

A decorative graphic on the left side of the slide, consisting of a network of white lines and small circles on a dark blue background, resembling a circuit board or neural network.

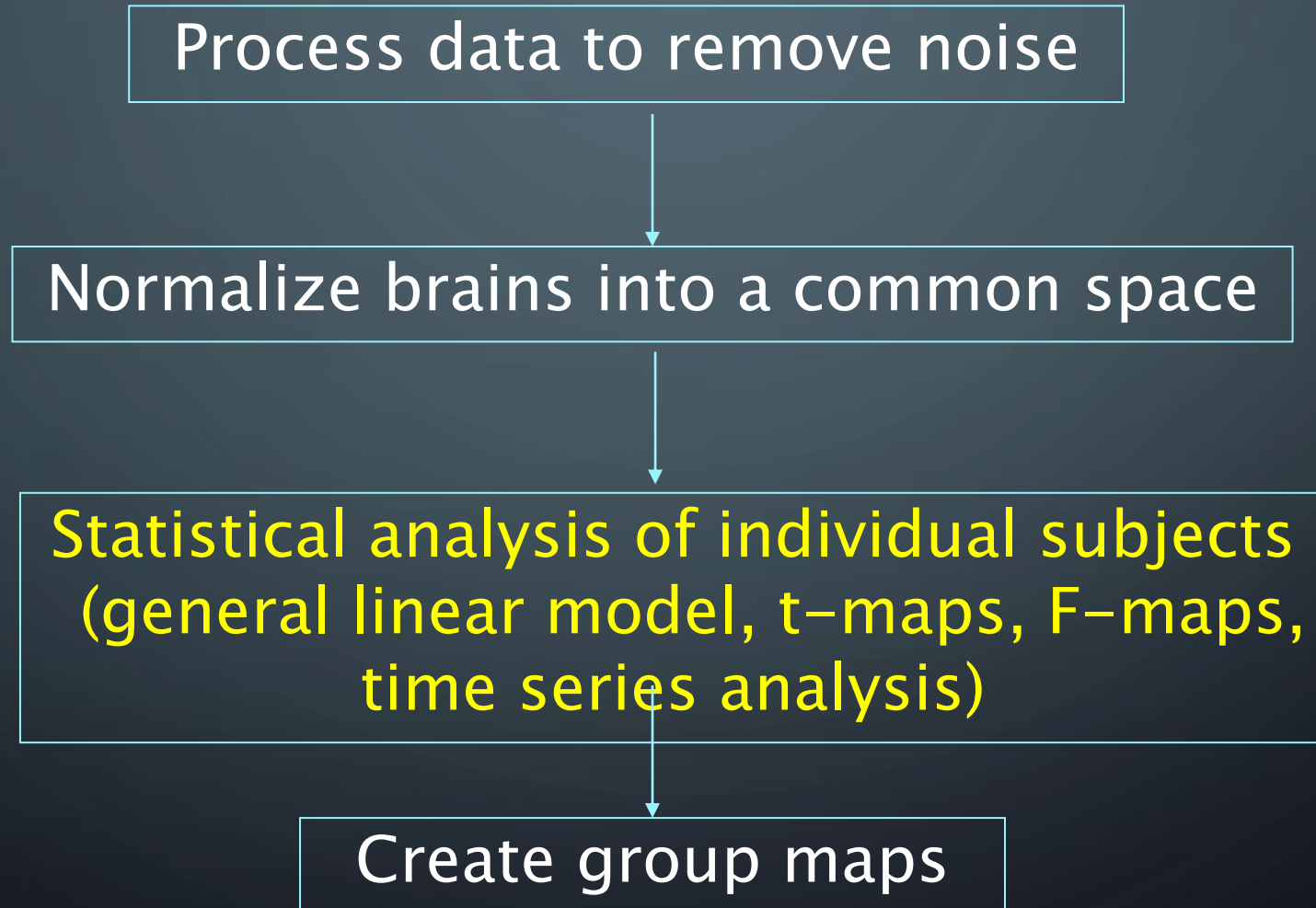
LECTURE 10: MODELING THE FMRI DATA AT THE FIRST LEVEL

NOVEMBER 15, 2016

OVERVIEW

1. Overview/History of fMRI
2. How it Works /Basic Principles
 - The Blood Oxygen–Level Dependent Signal (BOLD)
 - How the BOLD signal is measured
1. How to collect the data/ Design of experiments
2. How to analyze the data
3. Group analysis
4. Limitations and problems

EXAMPLE OF FMRI DATA ANALYSIS STEPS

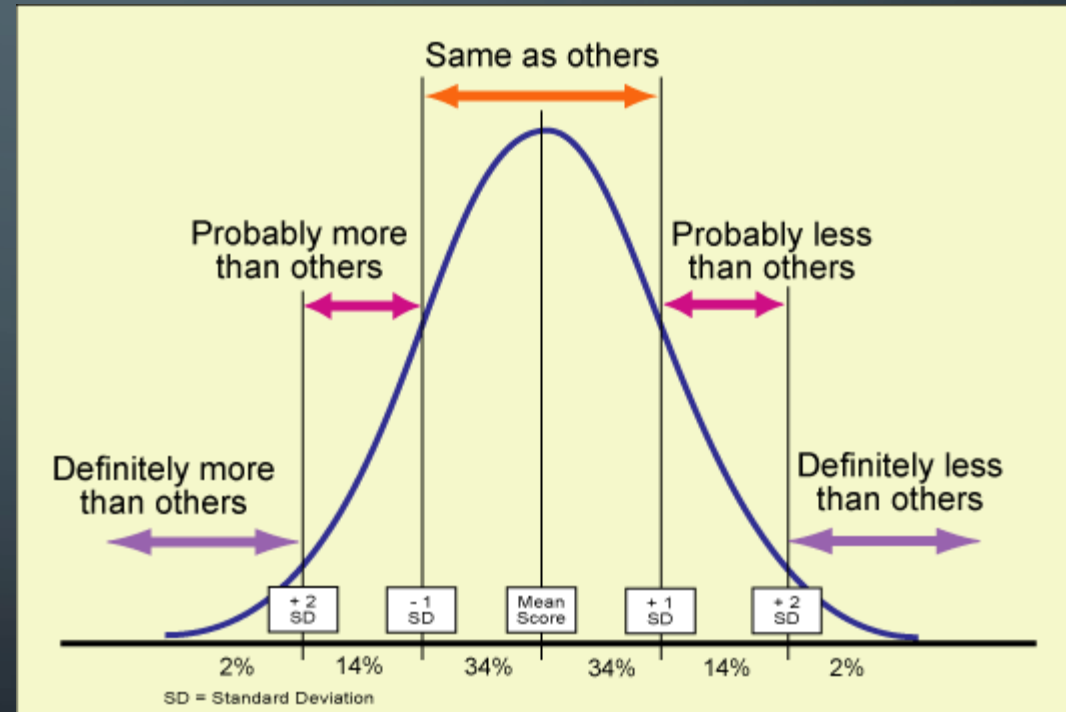


WHAT CONSTITUTES ACTIVATION

- Statistics help us to answer the following:
 - How do we determine whether an area of the brain is activated by our task?
 - How confident are we that the areas we find are activated by our task?
 - Are the results in my group of subjects applicable to the wider population?
 - Are the differences between groups of subjects significant?

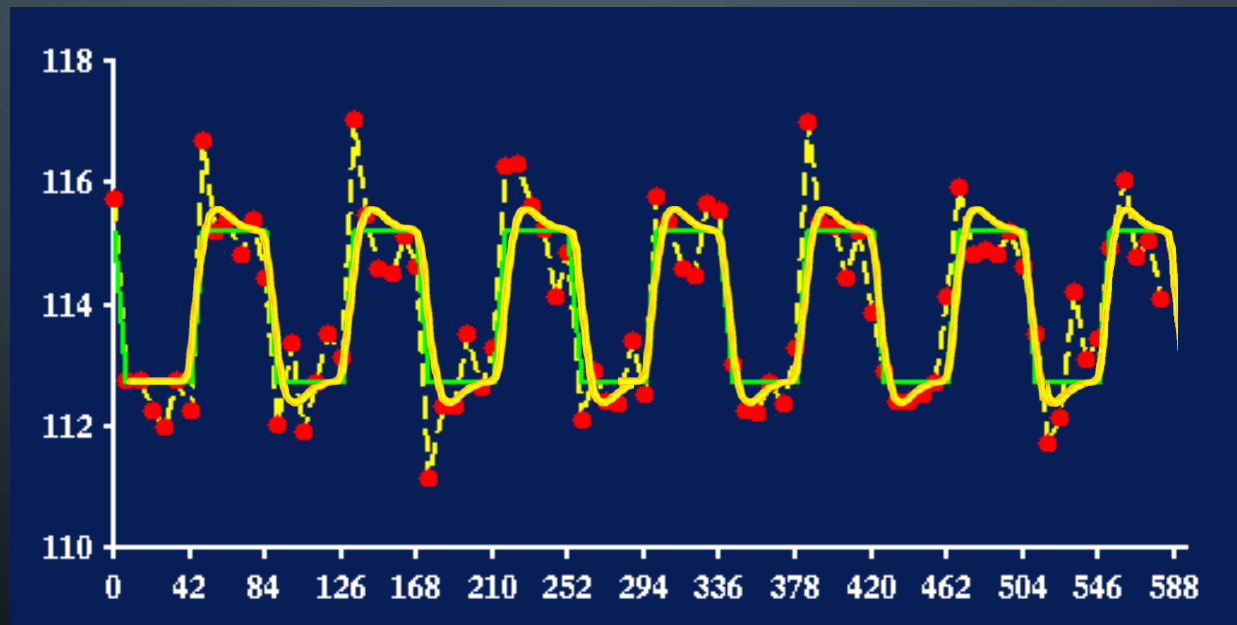
“BRAIN ACTIVATION”

- We fit the general linear model (GLM) to each voxel in the brain
- Use the estimated model parameter to test for “activation”
- “Activation” means the voxel exceeded a certain threshold



CORRELATION/REGRESSION

- Use a representative waveform representing the on-off periodicity of the design convolved with HRF and correlate that with MRI signal change across each of the scans



GENERAL LINEAR MODEL & FMRI

How does GLM apply to fMRI experiments?

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

Observed = Predictors * Parameters + Error

BOLD = Design Matrix * Betas + Error

Elliot Freeman, ICN, "Idiot's guide to the general linear model & fMRI. fMRI model, Linear Time Series, Design Matrices, Parameter estimation, *&%@!"

SIMPLE FMRI GLM

- $x(t)$ is the block design convolved with a model of the HRF

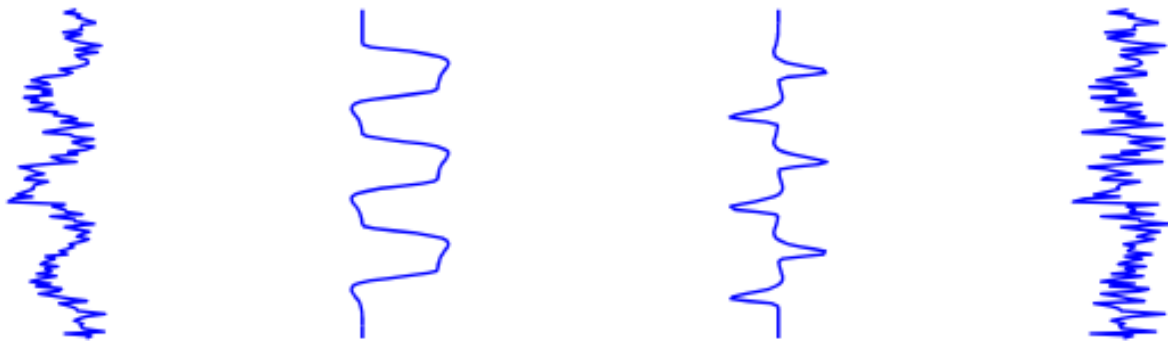
$$y(t) = \beta x(t) + \varepsilon(t)$$



GLM – ADDING REGRESSORS

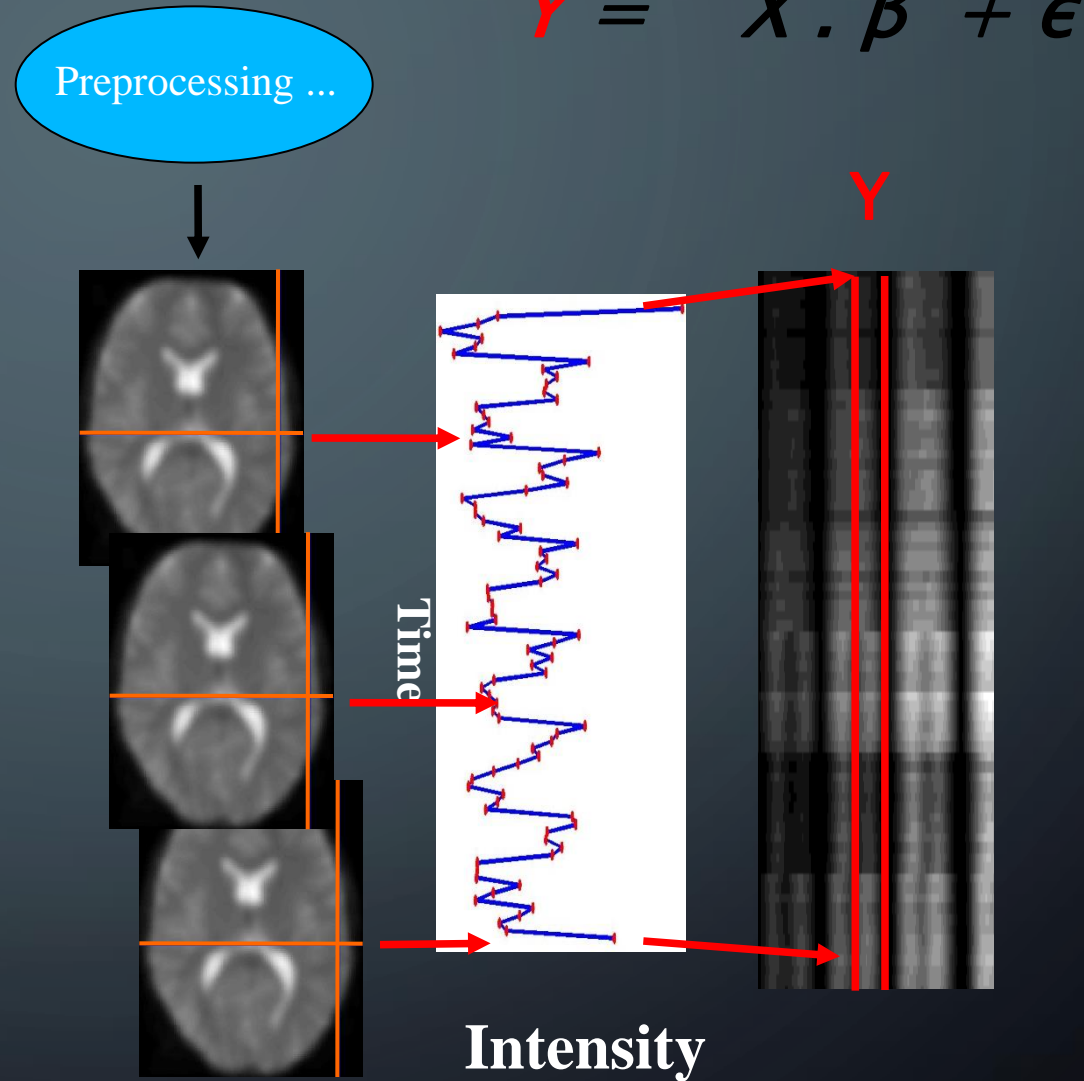
- $x_2(t)$ adds Temporal Derivatives – allows for shifts in transition points of block design

$$y(t) = \beta_1 x_1(t) + \beta_2 x_2(t) + \varepsilon(t)$$



Observed data

- Y is a matrix of BOLD signals:
- Each column represents a single voxel sampled at successive time points.

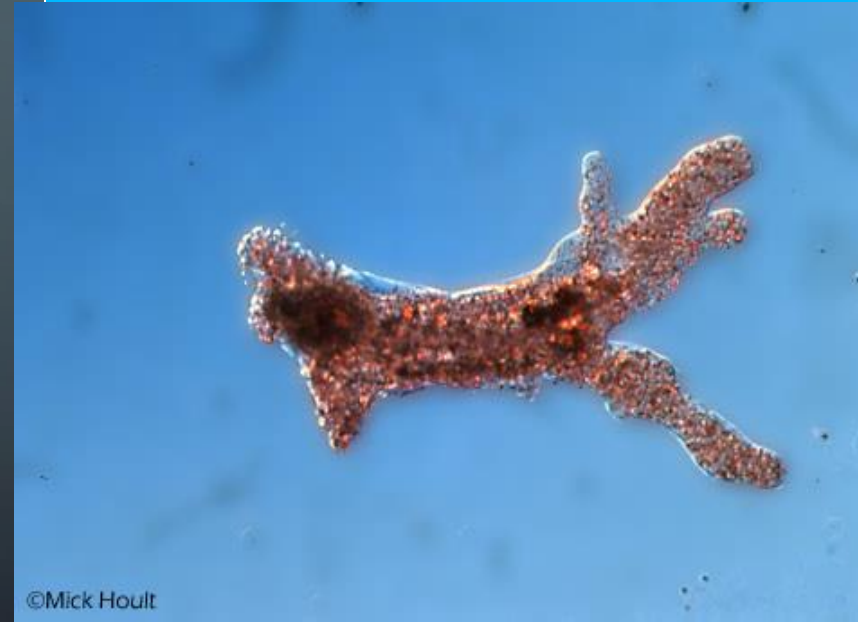
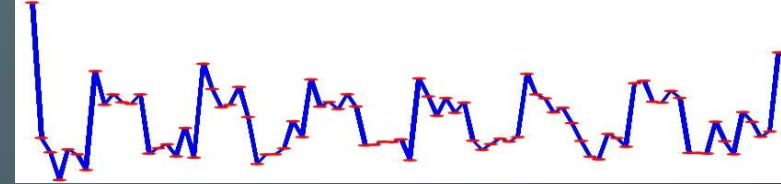


Elliot Freeman, ICN, "Idiot's guide to the general linear model & fMRI. fMRI model, Linear Time Series, Design Matrices, Parameter estimation, *&%@!"

UNIVARIATE ANALYSIS

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

- Each voxel considered as independent observation
- Analysis of individual voxels over time, not groups over space
- SPM would still work on an Amoeba!

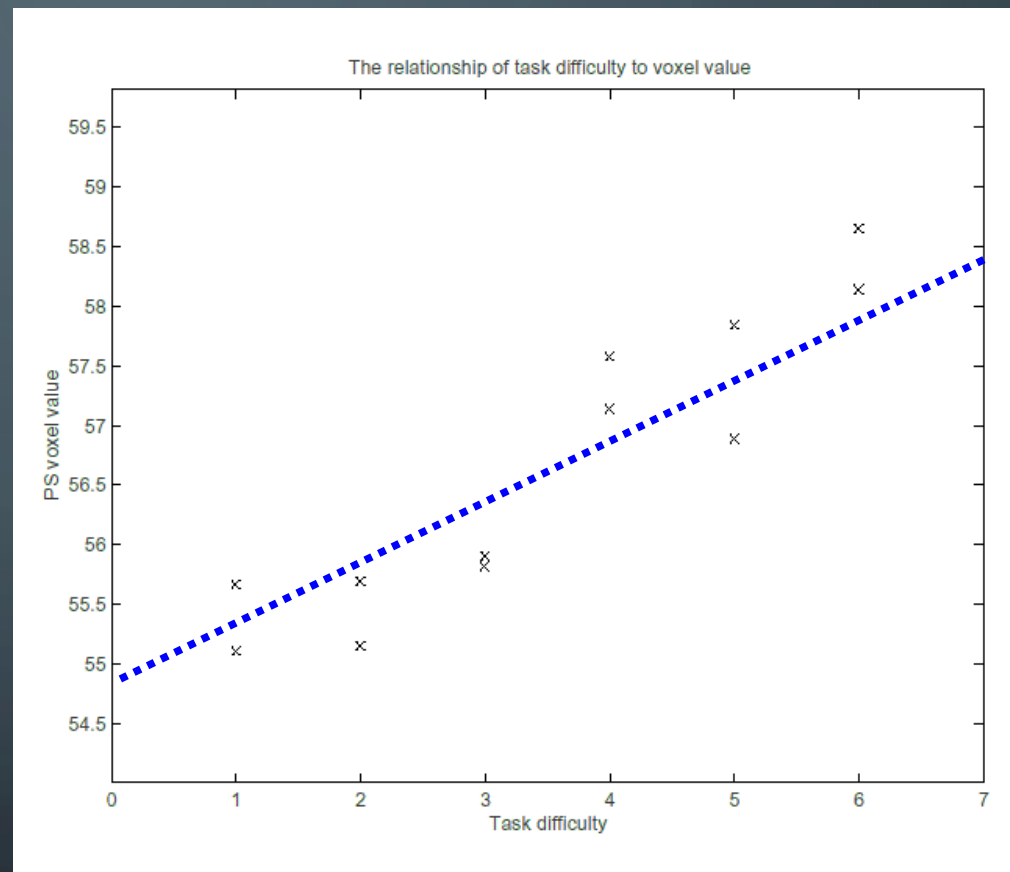


Elliot Freeman, ICN, "Idiot's guide to the general linear model & fMRI. fMRI model, Linear Time Series, Design Matrices, Parameter estimation, *&%@!"

CONTINUOUS PREDICTORS

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

	Y	X
Scan no	Voxel 1	Task difficulty
1	57.84	5
2	57.58	4
3	57.14	4
4	55.15	2
5	55.90	3
6	55.67	1
7	58.14	6
8	55.82	3
9	55.10	1
10	58.65	6
11	56.89	5
12	55.69	2



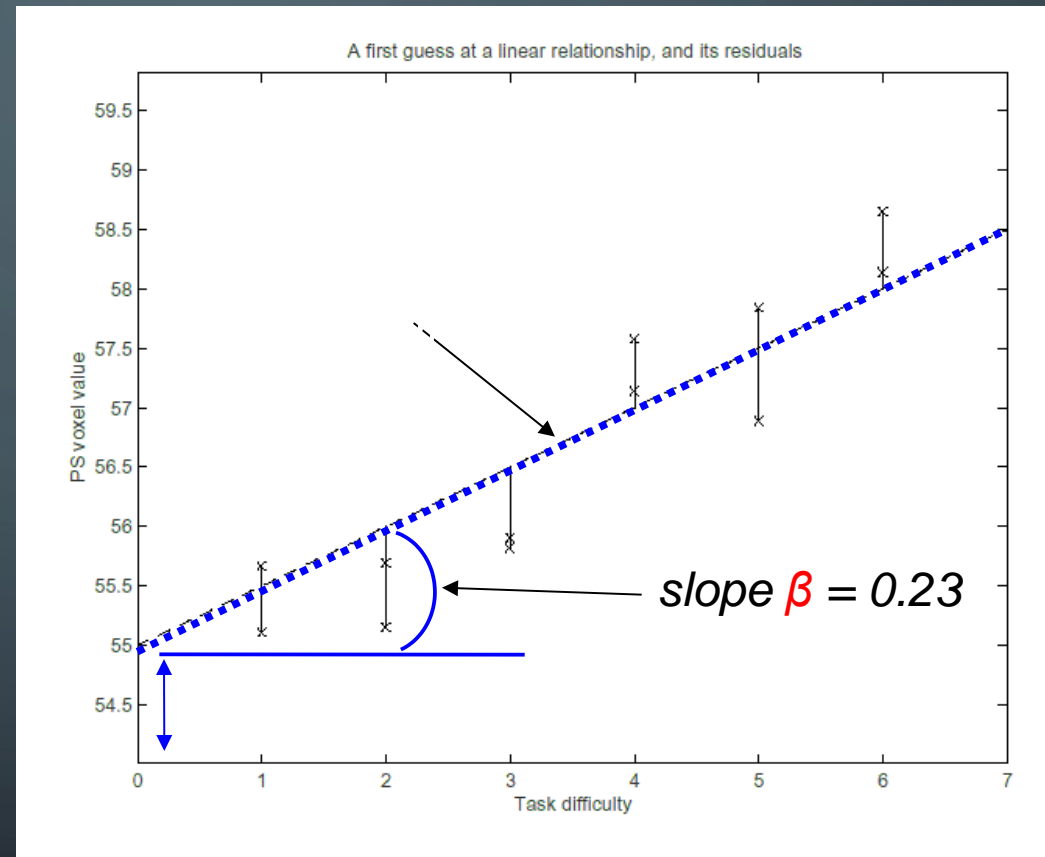
X can contain values quantifying experimental variable

Elliot Freeman, ICN, "Idiot's guide to the general linear model & fMRI. fMRI model, Linear Time Series, Design Matrices, Parameter estimation, *&%@!"

Parameters & error

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

- β : slope of line relating X to Y
 - ‘how much of X is needed to approximate Y ?’
 - the best estimate of β minimizes ϵ : deviations from line

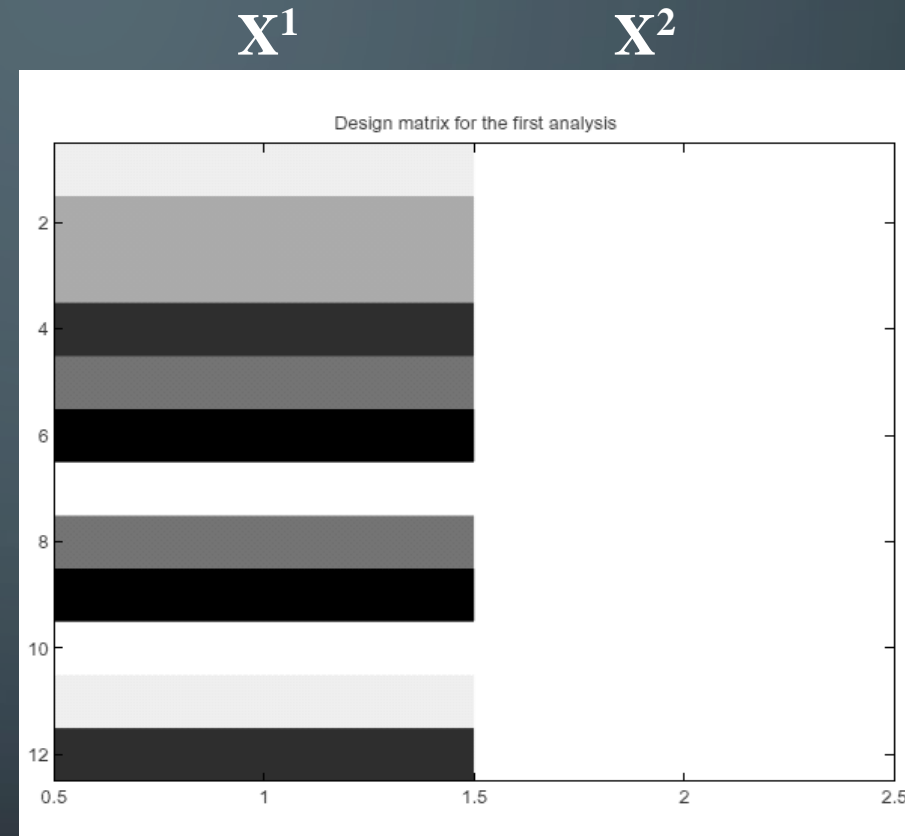


Elliot Freeman, ICN, “Idiot’s guide to the general linear model & fMRI. fMRI model, Linear Time Series, Design Matrices, Parameter estimation, *&%@!”

DESIGN MATRIX

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

	Y	X ¹	X ²
Scan no	Voxel 1	Task difficulty	Constant variable
1	57.84	5	1
2	57.58	4	1
3	57.14	4	1
4	55.15	2	1
5	55.90	3	1
6	55.67	1	1
7	58.14	6	1
8	55.82	3	1
9	55.10	1	1
10	58.65	6	1
11	56.89	5	1
12	55.69	2	1



Matrix represents values of X

Different columns = different predictors

Matrix formulation

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

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{pmatrix} = \begin{pmatrix} X^1_{(t1)} & X^2_{(t1)} \dots & X^L_{(t1)} \\ X^1_{(t2)} & X^2_{(t2)} \dots & X^L_{(tS)} \\ \vdots & \vdots & \vdots \\ X^1_{(tN)} & X^2_{(tN)} \dots & X^L_{(tN)} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_L \end{pmatrix} + \begin{pmatrix} \epsilon_{(t1)} \\ \epsilon_{(t2)} \\ \vdots \\ \epsilon_{(tN)} \end{pmatrix}$$

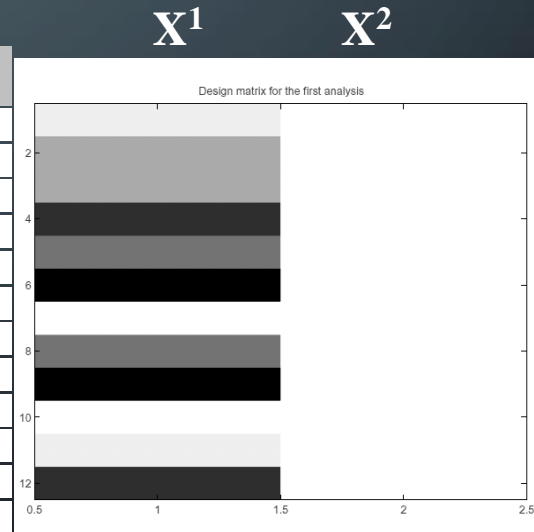
$$\hat{Y}_1 = (5 * \beta_1) + (1 * \beta_2)$$

$$\hat{Y}_2 = (4 * \beta_1) + (1 * \beta_2)$$

...

$$\hat{Y}_N = (X^1_{(tN)} * \beta_1) + (X^2_{(tN)} * \beta_2)$$

(t)	Y	X ¹	X ²
time (t)	Voxel 1	Task difficulty	Constant variable
1	57.84	5	1
2	57.58	4	1
3	57.14	4	1
4	55.15	2	1
5	55.90	3	1
6	55.67	1	1
7	58.14	6	1
8	55.82	3	1
9	55.10	1	1
10	58.65	6	1
11	56.89	5	1
12	55.69	2	1



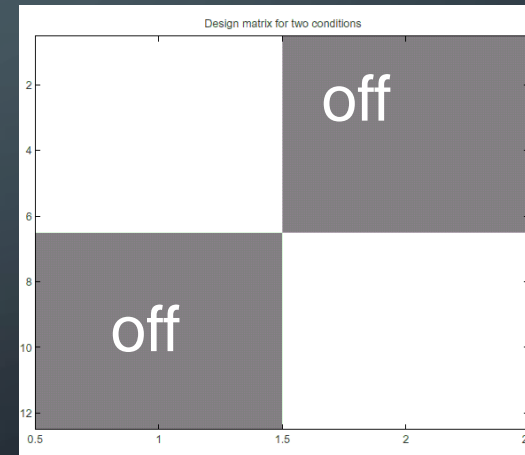
Elliot Freeman, ICN, "Idiot's guide to the general linear model & fMRI. fMRI model, Linear Time Series, Design Matrices, Parameter estimation, *&%@!"

Parameter estimation and stats

- Find betas (by least squares estimation)
 - $Y = \beta X \rightarrow "B = Y / X"$ (B= estimated β)
 - Matlab magic:
`>> B = inv(X) * Y`
- Now find error term:
 - $e = Y - (X * B)$
- ...and use these results for statistics:
 - $t = \text{betas} / \text{standard error}$

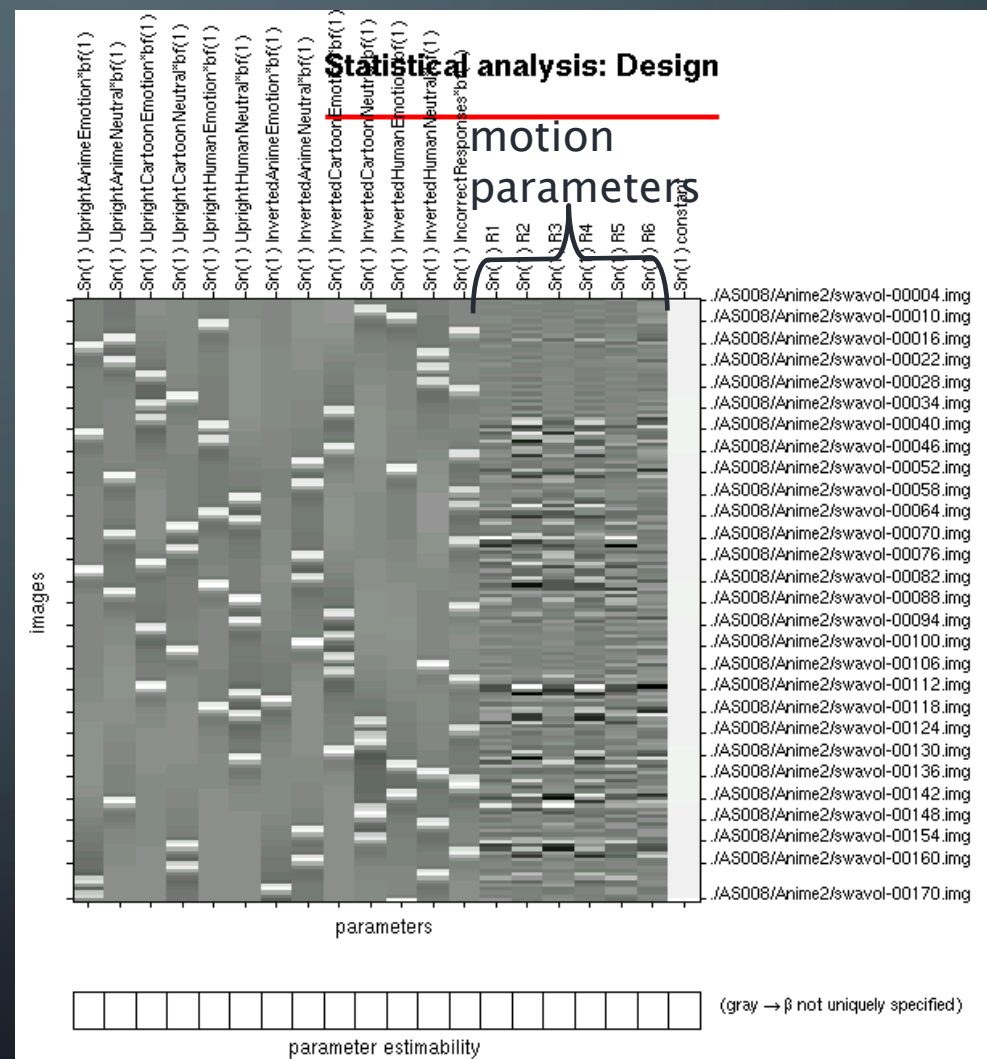
Covariates vs. conditions

- **Covariates:**
 - parametric modulation of independent variable
 - e.g. task-difficulty 1 to 6
– regression: $\beta = \text{slope}$
- **Conditions:**
 - specify time of onset and duration
 - e.g. integers 0 or 1: 'off' or 'on'
–> ANOVA: $\beta = \text{effect mean}$



DESIGN MATRIX WITH MOTION AS REGRESSORS

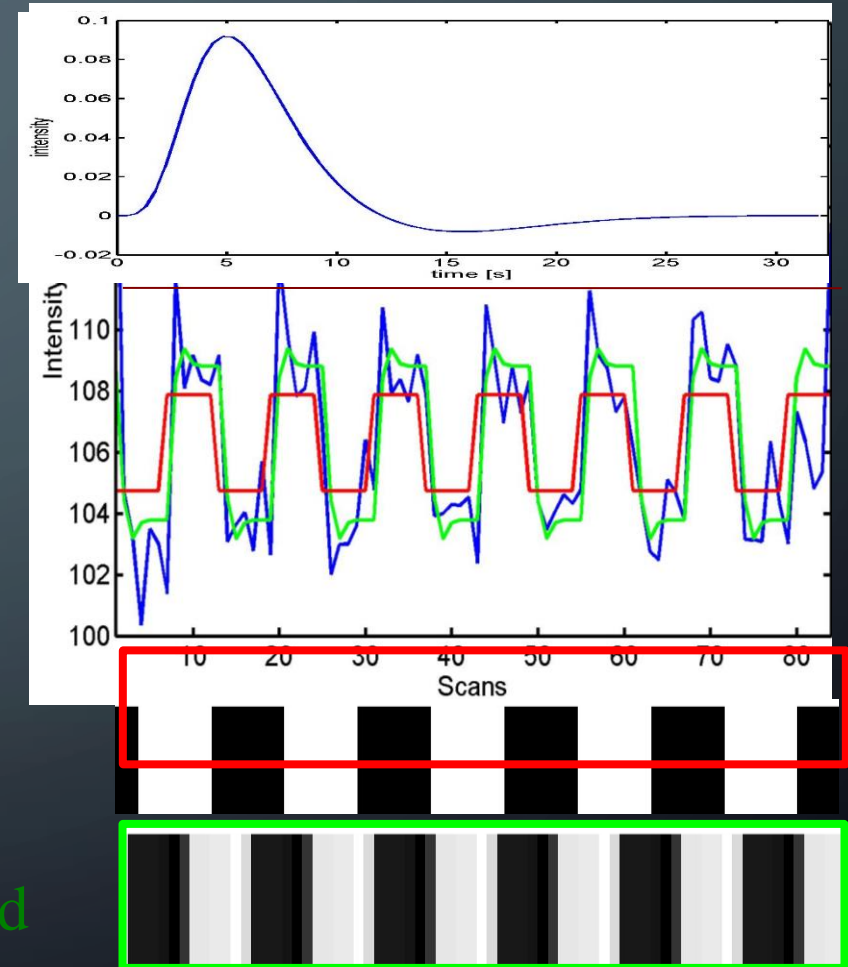
- Example includes transformation parameters from motion correction as nuisance regressors
 - Useful for removing noise due to head motion
 - However, if motion is correlated with task then this will reduce statistical significance



MODELLING HEMODYNAMICS

HRF basic
function

- Brain does not just switch on and off!
- Reshape (convolve) regressors to resemble HRF



Original

HRF
Convolved

Elliot Freeman, ICN, "Idiot's guide to the general linear model & fMRI. fMRI model, Linear Time Series, Design Matrices, Parameter estimation, *&%@!"

SUMMARY: REVERSE COOKERY

- You start with the finished product and want to know how it was made
 - You specify which *ingredients* to add (*design matrix variables*)
 - For each ingredient, GLM finds the *quantities* (*betas*) that produce the best reproduction (*model*)
 - Now you can compare your recipe with others (*null hypothesis*) to see if they differ! (*statistical tests*)



Elliot Freeman, ICN, “Idiot’s guide to the general linear model & fMRI. fMRI model, Linear Time Series, Design Matrices, Parameter estimation, *&%@!”

WHAT CONSTITUTES ACTIVATION

- Statistics help us to answer the following:
 - How do we determine whether an area of the brain is activated by our task?
 - How confident are we that the areas we find are activated by our task?
 - Are the results in my group of subjects applicable to the wider population?
 - Are the differences between groups of subjects significant?

FORMAL STATEMENT OF A HYPOTHESIS

- Research question is framed as a hypothesis
- Null hypothesis assumes that the hypothesis is not true
- Statistics aim to disprove the null hypothesis and thus accept the research hypothesis
- In fMRI we are testing difference in BOLD signal between two conditions

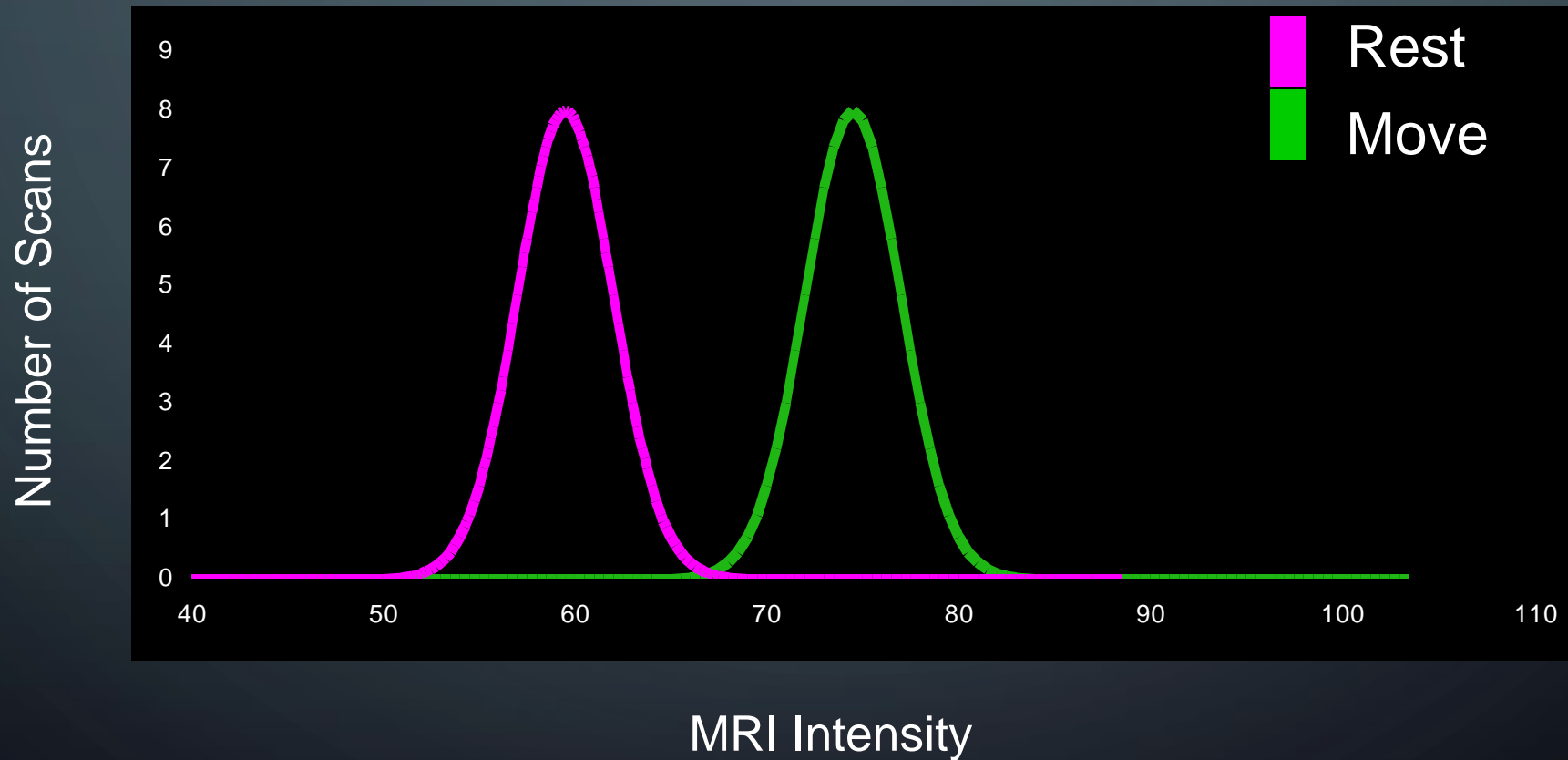
$H_1: \text{Condition}_1 \neq \text{Condition}_2$

$H_0: \text{Condition}_1 = \text{Condition}_2$

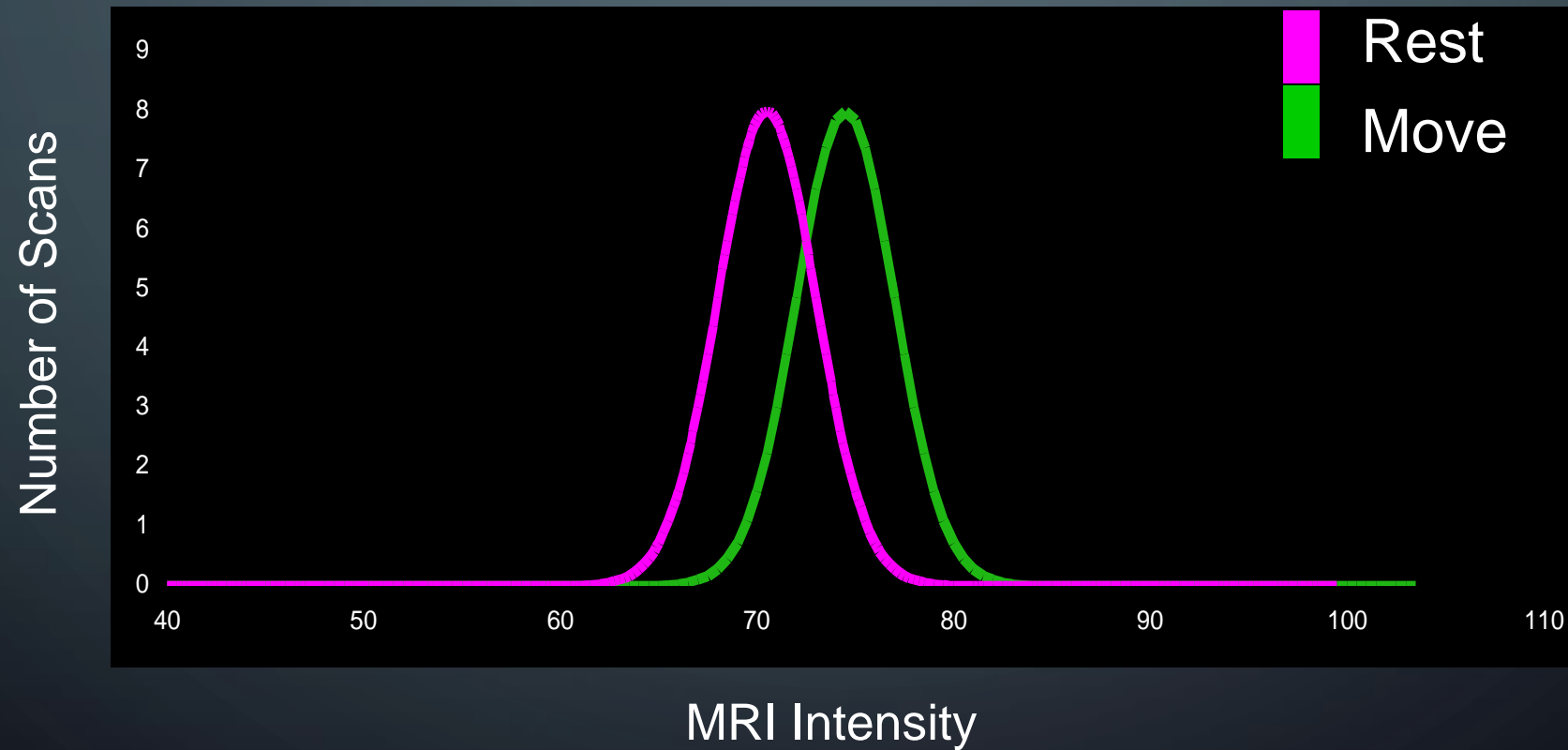
FMRI EXAMPLE

- Hypothesis: “Moving the thumb will cause an increase in neuronal activity which we detect with BOLD signal changes”
- Null Hypothesis: “Moving the thumb will *NOT* cause an increase in neuronal activity which we detect with BOLD signal changes”
 - Experimental condition – moving thumb
 - Control condition – thumb not moving
 - Outcome measure – MRI signal changes

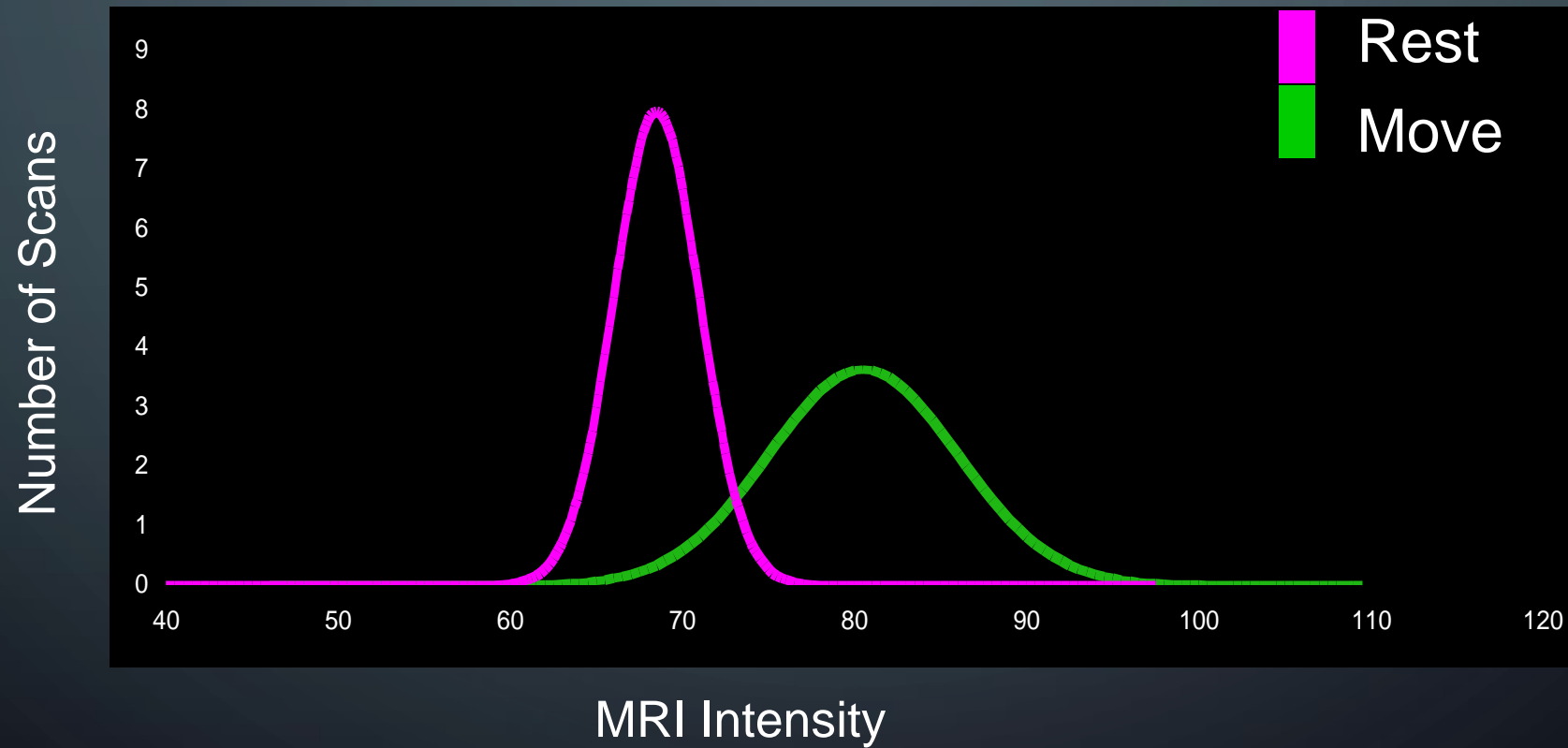
HYPOTHESIS TESTING



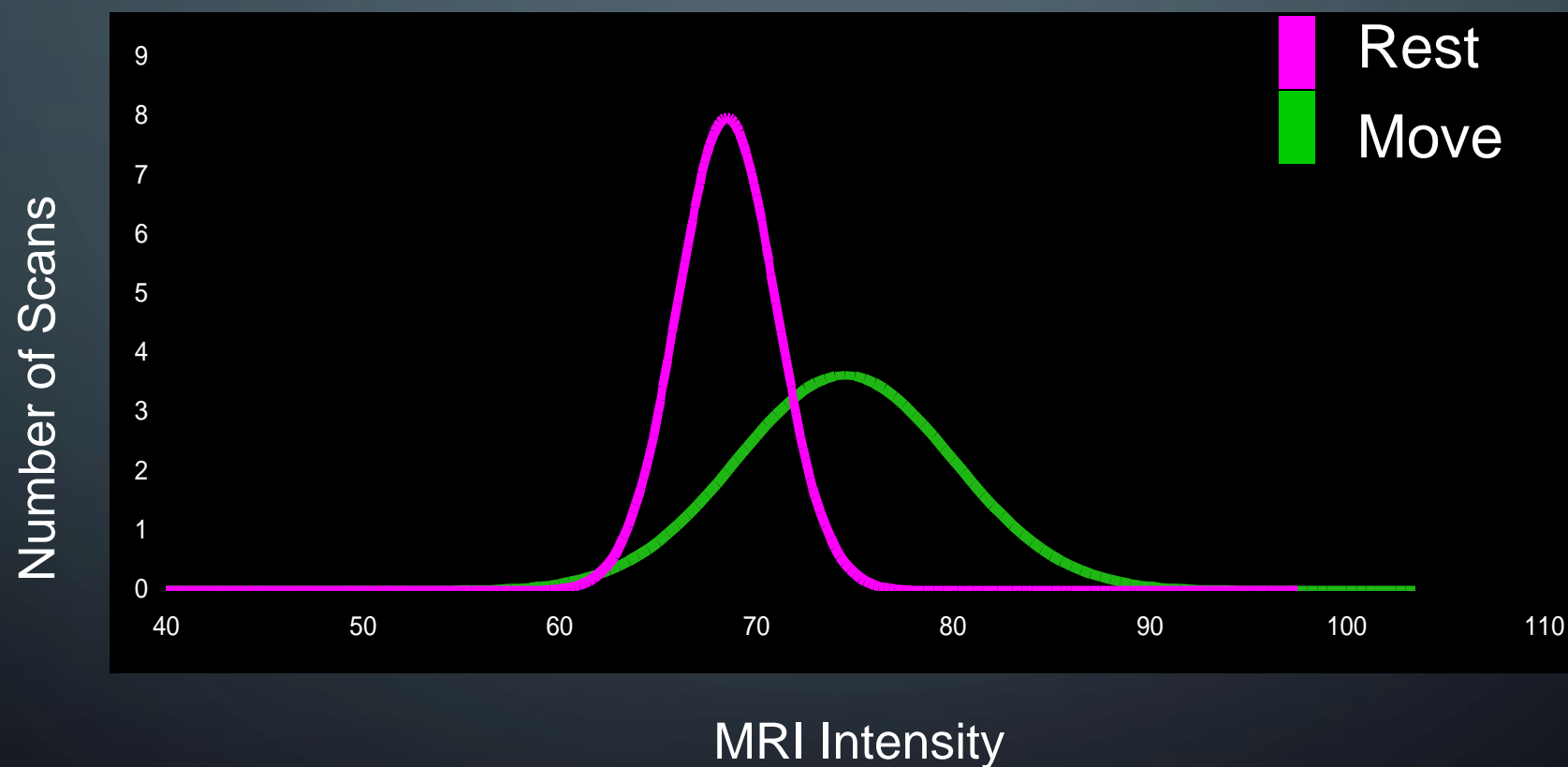
HYPOTHESIS TESTING



HYPOTHESIS TESTING



HYPOTHESIS TESTING

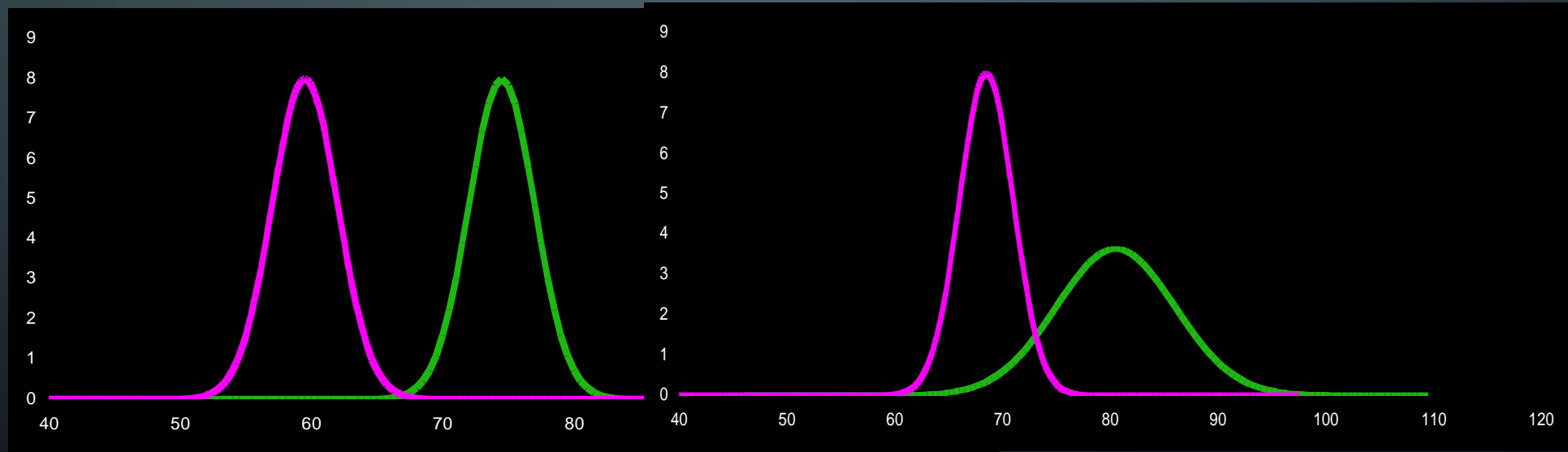


HYPOTHESIS TESTING

- Two factors describe how much effect the experimental condition had:
 - Difference between the mean intensities of each condition
 - Degree of overlap in intensities

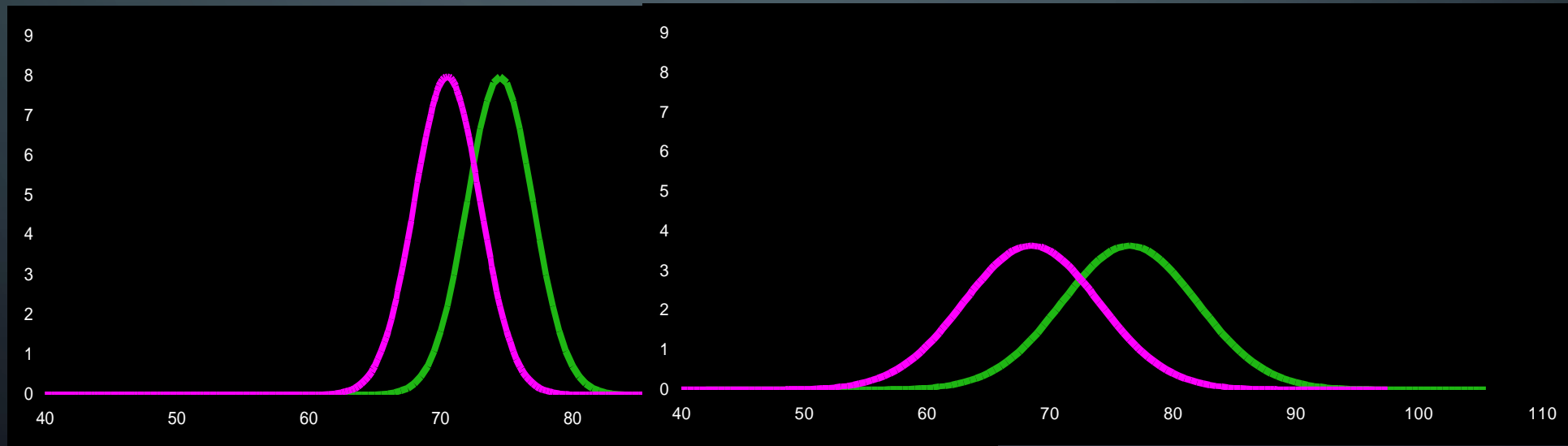
HYPOTHESIS TESTING

- Experimental condition has an effect



HYPOTHESIS TESTING

- Experimental condition has *no* effect



THE T-TEST

- Formally incorporates our intuitive sense of when there is an effect
- Based on a measure of the distance between the two means and the spread of each condition

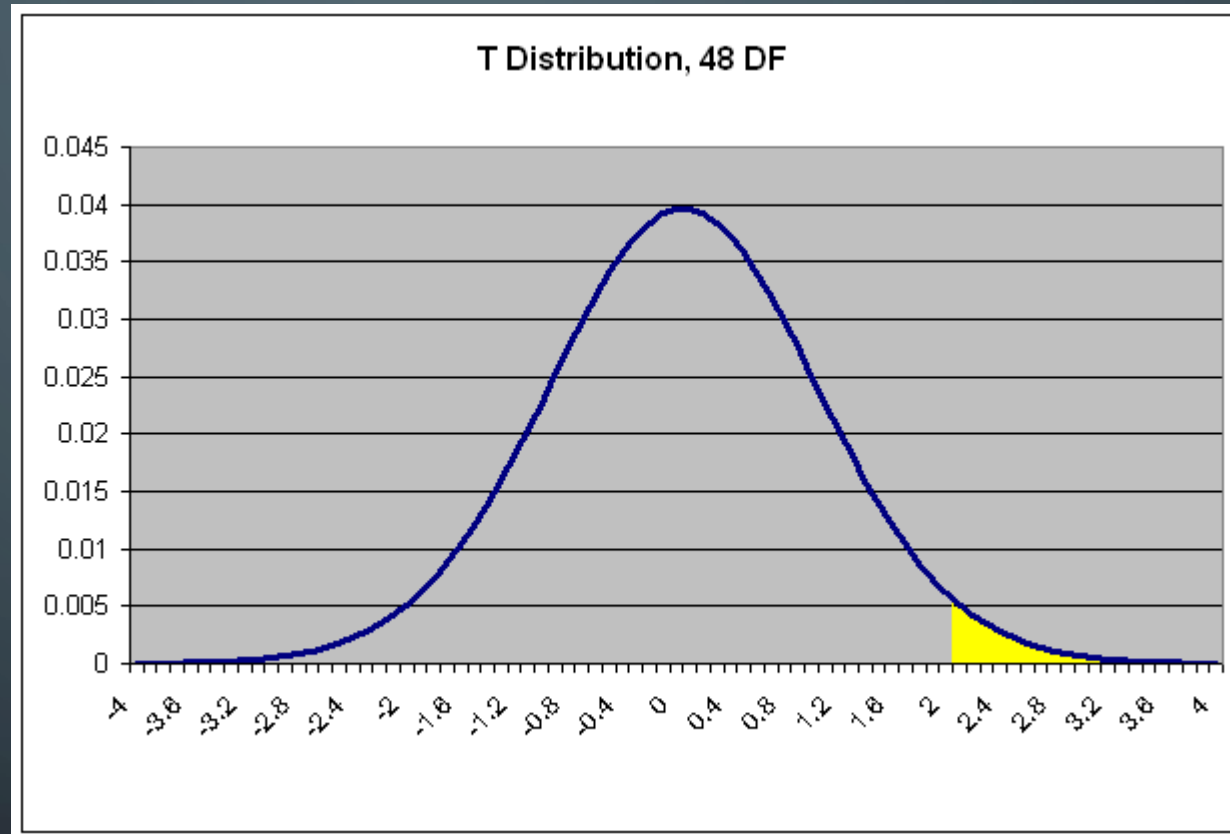
$$t = \frac{(m_1 - m_2)}{\sqrt{(\sigma_1^2 + \sigma_2^2)}}$$

- We use our Beta values in these statistical measures!

T-STATISTICS AND P-VALUES

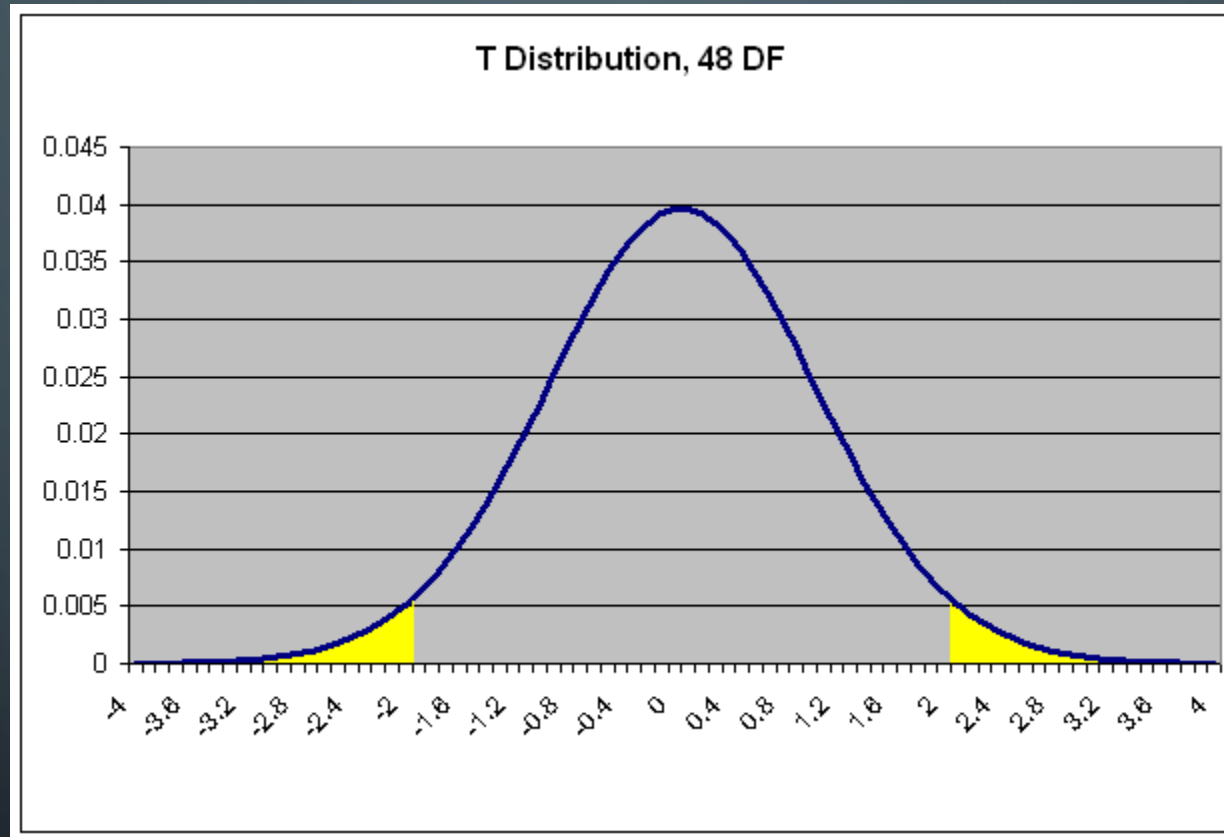
- The p-value for a t-statistic gives the probability that the difference between the experimental and control conditions arose by chance
- Typically $p < 0.05$ is considered minimum cut-off for significance (i.e. alpha is set at $p < 0.05$)
- Statistics tables list the p-values for each t-statistic based on the df, degrees of freedom, (single subject analysis $df = \text{total number of scans} - 1$)

ONE-TAILED TEST



Yellow area under the curve is about 0.025 (for $t=2$).

TWO-TAILED TEST



Yellow area under the curve is about 0.05 (for $t=2$).

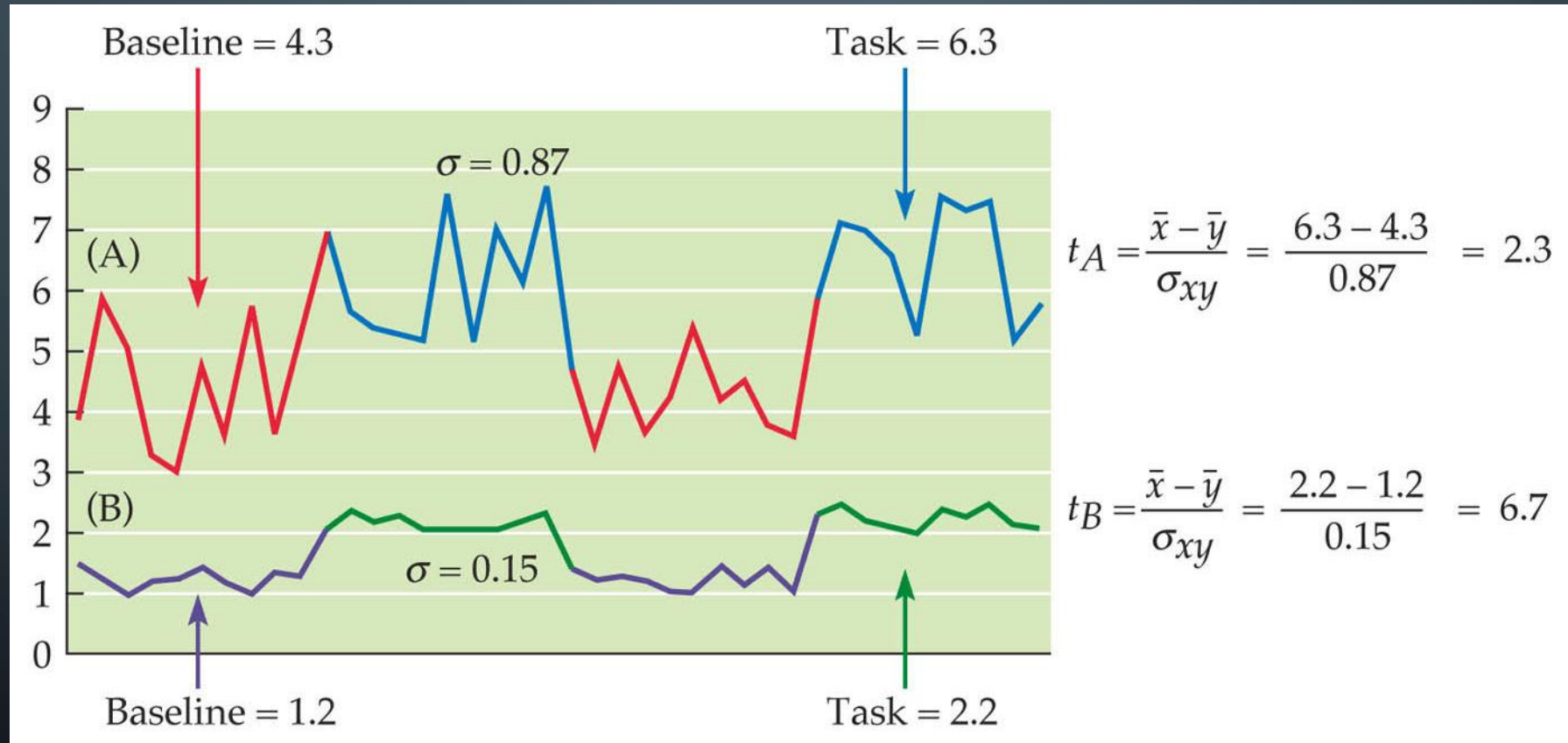
T-STATISTICS AND P-VALUES

Ex.

Suppose we find a voxel in which the t-statistic is 3.3 and there were 20 scans at rest and 20 scans while moving thumb.

The probability that the difference in the MRI signal in this voxel *is not due* to the movement of the thumb is 0.002 two-tail or 0.001 one-tail.

EXAMPLES OF THE AFFECT OF VARIANCE



T-STATISTICS, P-VALUES, & Z-SCORES

- Unlike t-statistics, the p-value for Z-scores, which are based on the normal distribution, does not change depending on the number of scans
- In functional imaging it is common to convert the t-statistic to a Z-score since it is easier to compare across studies (not dependent on degrees of freedom)

PROBLEMS WITH THE T-TEST

- Systematic differences such as artifacts can create apparent significant differences where none exists
- Disregards any temporal characteristics of the data since only means are compared
- Assumption of t-test is that the data for both conditions is normally distributed – usually though not always true
 - Smoothing helps make data normally distributed

MAKING ERRORS

- Two types of errors:
 - Type 1: Activation is true, but we mistakenly reject it (False positive)
 - Type II: Activation is false, but we fail to reject it (False negative)

	Disease or Condition	No Disease or Condition
Test Positive	A True Positive	B False Positive
Test Negative	C False Negative	D True Negative

MULTIPLE COMPARISON PROBLEM

- If more than one hypothesis test is performed, chance of making errors is even greater
- The more tests performed, the greater the likelihood of errors
- Say we have 32 slices, 64x64 voxels in xy plane
- If we test $64 \times 64 \times 32 = 131,072$ comparison tests!!!
- Which voxels are actually significant?
- Want to balance sensitivity (true positive rate) and specificity (true negative rate)

MEASURE OF FALSE POSITIVES

- There are ways to quantify the amount of false positives
 - Family Wise Error Rate (FWER) – control the probability of false positives
 - False Discovery Rate (FDR) – control the proportion of false positives among rejected tests

FAMILY-WISE ERROR RATE

- Control the probability of making one or more Type I errors in a family of tests
- Basically adjusting p-values for the number of hypothesis tests performed
- FWER controlling methods include:
 - Bonferroni correction
 - Random Field Theory
 - Permutation tests

How to avoid false associations?

Applying m independent statistical tests with significance level α , a probability of at least one false association should be

$$1 - (1 - \alpha)^m < 0.05$$



$$\alpha = 1 - (1 - 0.05)^{1/m} \approx \frac{0.05}{m}$$

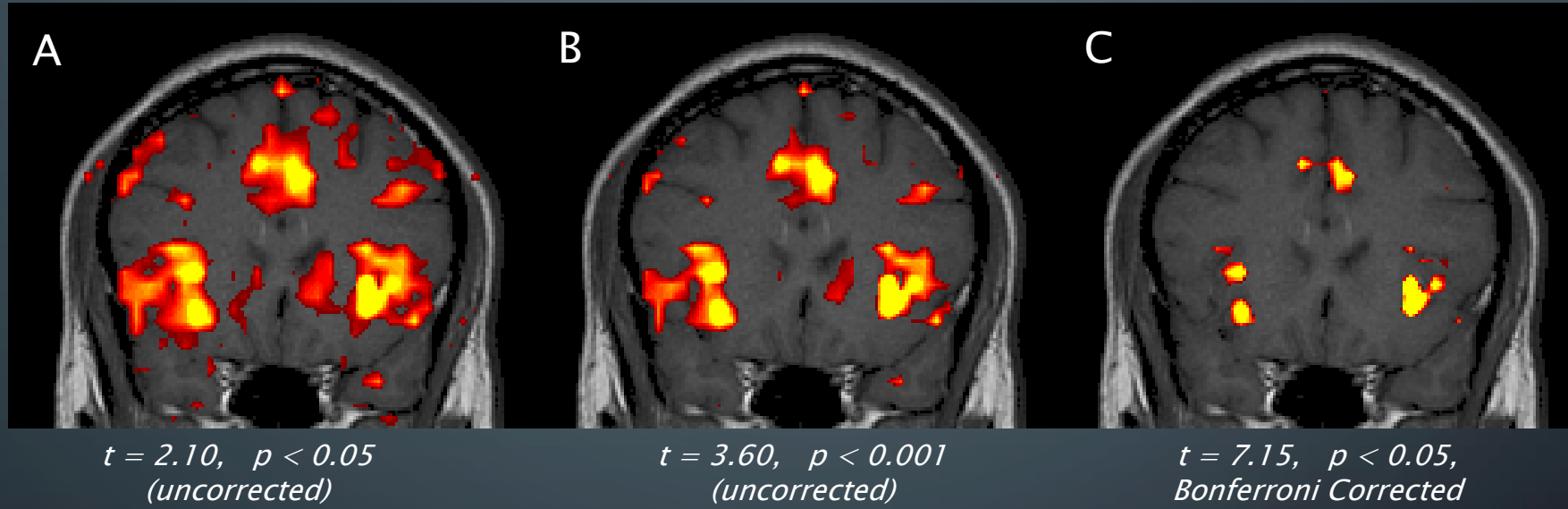
Carlo Bonferroni (1935):

When applying m independent statistical test, only significant results are results with

$$p < \frac{0.05}{m}$$

BONFERRONI CORRECTION

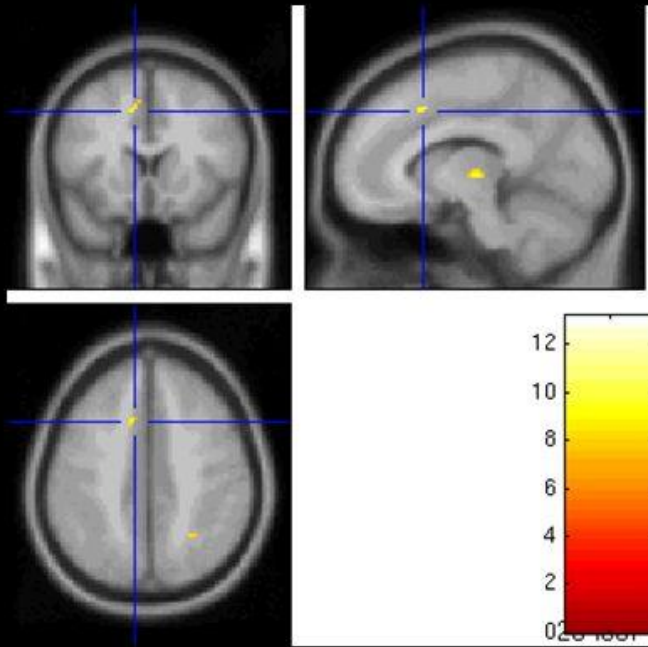
- Very simple method for ensuring that the overall Type I error rate of α is maintained when performing m independent hypothesis tests
- Rejects any hypothesis with $p\text{-value} \leq \alpha/m$:
- For example, if we want to have an experiment wide Type I error rate of 0.05 when we perform 10,000 hypothesis tests, we'd need a $p\text{-value}$ of $0.05/10000 = 5 \times 10^{-6}$ to declare significance
- Problem: Extremely conservative – often fail to find results



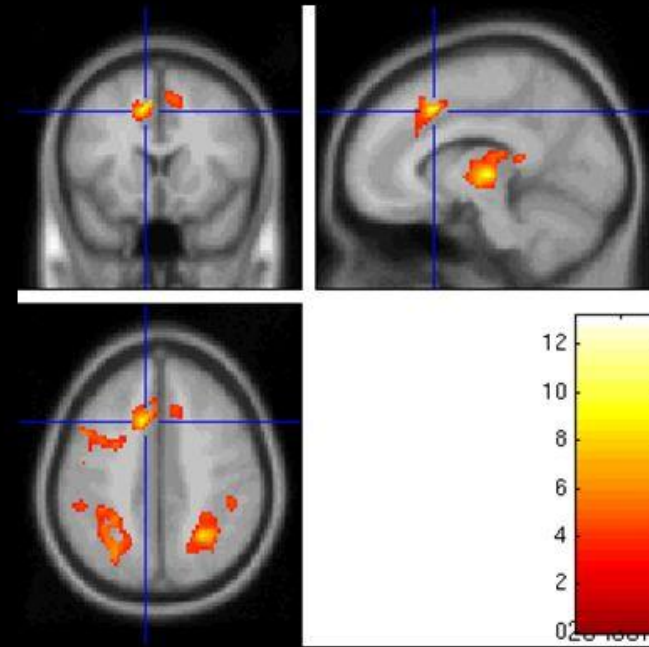
FALSE DISCOVERY RATE (BENJAMINI, HOCHBERG, 1995)

- Order tests according to p-value :
- $p_1 < p_2 < \dots < p_m$.
- For FDR control on α level (e.g. 0.05),
we find $j^* = \max \left\{ j : p_j \leq \frac{j}{m} \alpha \right\}$
- Differences are assumed to be significant for $j = 1, \dots, j^*$.
- For $j > j^*$ differences are assumed not to be significant

FWE vs FDR Working Memory Example



FWE Perm. Thresh. = 7.67
58 voxels



FDR Threshold = 3.83
3,073 voxels

ANALYSIS OUTPUTS

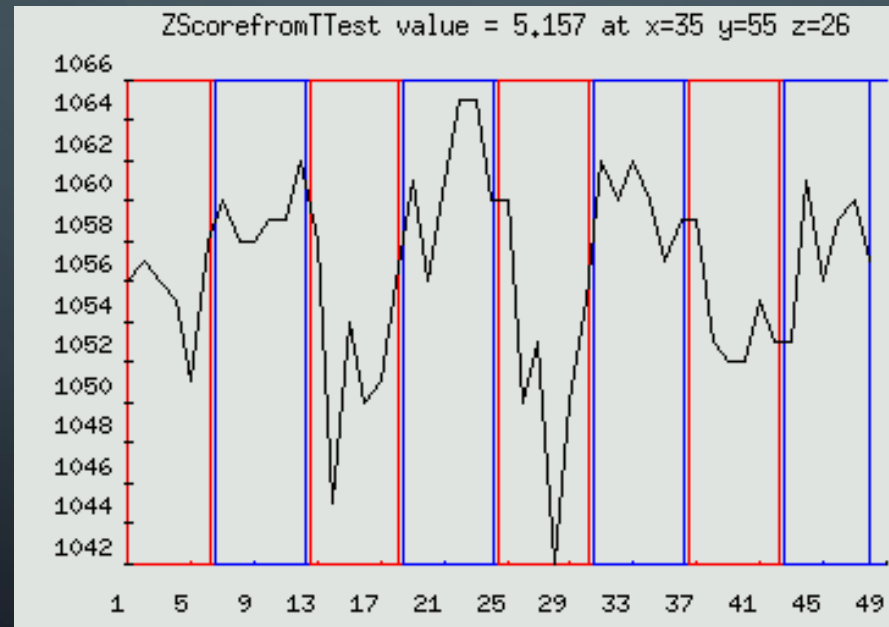
- Output from analysis software will typically include some visual representation of the results and tables of areas of activation
- Variety of tools available to interrogate and visualize results

LOCAL MAXIMA REPORTING

- Utility that generates a list of coordinates that correspond to the highest values in the statistical map grouped by 'cluster'
- Clusters are defined by spatially contiguous set of voxels above a statistical threshold (p-value)
- SPM reports the maxima within each cluster and up to 2 sub-maxima at least 8 mm from the other maxima

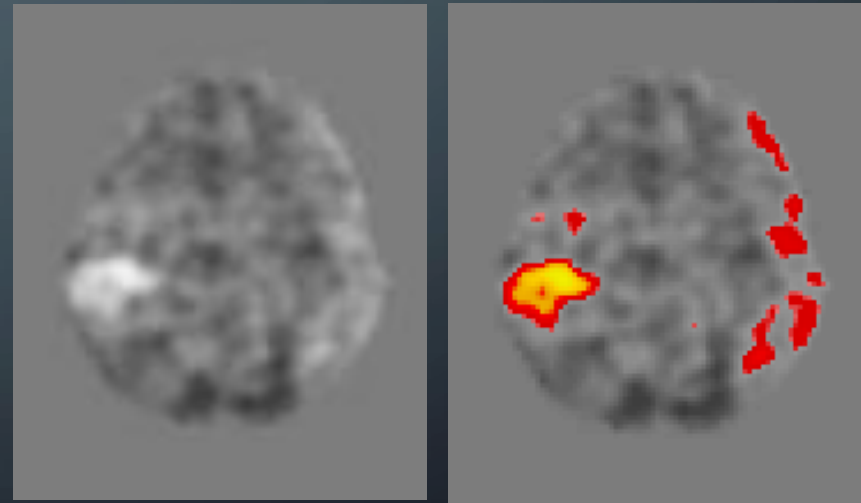
VOXEL SURFING/PLOTTING

- Used to examine how well the changes in the MRI signal follow the on-off characteristics of the task



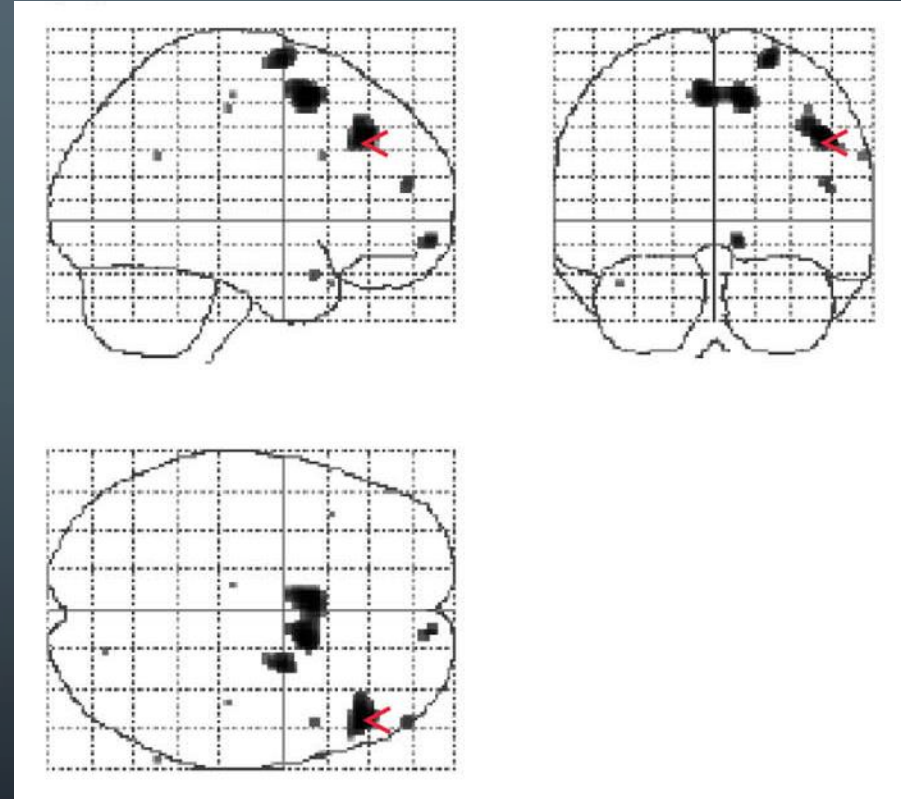
BASIC DISPLAY OF RESULTS

- Simply display all of the t-statistics or other statistic in gray scale or with color coding
- Useful for getting an overall sense of the results
- Can see the data in its most basic form
- Use threshold of 0.1 or higher in SPM to get similar type of display



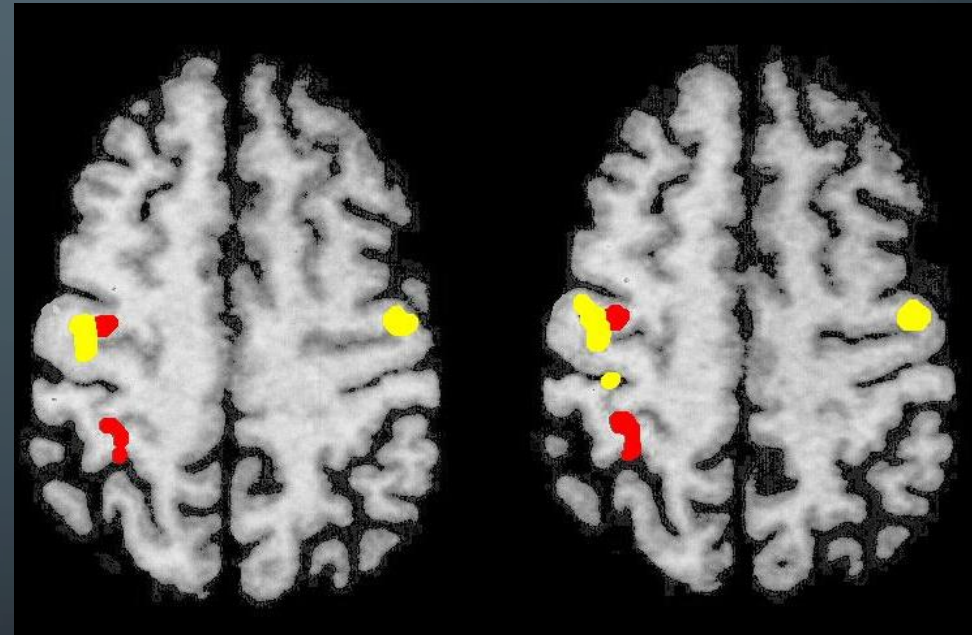
GLASS BRAIN (SPM)

- Glass brain is a maximum intensity projection (MIP) generated for all three orthogonal planes
- Quick way to see what is activated when there are only a few areas



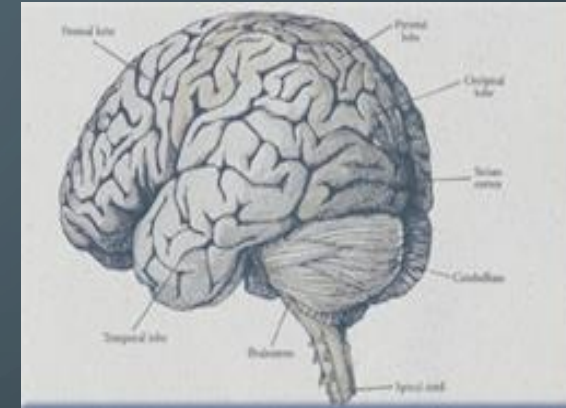
FUSION OF FUNCTIONAL RESULTS AND ANATOMICAL IMAGE

- Use co-register to overlay onto a subject's own anatomical scan
- SPM has option to overlay onto standard brain in MNI space

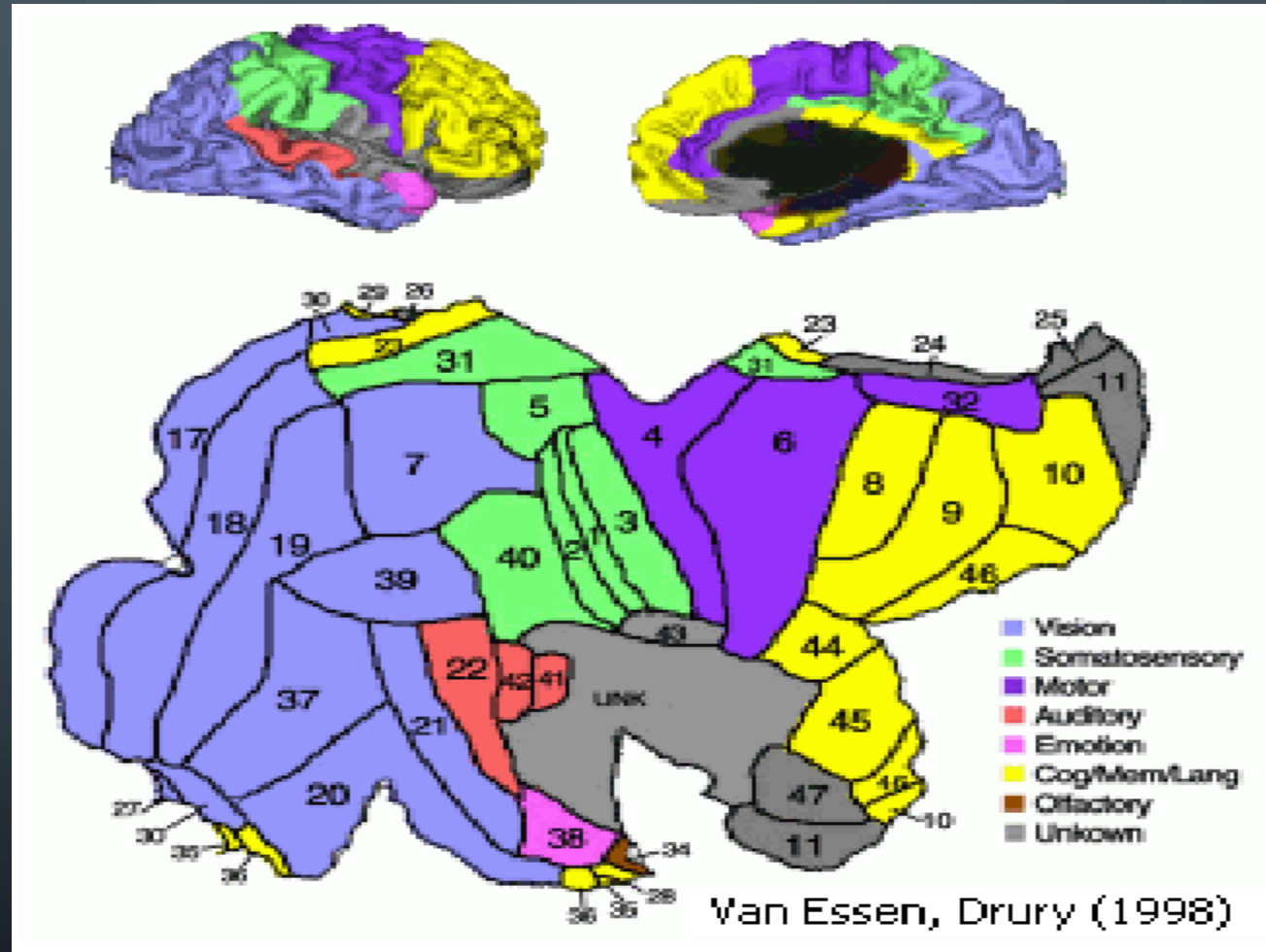


WHAT DO WE DO WITH THIS “ACTIVATION INFORMATION”?

- Whole brain analysis
 - Explore areas throughout that are active
 - Compare subjects or groups
- Region of interest (ROI) analysis
 - Explore specific areas of interest in the brain
 - Percent activation in the area –
(number of active voxels/total number of voxels)
 - Explore “level of activation” as a function of another variable
(example, IQ)

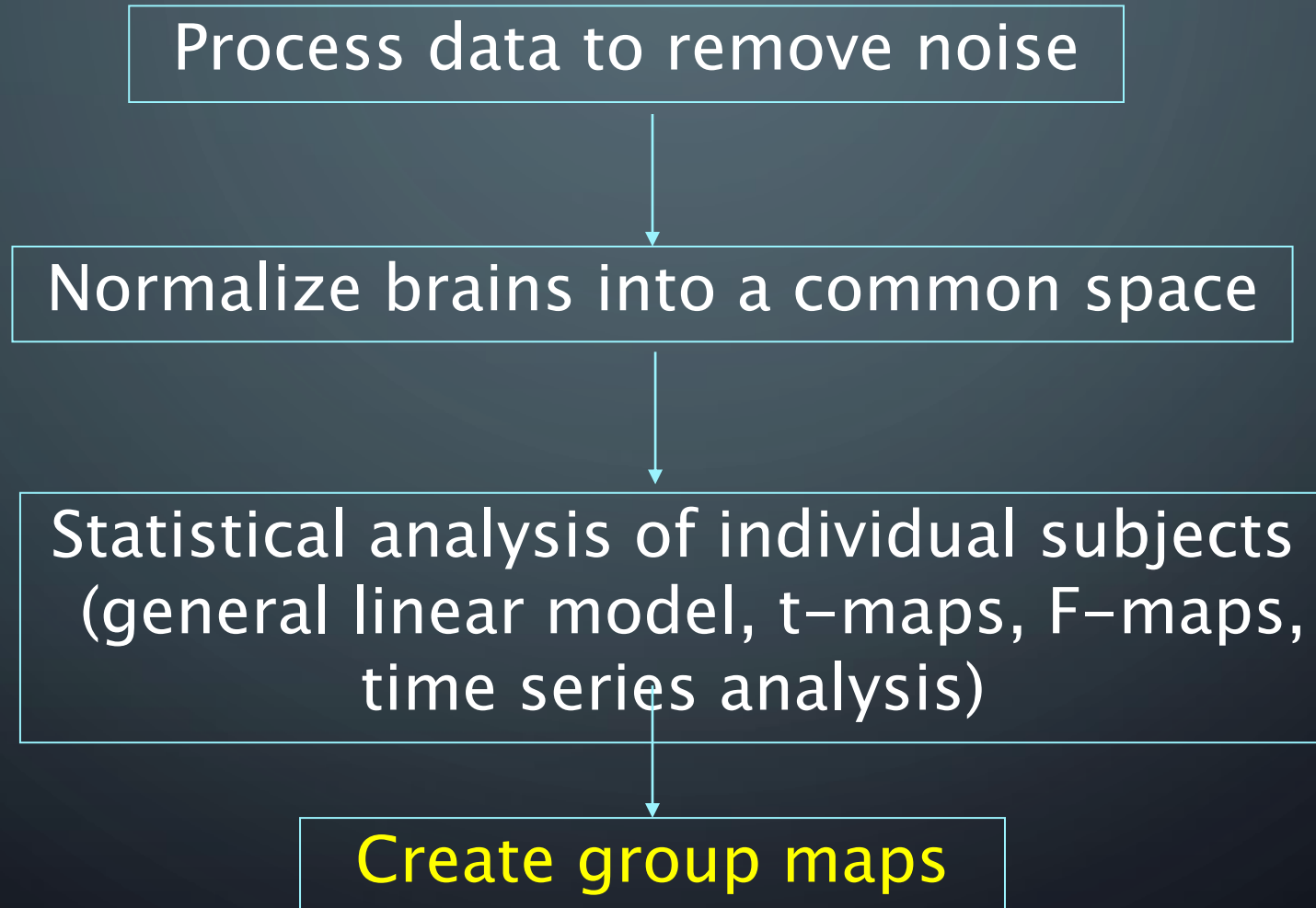


Region of Interest Analysis



Map of Brodmann's Areas

EXAMPLE OF FMRI DATA ANALYSIS STEPS

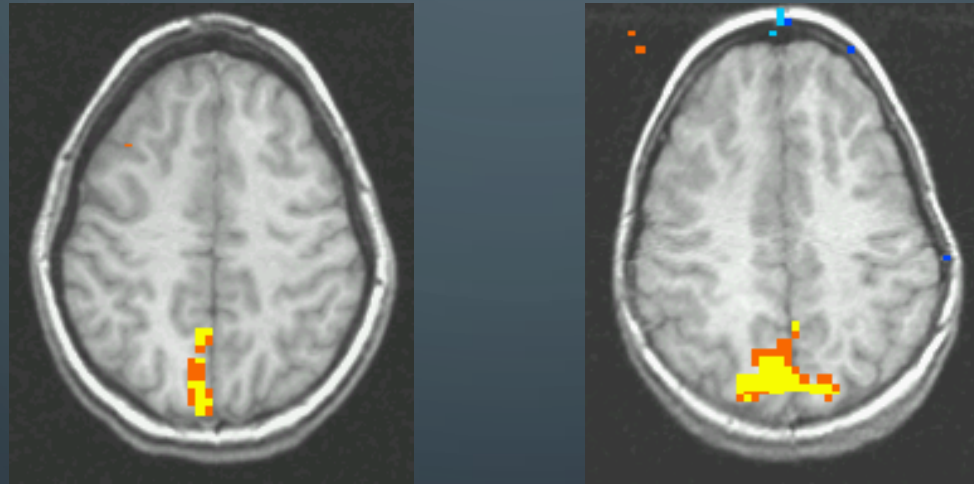


PROBLEMS OF GROUP MAPS

1. Limited number of subjects
2. All brains must be “warped” into a normalized space
3. Need a method for “pooling” statistical data
4. Levels of activation vary across subjects

FMRI GROUP ANALYSIS

- Can look at activation for a single subject
- Can compare between 2 subjects



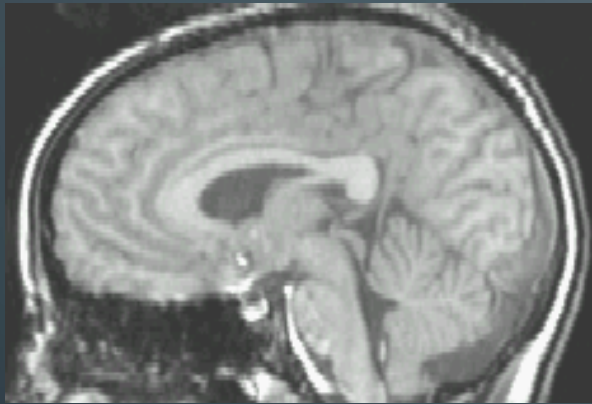
- What if we want to compare between 2 groups of subjects?
- Need more subjects for statistical power

BRAIN NORMALIZATION

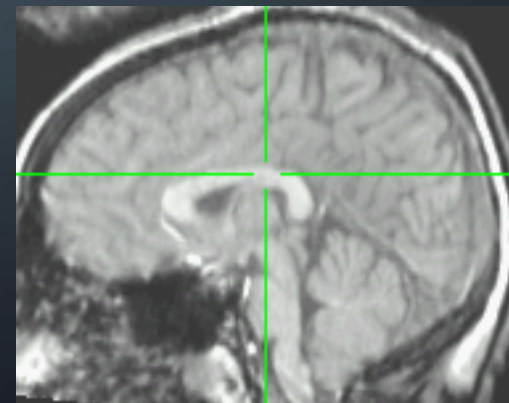
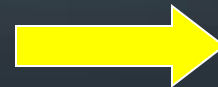
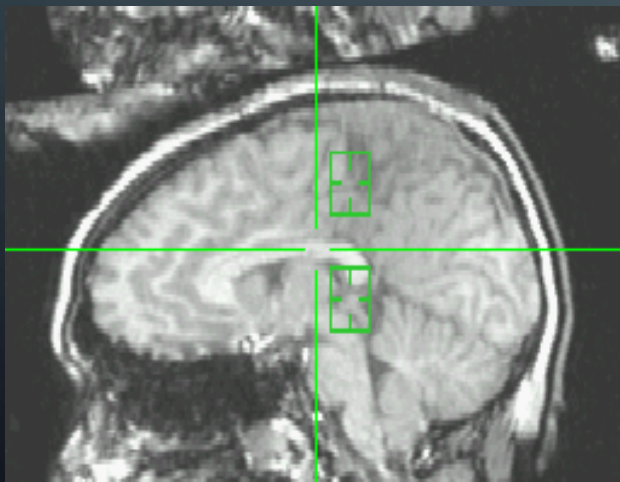
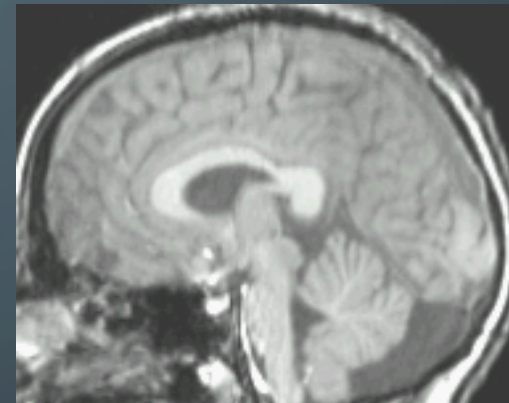
- Warp all brains into a “normalized” space by lining up brain landmarks
- Talairach and Tournoux Atlas, Montreal Neurological Institute (MNI) template
- Although this is a standard procedure in fMRI group analysis, it introduces *spatial smoothing and other errors*

“TALAIRACHING” OF BRAINS

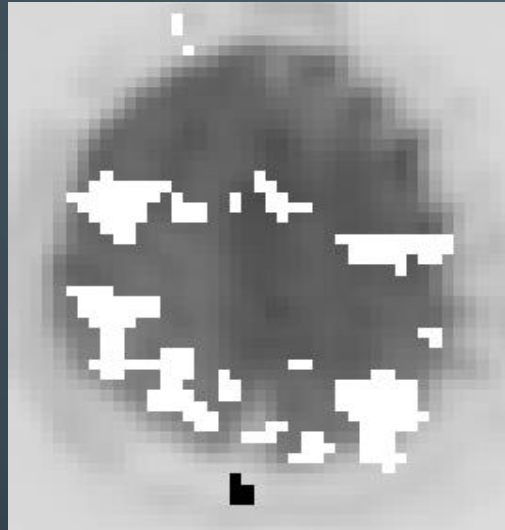
BEFORE



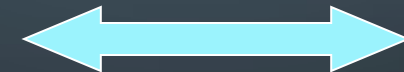
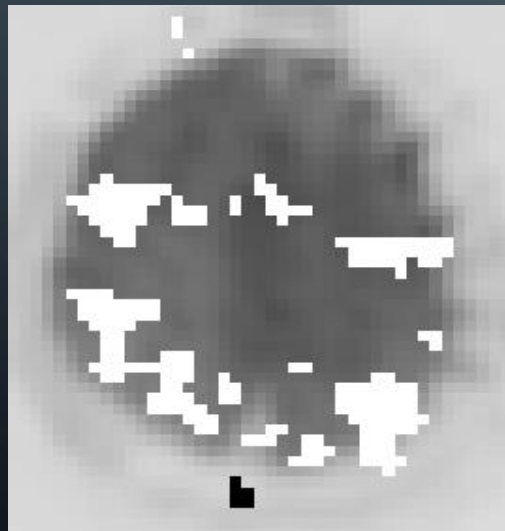
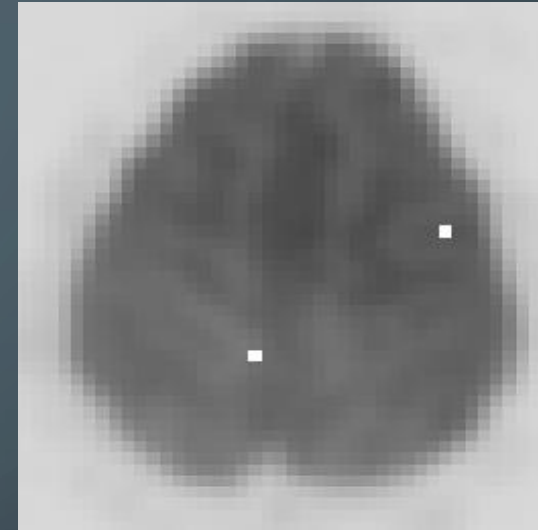
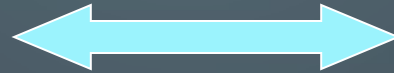
AFTER



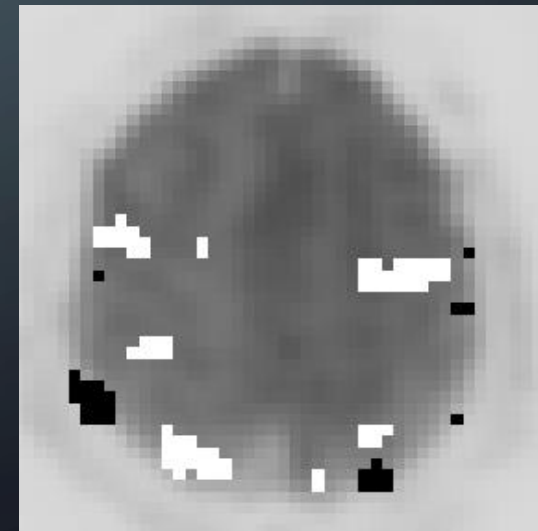
EXAMPLE OF VARIATIONS IN ACTIVATION



Same Task
Different Subjects



Same Subject
Same Task
Different Years



HOW DO WE COMBINE STATISTICAL INFORMATION?

- Individual subject results are already based on statistical processing
- Combining Tests (T values, P values)
 - *Fisher's Method (1950)*
 - *Tippett's Method (1931)*
 - *The Stouffer Method (1949)*
 - *Averaging T-maps*
 - *Worsley and Friston (2000)*
- Combined estimation or “meta-analysis”
 - *Fixed effects and random effects models*

COMBINING METHODS

For k independent tests of a particular null hypothesis, with corresponding P and T values...

Fisher's Method

$$T_F = -2 \sum_{i=1}^k \log P_i$$

Stouffer Method

$$T_s = \sum_{i=1}^k \frac{\Phi^{-1}(1 - P_i)}{\sqrt{k}}$$

Averaging T-maps

$$T_A = \sum_{i=1}^k \frac{T_i}{\sqrt{k}}$$

The Random Effects Model

$$y_i = (\theta + e_i) + \varepsilon_i$$

$$\hat{\theta}^* = \frac{\sum_{i=1}^k w_i^* y_i}{\sum_{i=1}^k w_i^*}$$

$$w_i^* = \frac{1}{(s_i^2 + \hat{\sigma}_\theta^2)}$$

WHICH METHODS ARE “BEST”?

- For robust results without excessive information smoothing –
 - *Fixed or Random effects*
 - *The Stouffer Method*
 - *Fisher’s Method*
- How sensitive are these to the effects of individual subjects?

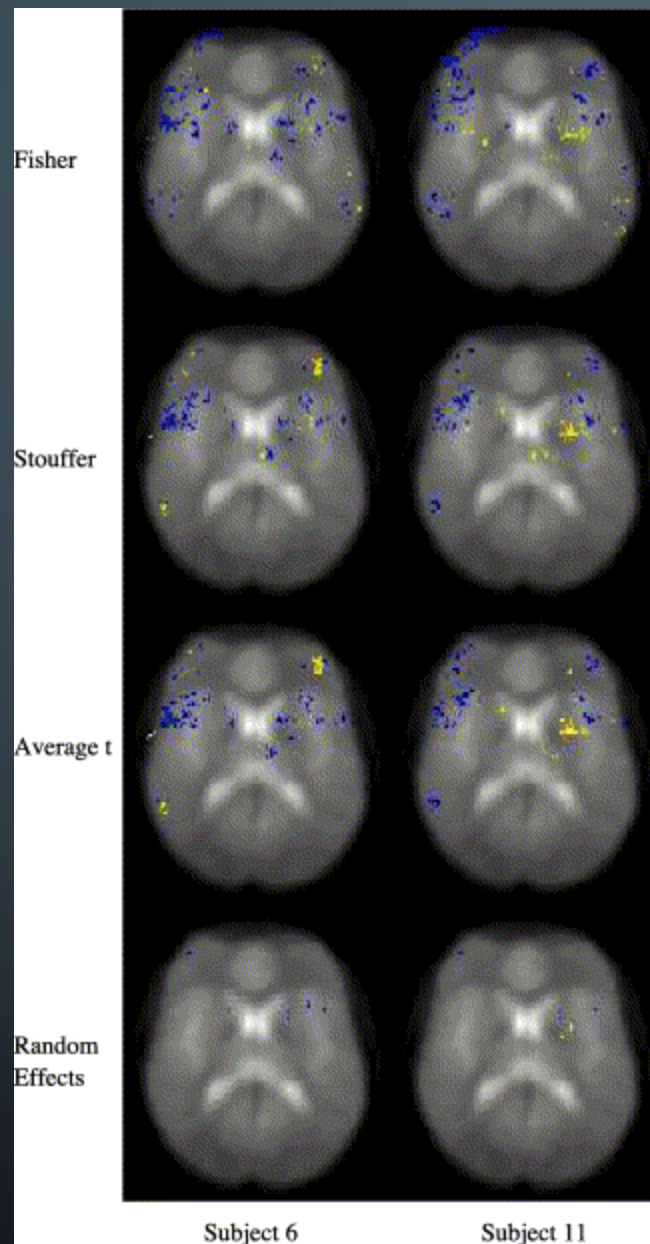
SENSITIVITY ASSESSMENT OF GROUP MAPS

- 4 combining methods, 11 subjects
- Used “delete one subject diagnostics” (jack-knifing) to create “leave one out” maps
- Recomputed group maps
- Examined changes in the resulting maps for each method as each subject was left out

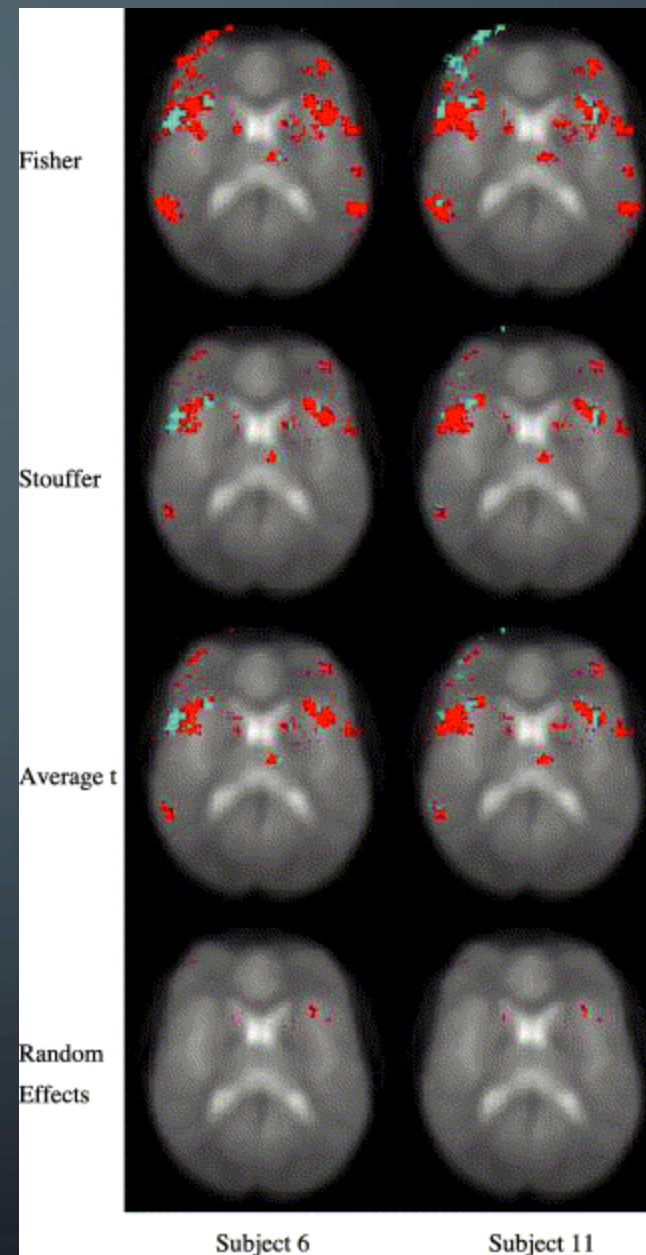
DISPARITY MEASURES FOR “LEAVE ONE OUT” MAPS

Using a standard threshold, compared each “leave one out” map to the complete group map (CGM), examining:

- Enhancing voxels – *# voxels added to CGM*
- Diminishing voxels - *# voxels removed from CGM*
- Relative effect – $(\text{\# enhancing voxels} + \text{\# diminishing voxels}) / \text{\# voxels in CGM}$
- Percent overlap – *# voxels in the subject's individual t-map also present in CGM*



Difference Maps



Overlap Maps

DISPARITY TABLES

Study	Enhancing voxels	Diminishing voxels	Relative effect	Percent overlap
1	319	9	0.47	0.2741
2	170	62	0.34	0.0539
3	86	190	0.40	0.0233
4	185	47	0.34	0.1108
5	155	29	0.27	0
6	266	55	0.47	0.2230
7	64	175	0.35	0.0029
8	249	51	0.44	0.2784
9	132	93	0.33	0.0204
10	199	76	0.40	0.1005
11	205	89	0.43	0.1589
Mean	184.55	79.64	0.39	0.11
SD	75.88	56.49	0.06	0.11

Results from the Stouffer Method

Fisher

Study	Enhancing voxels	Diminishing voxels	Relative effect	Percent overlap
1	432	50	0.30	0.1857
2	176	113	0.18	0.0414
3	83	169	0.16	0.0113
4	307	56	0.23	0.0909
5	140	106	0.15	0.0038
6	363	42	0.25	0.1242
7	82	134	0.13	0.0050
8	673	14	0.43	0.3683
9	156	146	0.19	0.0176
10	254	106	0.23	0.0928
11	399	75	0.30	0.2114
Mean	278.64	91.91	0.23	0.10
SD	179.94	48.35	0.09	0.11

Random Effects

Study	Enhancing voxels	Diminishing voxels	Relative effect	Percent overlap
1	10	8	0.60	0.0333
2	13	2	0.50	0
3	6	42	1.60	0
4	20	2	0.73	0.1000
5	18	2	0.67	0
6	18	1	0.63	0
7	7	17	0.80	0
8	14	6	0.67	0.0333
9	9	3	0.40	0
10	11	5	0.53	0.0333
11	8	16	0.80	0.0333
Mean	12.18	9.45	0.72	0.02
SD	4.81	12.14	0.32	0.03

RESULTS AND CONCLUSIONS

Most
sensitive



Least
sensitive

- Fisher's Method – *exploratory studies*
- The Stouffer method/ Averaging of T-maps – *general use*
- Random effects – *large subject groups*