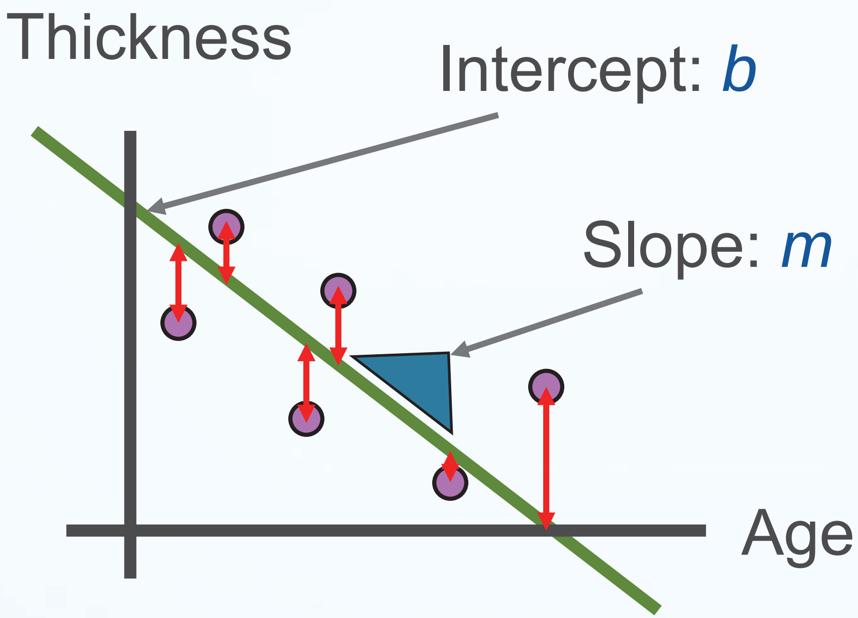
BUT… if we have the same *m* and *b* for all data points, we will have errors:



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



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

Y = X*b+n

$y1 = 1*b + x1*m + n1$

$y2 = 1*b + x2*m + n2$

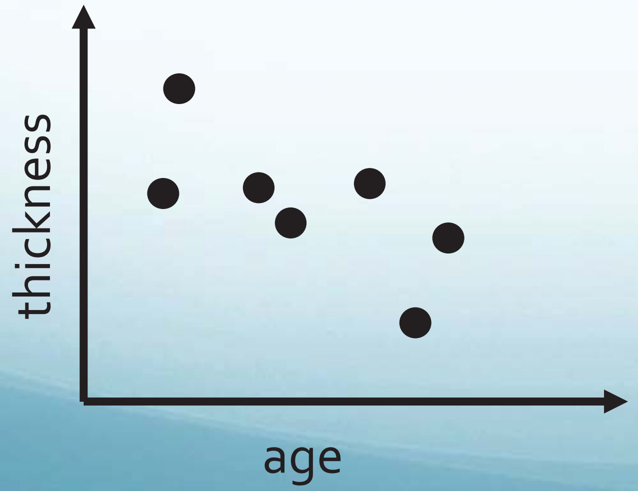
$y3 = 1*b + x3*m + n3$

$y4 = 1*b + x4*m + n4$

- 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

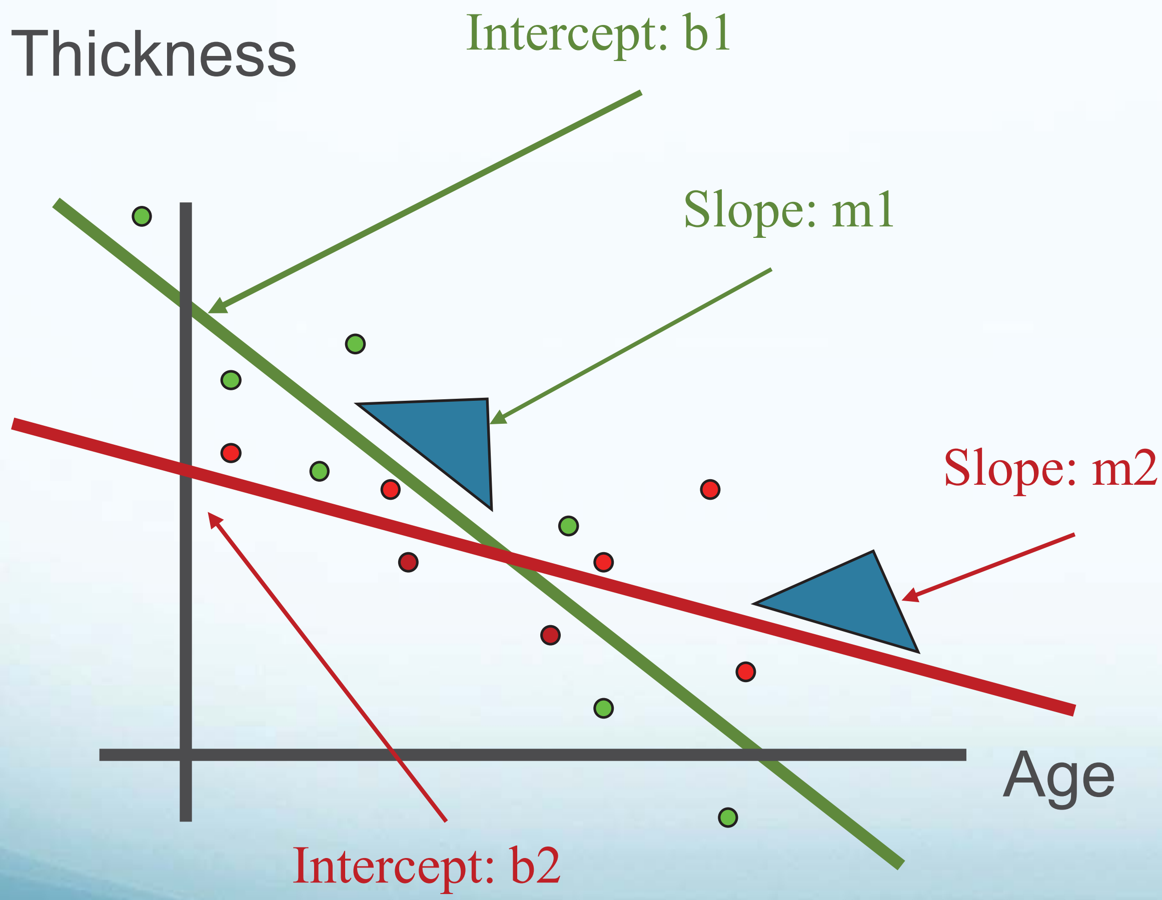
Putting it all together

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

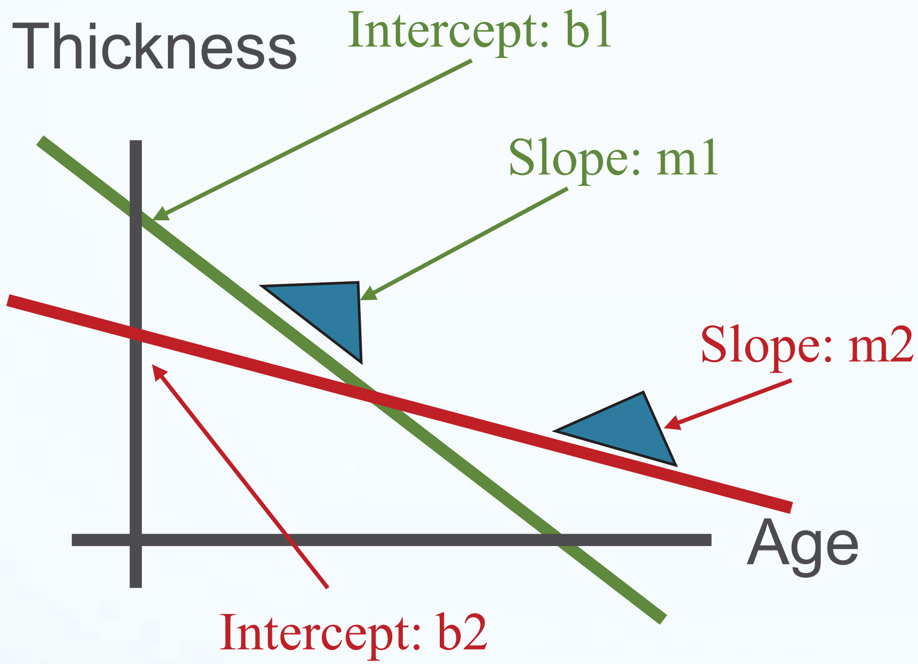
3. We created a **contrast matrix: C** to test our hypothesis for:
    1. Direction of effect: **g = C*β**
    2. Significance of effect: **t-test**

Two Groups



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

Two Groups



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

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

$$y11 = 1*b1 + 0*b2 + x11*m1 + 0*m2 + n11$$
$$y12 = 1*b1 + 0*b2 + x12*m1 + 0*m2 + n12$$
$$y21 = 0*b1 + 1*b2 + 0*m1 + x21*m2 + n21$$
$$y22 = 0*b1 + 1*b2 + 0*m1 + x22*m2 + n22$$

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

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