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?

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.)

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

Fundamentals: Probability

- Probability is the long run frequency of an event
- P{"Heads"} = 1/2
  - If you flip a coin repeately, in the long run, the fraction of heads will converge to 1/2.
- P{"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.

Fundamentals: Random Variables vs Parameters

- Random Variable
  - A quantity that is different each time it is observed
  - Ex: X = "Response Time in ms"

    $\mathsf{P}\{X > 1200\} = 0.31$

Fundamentals: Random Variables vs Parameters

- Distribution of a Random Variable
  - Density function: $f(t)$
    - The relative frequncy of different random values
  - Measure area under $f(\cdot)$ to find probabilities
    - $\mathsf{P}\{1200 < X < 5000\} = \int_{1200}^{5000} f(x)dx =$ D:Int

Fundamentals: Random Variables vs Parameters

- Distribution of a Random Variable (con't)
  - ⇀ Distribution function: $F(x) = \mathsf{P}\{X \le x\}$
    - – Chance of observing X at or below x D:Int
- Notation
  - ⇀ Capital letters for random variables before observation
  - ⇀ Lower case for a particular observed value

Fundamentals: Independence

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

    $\mathsf{P}\{X \le x, Y \le y\} = \mathsf{P}\{X \le x\}\mathsf{P}\{Y \le y\}$

    Equivalently

    $f(x, y) = f(x)f(y)$
- Statistical Independence, informally
  - ⇀ Random variables X and Y independent when knowlege about X tells you nothing about Y

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*

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 < 1000ms

    $\mathcal{H}$: $\mu = 1200$ ms – The RT mean is 1200 ms
  - ⇀ Not a statment about data.
- Just like a parameter, hypotheses are fixed and unkown
  - ⇀ Taking a probability of a hypothesis is nonsense because it is not random
    - – $\mathsf{P}\{\mathcal{H}\}$

      It's either zero or one, but we can never know which.

Fundamentals: Statistics as Estimators

- A statistic is any function of the data
  - ⇀ Let x_i be ith RT observation, $i = 1, \ldots, n$
  - ⇀ 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:Smp.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

Fundamentals: Statistics as Tests

- To test a hypothesis we construst "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 $\mathcal{H}_0 : \mu = 1200ms$
  - ⇀ Alternative hypothesis expresses outcome of interest
    - – $\mathcal{H}_A$: $\mu > 1200$, or
    - – $\mathcal{H}_A$: $\mu \neq 1200$

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 $\mu = 1200$.

Fundamentals: Hypothesis Tests

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

Fundamentals: Hypothesis Tests

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

Fundamentals: Hypothesis Tests

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

Fundamentals: Type I & Type II Error

- Two rights, two wrongs

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

- 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

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 __________________

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?

Modeling: General Enterprise

- "All models are wrong, some are useful"
  - Really, everything we can do is approximate
  - The game is finding a accurate, but belivable model
- One simple model
  - Reponse time data x_1, x_2, ..., x_n
  - Reponse time model μ_1, μ_2, ..., μ_n
    - n data points, n parameters.
    - A very accurate model! Perfect fit!
- Parsiomony is key
  - How few parameters successfuly describe the data?
  - Between two equally-good models, always prefer simplier one, one with fewer parameters

Models: Linear Regression & Least Squares

- Simplest Model
  - $y_i = \beta_1 + \beta_2 x_i + \epsilon_i$ D:4ex
  - 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

Models: Linear Regression & Least Squares

- How to estimate β_1 β_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

Models: Linear Regression & Least Squares

- Least Squares
  - Find β's that minimize
  - $\sum_i e_i^2 = \sum_i (y_i - \beta_1 - \beta_2 x_i)^2$
  - Taking deriv's w.r.t. β's and setting to zero...
    $\hat{\beta}_1 = \bar{y}$
    $\hat{\beta}_2 = \sum_i (x_i - \bar{x}) y_i / \sum_i (x_i - \bar{x})^2$ N:Hat

Models: Multiple Regression

- Less Simple Model
  - ⇀ $y_i = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + \cdots + \beta_p x_{pi} + \epsilon_i$
  - ⇀ Where
    - – x_{pi} is pth predictor for subject i
    - – β_p is pth parameter
- Least Squares
  - ⇀ Principal still works
    - – But equations look like hell, unless
- Use Matrix Formuation to Simplify

Models: Linear Algebra Review

- Matrix multiplication
- Inverse
- Transpose
  - ⇀ Different notation

    A' is the same as $A^\top$
- Transpose & Multiplication
  - ⇀ Transpose permutes multiplication
  - ⇀ $(AB)' = C'B'$
  - ⇀ $(A'BC)' = C'B'A$

$$
\begin{array}{ccccccccccc}
y_1 & = & \beta_1 x_{11} & + & \beta_2 x_{21} & + & \cdots & + & \beta_p x_{p1} & + & \epsilon_1 \\
\vdots & & \vdots & & \vdots & & & & \vdots & & \vdots \\
y_i & = & \beta_1 x_{1i} & + & \beta_2 x_{2i} & + & \cdots & + & \beta_p x_{pi} & + & \epsilon_i \\
\vdots & & \vdots & & \vdots & & & & \vdots & & \vdots \\
y_n & = & \beta_1 x_{1n} & + & \beta_2 x_{2n} & + & \cdots & + & \beta_p x_{pn} & + & \epsilon_n
\end{array}
$$

$$
\begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{p1} \\ \vdots & \vdots & & \vdots \\ x_{1i} & x_{2i} & \cdots & x_{pi} \\ \vdots & \vdots & & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{pn} \end{bmatrix} \times \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_i \\ \vdots \\ \epsilon_n \end{bmatrix}
$$

$$
Y \quad = \quad X \quad \beta \quad + \quad \epsilon
$$

Models: GLM in Matrix Form

- Matricies yeild a consise statement of mulitple regression
  - ⇀ This form is so useful, so general, it's called...
- General Linear Model
  - ⇀ Now can concisely write β's that satisfy least squares principal
  - ⇀ $\hat{\beta} = (X'X)^{-1}X'Y$
  - ⇀ In fact, easy to derive

    $$Y = X\beta$$

Models: GLM & Basic Stats

- General Linear Model encompases many statandard statistics
  - ⇀ One sample t-test
  - ⇀ Two sample t-test
  - ⇀ Paired t-test
  - ⇀ Anova
  - ⇀ AnCova
  - ⇀ Correlation
  - ⇀ Linear regression
  - ⇀ Multiple regression
  - ⇀ F-tests

Models: GLM & Simple Correlation

Covariate, $x_1, x_2, \ldots, x_n$.

$$
X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_i \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}
$$

Implies a model of $y_i = \beta_1 + \beta_2 x_i$

Models: GLM & AnCova

- AnCova - Analysis of Covariance
  - ⇀ For a discrete and a continuous predictor
- Response Time Example (again)
  - ⇀ Discrete: Two groups, n_1 young n_2 old
  - ⇀ Continuous: Accuracy measure for each subject (x_i)

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

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 βs
  - ⇀ Can performan hypothesis test on $c\hat{\beta}$
- AnCova Example
  - ⇀ Interest was in Age effect
  - ⇀ Appropriate contrast
    $c = [0 \;\; 1 \;\; 0]$
    $c\beta = \beta_2$

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

$$X = \begin{bmatrix} 1 & 1 & 0 & x_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & x_{n_1} \\ 1 & 0 & 1 & x_{n_1+1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & x_{n_1+n_2} \end{bmatrix}$$

- ⇀ 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!

Models: GLM & Estimability

- An estimable contrast is guarunteed to have a unique value
  - ⇀ Even *if* an infinte number of β's yeild 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

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 principal
  - ⇀ But not GLM
  - ⇀ Requires nonlinear optmization

Models: Quiz

- Following from the fundamentals quiz, let $Y = (a_1, \ldots, a_{20}, b_1, \ldots, b_{20})^\top$.

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

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?

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

Inference: Linear Algebra & Random Variables

- The mean or expected value operates elementwise.

$$\mathsf{E}(Y) = \mathsf{E}\left(\begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix}\right) = \begin{bmatrix} \mathsf{E}(y_1) \\ \vdots \\ \mathsf{E}(y_i) \\ \vdots \\ \mathsf{E}(y_n) \end{bmatrix}$$

- Expectation is linear
  - ⇀ $\mathsf{E}(AY + \epsilon) = A\mathsf{E}(Y) + \mathsf{E}(\epsilon)$
    Where A is a fixed matrix

Inference: Linear Algebra & Random Variables

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

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

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

Inference: GLM & Normality

- Normality is a magical distribution
  - ⇀ Averages of *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

    $\mathcal{N}(,)$ is the Normal distribution

    ϵ_i and ϵ_j independent for $i \neq j$
  - ⇀ σ^2 is the variance, the magnitude of randomness

Inference: GLM & Normality

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

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!

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: $\mathsf{E}(Y) = \mathsf{E}(X\beta + \epsilon) = X\beta + \mathbf{0}$
  - ⇀ Var of Y: $\mathsf{Var}(Y) = \mathsf{Var}(X\beta + \epsilon) = \mathbf{0} + \sigma^2 I_n$

    $Y \sim \mathcal{N}(X\beta, \sigma^2 I_n)$

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 $\hat{\beta}$:

$\mathsf{E}(\hat{\beta}) = (X'X)^{-1}X'\mathsf{E}(Y) = (X'X)^{-1}X'X\beta = \beta$

Variance of $\hat{\beta}$?

$$\begin{aligned}
\mathsf{Var}(\hat{\beta}) &= \mathsf{Var}\left((X'X)^{-1}X'Y\right) \\
&= (X'X)^{-1}X'\mathsf{Var}(Y)\left((X'X)^{-1}X'\right)' \\
&= (X'X)^{-1}X'(\sigma^2 I_n)\left((X'X)^{-1}X'\right)' \\
&= (X'X)^{-1}X'(\sigma^2 I_n)X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}
\end{aligned}$$

Inference: GLM & Normality

- All of that yeilds...

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

Inference: GLM & t-test

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

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

  - 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

Inference: GLM & t Distribution

- Why not Normal?
  A t ratio isn't Normal b/c σ^2 isn't known D:tvN

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

Inference: GLM & Extra Sums of Squares

- Often more than one possible model
- How to compare?
- Residual Sum of Squares (RSS)
  - $\sum_i e_i = \sum_i(y_i - \hat{y}_i)^2 = (Y - \hat{Y})'(Y - \hat{Y})$

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?

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

Inference: GLM & F-test

- Contrast representation of F tests
  - ⇀ Can define F-test with contrasts
  - ⇀ One contrast for each column to join null model
  - ⇀ Or, a collection of contrasts to be simultaneously tested as zero
- 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

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: ith subject's laser energy
- Use F test to compare RSS of this model with previous

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".