

Martin Lindquist

Department of Biostatistics
Johns Hopkins
Bloomberg School of Public Health

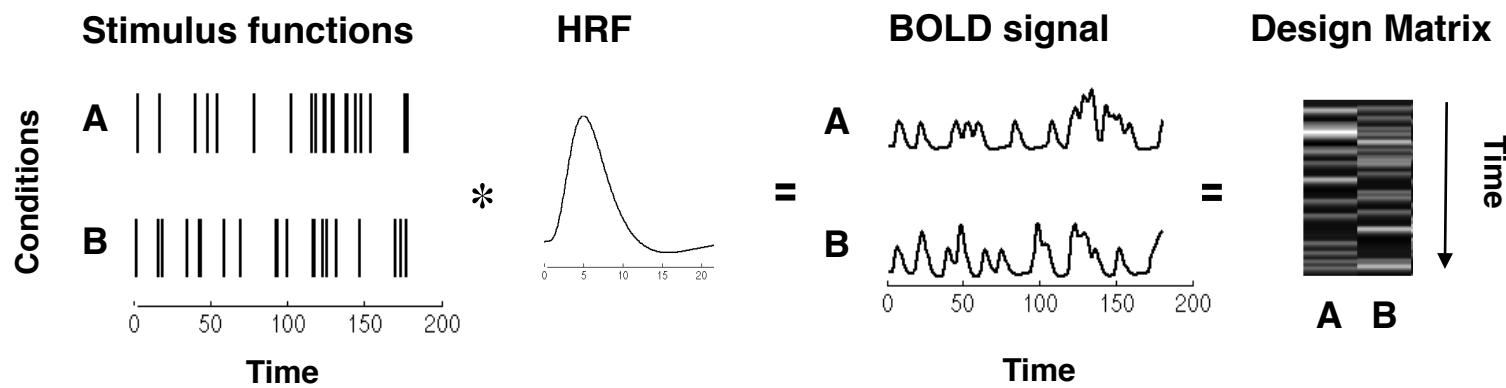
Tor Wager

Department of Psychology and
Neuroscience and the
Institute for Cognitive Science
University of Colorado, Boulder

Multiple Comparisons

Localizing Activation

1. Construct a model for each voxel of the brain.
 - “Massive univariate approach”
 - Regression models (GLM) commonly used.

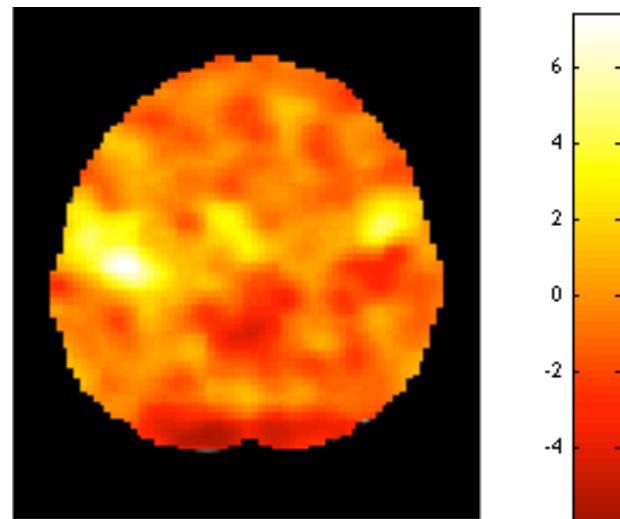


$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{V})$$

Localizing Activation

2. Perform a statistical test to determine whether task related activation is present in the voxel.

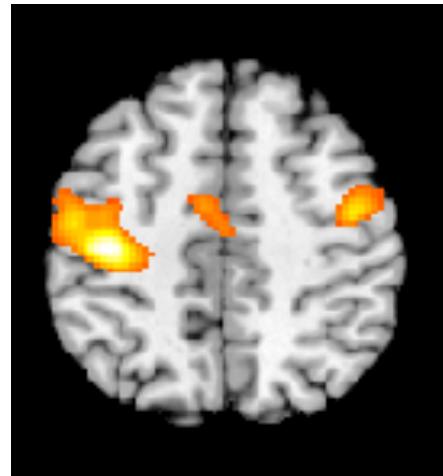
$$H_0 : \mathbf{c}^T \boldsymbol{\beta} = 0$$



Statistical image:
Map of t-tests
across all voxels
(a.k.a t-map).

Localizing Activation

3. Choose an appropriate threshold for determining statistical significance.



Statistical parametric map:
Each significant voxel is
color-coded according to
the size of its p-value.

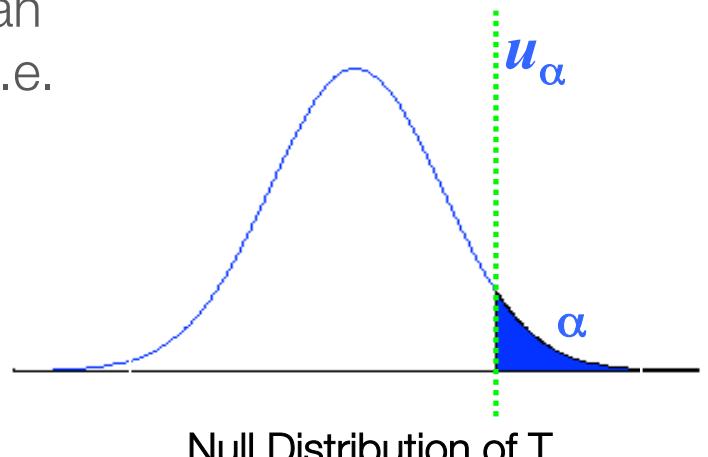
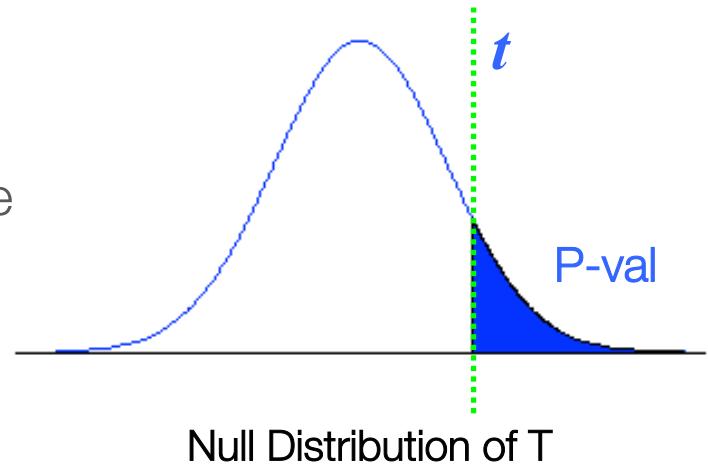
Hypothesis Testing

- Null Hypothesis H_0
 - Statement of no effect (e.g., $\beta_1 - \beta_2 = 0$).

- Test statistic T
 - Measures compatibility between the null hypothesis and the data.

- P-value
 - Probability that the test statistic would take a value as or more extreme than that actually observed if H_0 is true, i.e. $P(T > t | H_0)$.

- Significance level
 - Threshold u_α controls false positive rate at level $\alpha = P(T > u_\alpha | H_0)$



Making Errors

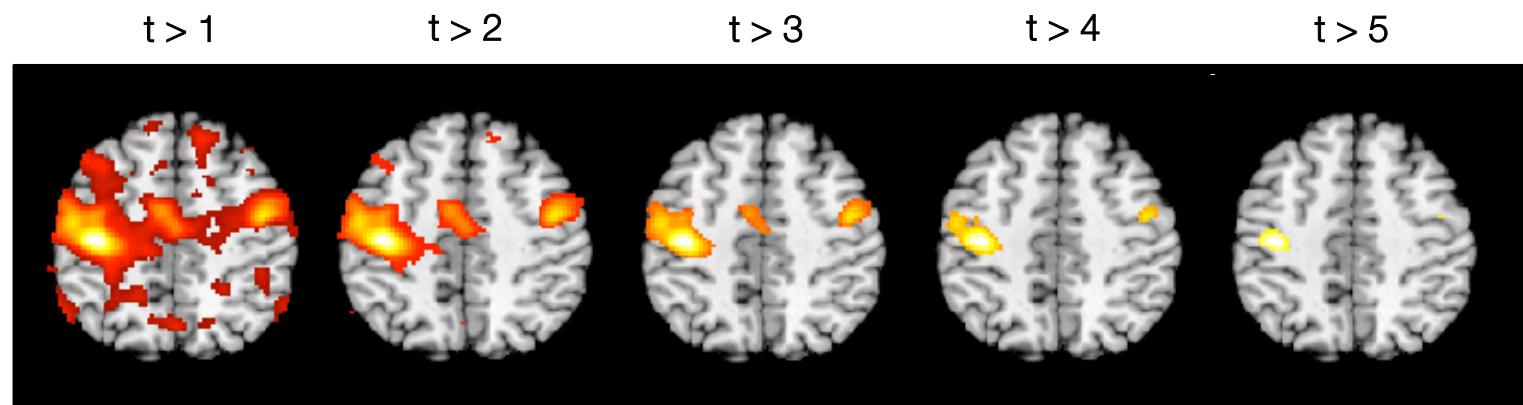
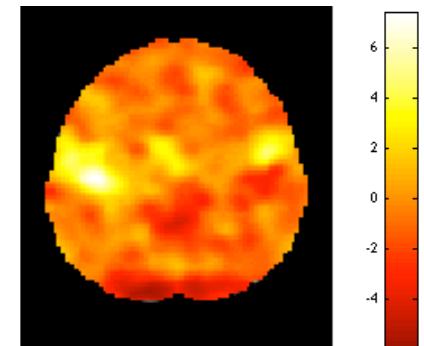
- There are two types of errors one can make when performing significance tests:
 - Type I error
 - H_0 is true, but we mistakenly reject it (False positive).
 - Controlled by significance level α .
 - Type II error
 - H_0 is false, but we fail to reject it (False negative)
- The probability that a hypothesis test will correctly reject a false null hypothesis is the **power** of the test.

Multiple Comparisons

- Choosing an appropriate threshold is complicated by the fact we are dealing with a **family of tests**.
- If more than one hypothesis test is performed, the risk of making at least one Type I error is greater than the α value for a single test.
- The more tests one performs, the greater the likelihood of getting at least one false positive.

Multiple Comparisons

- Which of 100,000 voxels are significant?
 - $\alpha=0.05 \Rightarrow 5,000$ false positive voxels
- Choosing a threshold is a balance between sensitivity ([true positive rate](#)) and specificity ([true negative rate](#)).



Measures of False Positives

- There exist several ways of quantifying the likelihood of obtaining false positives.
- Family-Wise Error Rate (FWER)
 - Probability of any false positives
- False Discovery Rate (FDR)
 - Proportion of false positives among rejected tests



End of Module



@fMRIstats