

Overview: Statistical Motivation

- What is the job of statistics?
 - To quantify uncertainty
 - To answering questions with data
- Why do we use statistics in fMRI?
 - People are noisy, MR scanners are noisy
 - To answering questions about the brain
 - *Where* is there experimentally-related variability?
 - *How* is such variability modulated by task subtlties?

1

Overview: Lies, Damn Lies & Statistical Parametric Mapping

- Purpose of these last two weeks of statistics?
 - To make you a thoughtful user of state of art tools
- Available “State of the art” tools
 - FSL - FMRIB Software Library, Oxford
 - C programs wrapped with tcl/tk scripts
 - VoxBo, Berkley
 - C programs, facilitated (?) with IDL scripts
 - SPM - Statistical Parametric Mapping, London
 - Matlab/C programs
- All can be abused by treating like a black box
 - Goal is to see “into” the black box (of SPM.)

2

Overview: Outline

- Fundamentals
 - Statistics, Hypothesis Tests, Type I Error
- Modeling
 - Basic stats & the general linear model
- Inference
 - t- & F-statistics & p-values w/ GLM
 - Note: If this is too slow, tell me... we can just do QA

3

Fundamentals: Probability

- Probability is the long run frequency of an event
- $P\{\text{“Heads”}\} = 1/2$
 - If you flip a coin repeatedly, in the long run, the fraction of heads will converge to 1/2.
- $P\{\text{“Response Time greater than 1200ms”}\} = 0.31$
 - If you repeat the experiment again and again, the fraction of experiments with $RT > 1200$ will converge to 0.31.

4

Fundamentals: Random Variables vs Parameters

- Random Variable
 - A quantity that is different each time it is observed
 - Ex: $X = \text{“Response Time in ms”}$
 - $P\{X > 1200\} = 0.31$
- Notation
 - Capital letters for random variables before observation
 - Lower case for a particular observed value
 - $P\{X > x\}$ chance of observing (a new) X at or above a particular x

5

Fundamentals: Random Variables vs Parameters

- Distribution of a Random Variable
 - Density function: $f(t)$
 - The relative frequency of different random values
 - Measure area under $f(\cdot)$ to find probabilities
 - $P\{X > 1200\} = \int_{1200}^{\infty} f(x)dx = \text{D.int}$

6

Fundamentals: Statistical Independence

- Statistical Independence, formally
 - Defined as property of distributions
 - Random variables X & Y are independent iff

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$
- Statistical Independence, informally
 - Random variables X and Y independent when knowledge about X tells you nothing about Y

7

Fundamentals: Random Variables vs Parameters

- Typically we assume a distribution to have specific form
 - Say reponse times are normally distributed...
 - $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2}(x - \mu)^2/\sigma^2)$

D:Dist

- Here μ & σ^2 are *parameters*

8

Fundamentals: Random Variables vs Parameters

- A parameter is *fixed, unknown* quantity
 - A parameter is not random
 - “Population parameter” (not data)
 - The parameter is a summary of a distribution
 - For the above distribution, the mean is μ
- Standard Summaries
 - Mean: Average or center of a distn
 - A measure of location
 - “First moment”
 - Variance: Average of squared variation from center
 - A measure of spread
 - “Second centered moment”
 - Covariance is a measure of association between two r.v.

Fundamentals: Hypotheses

- A hypothesis is a statement about a parameter
 - Note the special statistical meaning
 - For example,
 - $\mathcal{H}: \mu < 1000$ ms – The RT mean is < 1000 ms
 - $\mathcal{H}: \mu = 1200$ ms – The RT mean is 1200 ms
 - Not a statement about data.
- Just like a parameter, hypotheses are fixed and unknown
 - Taking a probability of a hypothesis is nonsense because it is not random
 - $P\{\mathcal{H}\}$
 - It’s either zero or one, but we can never know which.

10

Fundamentals: Statistics as Estimators

- A statistic is any function of the data
 - Let x_i be i th RT observation, $i = 1, \dots, n$
 - The sample mean: $\bar{x} = \frac{1}{n} \sum_i x_i$ is most common statistic
 - But x_3 , is also a statistic.
 - So is $(x_3 + x_5)/2$
 - So is $[x_3 \ x_5 \ x_6/x_8]$, a 3-vector
- Statistics usually are *estimators* of parameters
 - The sample mean is an estimate of the true RT mean N: Samp.v.Tr
 - x_3 is also an estimate of the RT mean
 - But \bar{x} is better
 - $[x_3 \ x_5 \ x_6/x_8]$ doesn’t estimate anything useful
- Estimating parameters is the bread & butter of statistics

11

Fundamentals: Statistics as Tests

- To test a hypothesis we construct “test statistics”
- Null Hypothesis
 - First a *Null Hypothesis* is defined, \mathcal{H}_0
 - The null expresses the default or “no effect” state
 - For example, if the expected response time is 1200ms, we might test $\mathcal{H}_0: \mu = 1200$ ms
 - Alternative hypothesis expresses outcome of interest
 - $\mathcal{H}_A: \mu > 1200$, or
 - $\mathcal{H}_A: \mu \neq 1200$

12

Fundamentals: Statistics as Tests

- Test Statistic
 - A test statistic summarizes evidence about \mathcal{H}_0
 - Typically, test statistic is small in magnitude when the hypothesis is true, and large when false
 - For example, to test \mathcal{H}_0 above, I might use statistic

$$T = \bar{X} - 1200$$
 ... which should be near zero if $\mu = 1200$.

13

Fundamentals: Hypothesis Tests

- Null Hypothesis \mathcal{H}_0
 - Default state
- Test Statistic T
 - Function of to-be observed data
- Observed Test Statistic t

14

Fundamentals: Hypothesis Tests

- p-values
 - A p-value summarizes the evidence against \mathcal{H}_0
 - p-value is chance of observing value more extreme than t under the null hypothesis
 - For a one-sided alternative - $\mathcal{H}_A : \mu > 1200$
 - p-value = $\mathbf{P}\{T \geq t | \mathcal{H}_0\}$ D:Dis
 - For a two-sided alternative - $\mathcal{H}_A : \mu \neq 1200$
 - p-value = $\mathbf{P}\{|T| \geq t | \mathcal{H}_0\}$ D:Dis

15

Fundamentals: Hypothesis Tests

- α 's
 - If we always "Reject \mathcal{H}_0 " when $p < \alpha$, then *false positive rate* is controlled at α

16

Fundamentals: Type I & Type II Error

- Two rights, two wrongs

	\mathcal{H}_A	\mathcal{H}_0
Reject \mathcal{H}_0		
Accept \mathcal{H}_0		

- All of hypothesis testing focuses on controlling Type I error
 - It's easy!
 - Null usually only consists of a single case
- Type II Error
 - Power = 1-P(Type II Error), so of great interest
 - Hard, since depends on unobserved alternative

17

Fundamentals: Quiz

- A study considers two possible treatments for bipolar subjects, drug A and drug B. For group A, let a_i be subject i 's time until a manic episode in days from the start of treatment; let b_j be subject j 's time until manic episode. There are 20 subjects in each group.
- I-1. The average of group A, $\bar{a} = \frac{1}{20} \sum_i x_i$ is a parameter. T | F?
- I-2. I want to test if drug A is better (longer time to manic episode) than drug B. I should test
 - \mathcal{H}_0 : _____
 - Where ___ is _____
 - and ___ is _____

18

Fundamentals: Quiz

- I compute a two sample t-test, T , and obtain a test statistic of 2.32 and a p-value is 0.013.

I-3. Choose an inequality and fill in the math:

$$P(_ \leq _ \geq _ | \mathcal{H}_0) = _.$$

Draw & label a picture of this expression.

I-4. The probability that the null hypothesis is true is 0.013. T | F?

19

Modeling: General Enterprise

- “All models are wrong, some are useful”
 - Really, everything we can do is approximate
 - The game is finding a accurate, but believable model
- One simple model
 - Reponse time data x_1, x_2, \dots, x_n
 - Reponse time model $\mu_1, \mu_2, \dots, \mu_n$
 - n data points, n parameters.
 - A very accurate model! Perfect fit!
- Parsimony is key
 - How few parameters successfully describe the data?
 - Between two equally-good models, always prefer simpler one, one with fewer parameters

20

Models: Linear Regression & Least Squares

- Simplest Model

$$\rightarrow y_i = \beta_1 + \beta_2 x_i + \epsilon_i$$

D:4cx

→ Where

- y_i is the data, say response times for subject i
- x_i is a predictor, say accuracy for subject i
- β_1 is
- β_2 is
- ϵ_i is mean zero error

21

Models: Linear Regression & Least Squares

- How to estimate $\beta_1 \beta_2$?

- Tweak β 's until some error metric is minimized
- What metric?

(Let $e_i = y_i - \hat{y}_i$, where \hat{y}_i is fitted value)

- Worst error? $\max_i e_i$?
 - Sum of absolute value of errors? $\sum_i |e_i|$?
 - Sum of squared errors? $\sum_i e_i^2$?
- Easiest theoretically

22

Models: Linear Regression & Least Squares

- Least Squares

→ Find β 's that minimize

$$\rightarrow \sum_i e_i^2 = \sum_i (y_i - \beta_1 - \beta_2 x_i)^2$$

→ Taking deriv's w.r.t. β 's and setting to zero...

$$\hat{\beta}_1 = \bar{y}$$

$$\hat{\beta}_2 = \frac{\sum_i (x_i - \bar{x}) y_i}{\sum_i (x_i - \bar{x})^2}$$

N:Hat

23

Models: Multiple Regression

- Less Simple Model

$$\rightarrow y_i = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + \dots + \beta_p x_{pi} + \epsilon_i$$

→ Where

- x_{pi} is p th predictor for subject i
- β_p is p th parameter

- Least Squares

→ Principal still works

- But equations for each $\hat{\beta}_j$ look like hell, unless

- Use Matrix Formulation to Simplify

24

- Design Matrix

$$X = \begin{bmatrix} 1 & 1 & x_1 \\ \vdots & \vdots & \vdots \\ 1 & 1 & x_{n_1} \\ 1 & -1 & x_{n_1+1} \\ \vdots & \vdots & \vdots \\ 1 & -1 & x_{n_1+n_2} \end{bmatrix}$$

→ Model for young:

$$y_i = \beta_1 + \beta_2 + \beta_3 x_i$$

→ Model for old:

$$y_i = \beta_1 - \beta_2 + \beta_3 x_i$$

31

Models: GLM & Contrasts

- Usually not interested in whole β vector
- Contrasts select an effect of interest
 - Contrast is a length- p row vector, c
 - $c\beta$ is a linear combination of β s
 - Can perform hypothesis test on $c\hat{\beta}$
- AnCova Example
 - Interest was in Age effect
 - Appropriate contrast
 - $c = [0 \ 1 \ 0]$
 - $c\beta = \beta_2$
 - Appropriate hypothesis
 - $\mathcal{H}_0 : c\beta = 0$

32

Models: GLM & Estimability

- Some GLM's are not uniquely determined
 - Sometimes it's easier to specify the model that way
- AnCova example revisited
 - Alternate design matrix

$$X = \begin{bmatrix} 1 & 1 & 0 & x_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & x_{n_1} \\ 1 & 0 & 1 & x_{n_1+1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & x_{n_1+n_2} \end{bmatrix} \quad \& \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}$$

33

Models: GLM & Estimability

- Model for young:
 - $y_i = \beta_1 + \beta_2 + \beta_4 x_i$
- Model for old:
 - $y_i = \beta_1 + \beta_3 + \beta_4 x_i$
- But there are an infinite number of solutions!
- For any β ,
 - I can add 42.9 to β_1 , and
 - I can subtract 42.9 from β_2 and β_3 and get the very same fit!

34

Models: GLM & Estimability

- An estimable contrast is guaranteed to have a unique value
 - Even if an infinite number of β 's yield same fit
- Technical condition
 - A contrast c is estimable if it is in the row space of X
- Simple condition
 - Usually sum to zero is sufficient
 - Software usually checks
- AnCova Example
 - In the above example, the contrast
 - $c = [0 \ 1 \ 0 \ 0]$ is not estimable, but
 - $c = [0 \ 1 \ -1 \ 0]$ is estimable

35

Models: Nonlinear Models

- Linear Models
 - Fit is a linear combination of the parameters
 - Ex: $y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + \epsilon_i$
 - This is a linear model
- Nonlinear models
 - Ex: $y_i = \beta_1 + 1/(\beta_2 + \beta_3 x_i) + \epsilon_i$
- Finding β 's hard
 - Still can use least squares principle
 - But not GLM
 - Requires nonlinear optimization

36

Models: Quiz

- Following from the fundamentals quiz, let

$$Y = (a_1, \dots, a_{20}, b_1, \dots, b_{20})^\top.$$

- II-1. Draw the GLM design matrix for the two sample t-test. Label the rows and columns.

37

Models: Quiz

- II-2. What is the relevant contrast for the hypothesis \mathcal{H}_0 in question I-2?
- II-3. I have the ages of the 40 subjects and am concerned about an age effect. What model can I fit?
- II-4. For each subject I have the results of a personality inventory which rates individuals on 20 axes (extrovert, sociopath, etc). With these 20 additional covariates what model can I fit?

38

Inference: Overview

- Statistical inference
 - Drawing conclusions from data
 - Performing hypothesis tests
 - Measuring uncertainty of parameter estimates
- Here come the assumptions
 - Note we haven't mentioned any assumptions so far?
 - Least squares's $\hat{\beta}$'s are valid without any
 - p-values need distributions

39

Inference: Linear Algebra & Random Variables

- The mean or expected value operates elementwise.

$$E(Y) = E \begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} E(y_1) \\ \vdots \\ E(y_i) \\ \vdots \\ E(y_n) \end{pmatrix}$$

- Expectation is linear
 - $E(AY + \epsilon) = AE(Y) + E(\epsilon)$
 - Where A is a fixed matrix

40

Inference: Linear Algebra & Random Variables

- For a vector, we need to now variance *and* covariances

$$\text{Var}(Y) = \begin{bmatrix} \text{Var}(y_1) & \text{Cov}(y_1, y_2) & \cdots & \text{Cov}(y_1, y_n) \\ & \text{Var}(y_2) & \cdots & \text{Cov}(y_2, y_n) \\ & & \ddots & \vdots \\ & & & \text{Var}(y_n) \end{bmatrix}$$

- The variance-covariance operator is not linear
 - $\text{Var}(AY) = A\text{Var}(Y)A'$

41

Inference: GLM & Normality

- Normality is a magical distribution
 - Averages of *almost any* distribution converge to Normal
 - Nice simple, symmetric distribution
- Normality and Least Squares
 - Normality and using principal of Maximum likelihood gives the same $\hat{\beta}$ estimators as least squares
 - Assume
 - X is fixed and known
 - $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, for each i
 - $\mathcal{N}(\cdot)$ is the Normal distribution
 - ϵ_i and ϵ_j independent for $i \neq j$
 - σ^2 is the variance, the magnitude of randomness

42

Inference: GLM & Normality

- In words
 - Errors are Normal, independently and identically distributed
- Graphically D:Gp)

43

Inference: GLM & Normality

- In matrix notation
 - $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_n)$
 - Where $\mathbf{0}$ is a n -vector of zeros
 - I_n is a $n \times n$ identity matrix
 - In this *multivariate* form, instead of variance σ^2 we have variance-covariance *matrix* $\sigma^2 I_n$
 - All off diagonals of I_n is zero, and hence each of the e_i 's are uncorrelated.
- Normality very powerful!

44

Inference: GLM & Normality

- Normal random variable shifted is Normal
 - We know
 - $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_n)$ and
 - $Y = X\beta + \epsilon$
 - So Y is Normal
 - Mean of Y : $\mathbf{E}(Y) = \mathbf{E}(X\beta + \epsilon) = X\beta + \mathbf{0}$
 - Var of Y : $\text{Var}(Y) = \text{Var}(X\beta + \epsilon) = \mathbf{0} + \sigma^2 I_n$
 - $Y \sim \mathcal{N}(X\beta, \sigma^2 I_n)$
- So Normality assumptions on the error is equivalent to Normality assumptions on the data!

45

Inference: GLM & Normality

- Linear combination of Normals is Normal
 - Recall $\hat{\beta} = (X'X)^{-1}X'Y$
 - $\hat{\beta}$ is just a linear combinations of Normal Y 's!
 - Mean of $\hat{\beta}$:
 - $\mathbf{E}(\hat{\beta}) = (X'X)^{-1}X'\mathbf{E}(Y) = (X'X)^{-1}X'X\beta = \beta$
 - Variance of $\hat{\beta}$?
 - $\text{Var}(\hat{\beta}) = \text{Var}((X'X)^{-1}X'Y)$
 - $= (X'X)^{-1}X'\text{Var}(Y)((X'X)^{-1}X)'$
 - $= (X'X)^{-1}X'(\sigma^2 I_n)((X'X)^{-1}X)'$
 - $= (X'X)^{-1}X'(\sigma^2 I_n)X(X'X)^{-1}$
 - $= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}$

46

Inference: GLM & Normality

- All of that yields...
 - $\hat{\beta} \sim \mathcal{N}(\beta, (X'X)^{-1}\sigma^2)$
- So Normality assumptions buy us alot
 - $\epsilon \sim \mathcal{N}(0, \sigma^2 I_n) \Rightarrow \hat{\beta} \sim \mathcal{N}(\beta, \sigma^2(X'X)^{-1})$

47

Inference: GLM & t-test

- t-test is a Signal-to-Noise measure
 - Ratio of estimate to standard deviation of estimate
- GLM & t-test
 - Contrast: $c\hat{\beta} \sim \mathcal{N}(c\beta, c(X'X)^{-1}c'\sigma^2)$
 - Contrast t-test:

$$T = \frac{c\hat{\beta}}{\sqrt{c(X'X)^{-1}c'\hat{\sigma}^2}}$$

- What's $\hat{\sigma}^2$?
- The mean squared error, an estimate of σ^2
 - $\hat{\sigma}^2 = \frac{1}{n-p} \sum_i (y_i - \hat{y}_i)^2 = \frac{1}{n-p} (Y - X\hat{\beta})'(Y - X\hat{\beta})$
- Result has t-distribution with $n - p$ degrees of freedom

48

Inference: GLM & t Distribution

- Why not Normal?

A t ratio isn't Normal b/c σ^2 isn't known

D:ivN

49

Inference: GLM & t-test - Pain Example

- Subjects burned with laser on the back of their hand

- Variables
 - Perceived pain intensity, rating 0 - 10
 - Laser energy setting

- Research Question

- What is relationship between laser energy and pain rating?
- Does it vary with sex?

- Design Matrix

- Column 1: Grand Mean
- Column 2: Laser energy for men
- Column 3: Laser energy for women

- Use t test to compare men & women

50

Inference: GLM & Extra Sums of Squares

- Often more than one possible model

- How to compare?

- Residual Sum of Squares (RSS)

$$\rightarrow \sum_i e_i^2 = \sum_i (y_i - \hat{y}_i)^2 = (Y - \hat{Y})'(Y - \hat{Y})$$

51

Inference: GLM & Extra Sums of Squares

- Extra Sums of Squares

- Let X_1 be a GLM with p_1 predictors
- Let $[X_1 X_2]$ be a GLM with $p = p_1 + p_2$ predictors
 - $Y = [X_1 X_2][\beta'_1 \beta'_2]' + \epsilon$
- Residual Sum of Squares under X_1
 - $RSS_1 = (Y - \hat{Y}_1)'(Y - \hat{Y}_1)$
- Residual Sum of Squares under $[X_1 X_2]$
 - $RSS_{12} = (Y - \hat{Y}_{12})'(Y - \hat{Y}_{12})$
- Bigger model, better fit, smaller errors
 - $RSS_{12} < RSS_1$
- But, significantly better fit?

52

Inference: GLM & F-test

- F-test

- Under $\mathcal{H}_0 : \beta_2 = 0$, the difference between RSS_1 and RSS_{12} should be small
 - No extra “real” variation to account for w/ X_2
- F-test assess magnitude of extra variation

$$F = \frac{(RSS_1 - RSS_{12})/p_2}{RSS_{12}/(n - p)} = \frac{(RSS_1 - RSS_{12})/p_2}{\hat{\sigma}^2}$$

Compares change in RSS to residual error

53

Inference: GLM & F-test

- Contrast representation of F tests

- Can define F -test with contrasts
- One contrast for each column of null model
 - $[1\ 0\ 0\ 0]$
 - $[0\ 1\ 0\ 0]$
 - $[0\ 0\ 1\ 0]$ tests $\mathcal{H}_0 : \beta_1 = \beta_2 = \beta_3 = 0$
- Or, a collection of contrasts to be simultaneously tested as zero
 - $[1\ -1\ 0\ 0]$
 - $[0\ 0\ 1\ -1]$ tests $\mathcal{H}_0 : \beta_1 - \beta_2 = \beta_3 - \beta_4 = 0$

- Don't have to explicitly separate X into $[X_1 X_2]$

- Can compare any two models as long as they are *nested*
- In that case “ p_2 ” is difference in # of parameters

54

Inference: GLM & F-test - Pain Example

- Subjects burned with laser on the back of their hand
 - Variables
 - Perceived pain intensity, rating 0 - 10
 - Laser energy setting
- Research Question
 - What is relationship between laser energy and pain rating?
 - Does it vary by individual?
- Design Matrix
 - Column 1: Grand Mean
 - Column $i+1$: i^{th} subject's laser energy
- Use F test to compare RSS of this model with previous

55

Inference: Quiz

- Following from the Models quiz, I fit an AnCova model with 3 parameters: (1) a group A mean β_1 , (2) a group B mean β_2 , (4) an age covariate β_3 .
- III-1. Draw this model: Age on x-axis, y (relapse time) on y-axis, and a slope for each group.
- III-2. In terms of β 's, what is intercept for group A?
- III-3. Is the contrast $[1 \ 0 \ 0]$ estimable? What it's interpretation?
- III-3. What is the contrast for the group effect?
- III-4. Write down the contrasts for the F-test of "no group effect and no age effect".

56