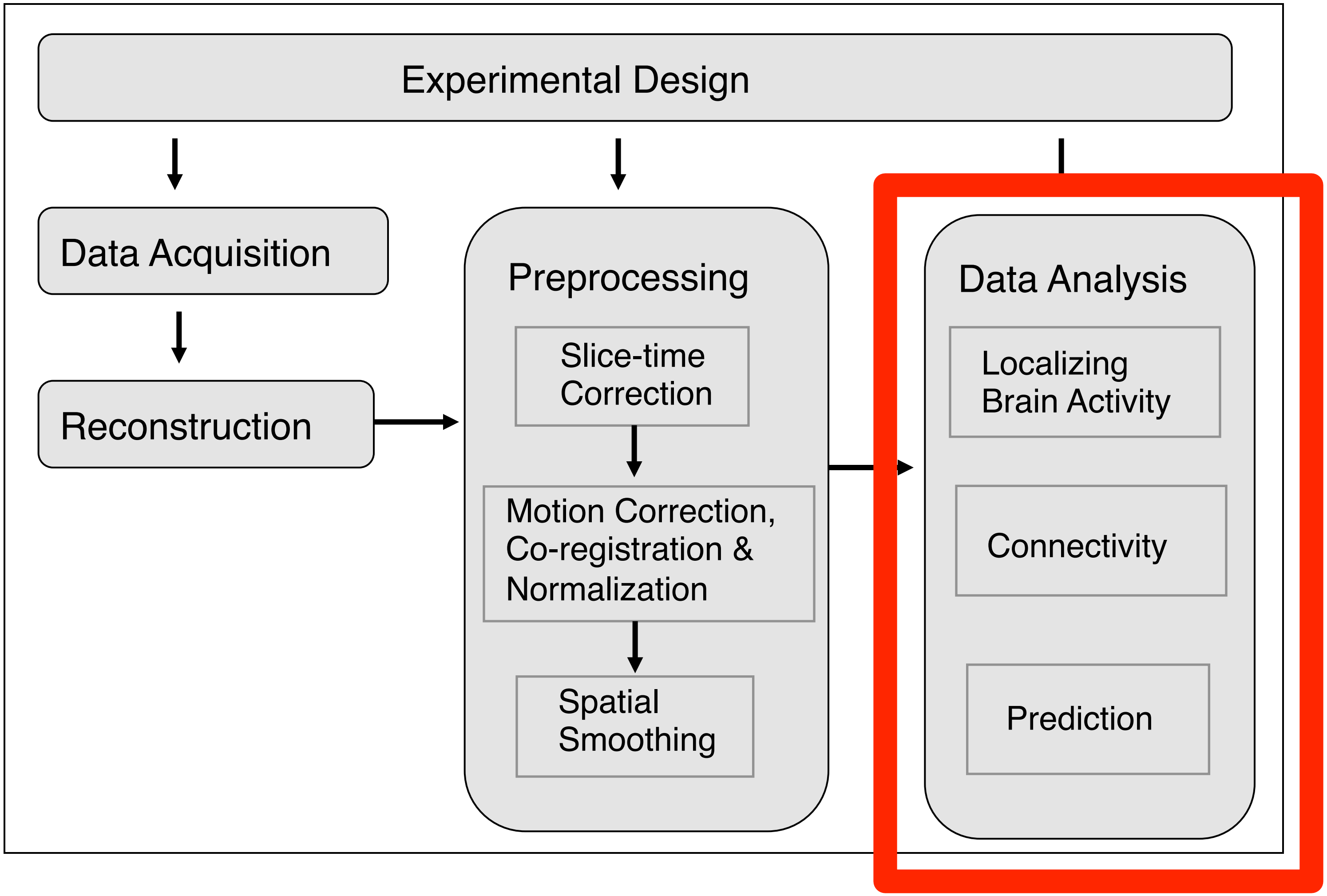
Module 10:
The General Linear Model

Data Processing Pipeline



Statistical Analysis

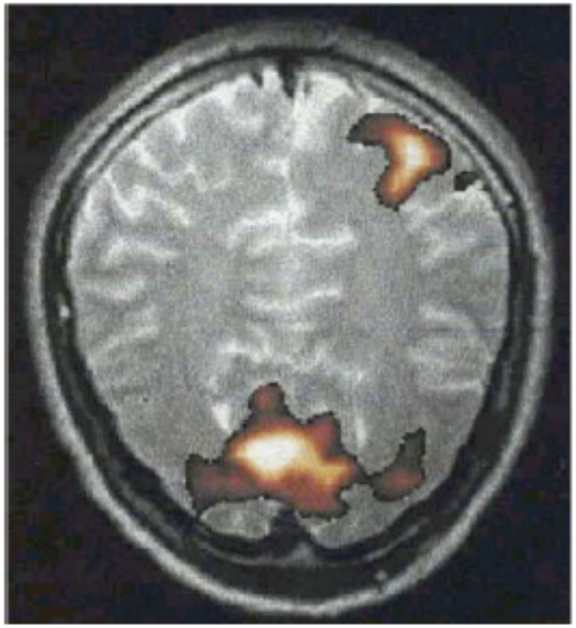
- There are multiple goals in the statistical analysis of fMRI data.

- They include:
  - localizing brain areas activated by the task;
  - determining networks corresponding to brain function; and
  - making predictions about psychological or disease states.

Human Brain Mapping

- The most common use of fMRI to date has been to localize areas of the brain that activate in response to a certain task.

- These types of human brain mapping studies are necessary for the development of biomarkers and increasing our understanding of brain function.



Massive Univariate Approach

- Typically analysis is performed by constructing a separate model at each voxel
  - The ‘massive univariate approach’.
  - Assumes an improbable independence between voxel pairs......

- Typically dependencies between voxels are dealt with later using random field theory, which makes assumptions about the spatial dependencies between voxels.

General Linear Model

- The general linear model (GLM) approach treats the data as a linear combination of model functions (predictors) plus noise (error).

- The model functions are assumed to have known shapes, but their amplitudes are unknown and need to be estimated.

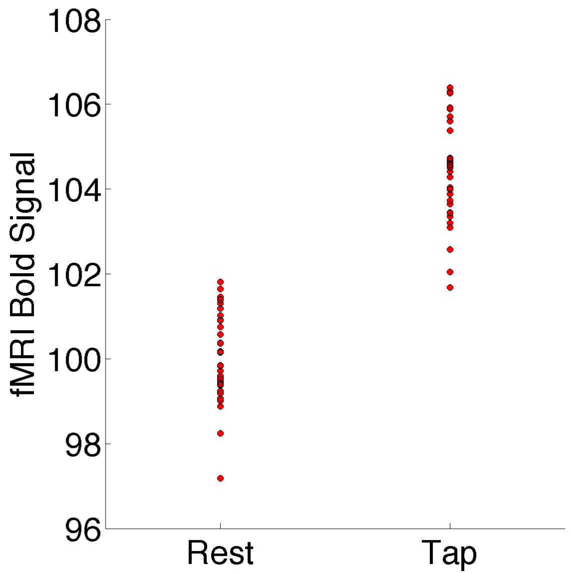
- The GLM framework encompasses many of the commonly used techniques in fMRI data analysis (and data analysis more generally).

Illustration

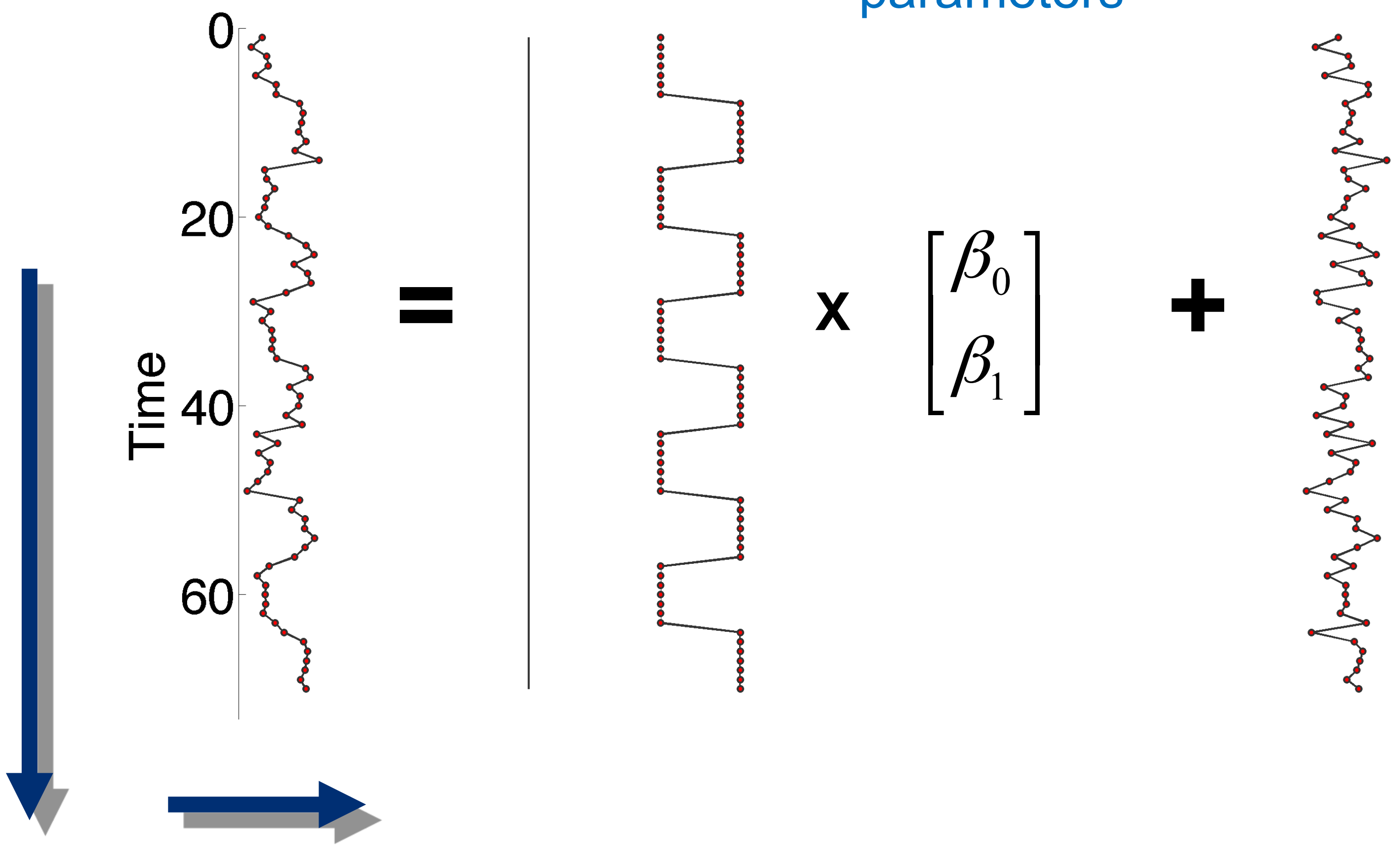
- Consider an experiment of alternating blocks of finger-tapping and rest.

- Construct a model to study data from a single voxel for a single subject.

- We seek to determine whether activation is higher during finger-tapping compared with rest.



fMRI Data | Design matrix | Model parameters | Residuals



BOLD signal | Intercept | Predicted task response

$$H_0 : \beta_1 = 0$$

GLM

A standard GLM can be written:

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

where

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1p} \\ 1 & X_{21} & \cdots & X_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & X_{np} & \cdots & X_{np} \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

fMRI Data

Design matrix

Regression coefficients

Noise

V is the covariance matrix whose format depends on the noise model.

The quality of the model depends on our choice of X and V.

Estimation

- If ε is i.i.d., then Ordinary Least Square (OLS) estimate is optimal

model **estimate**

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \longrightarrow \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

- If Var(ε) $= \mathbf{V}\sigma^2 \neq \mathbf{I}\sigma^2$, then Generalized Least Squares (GLS) estimate is optimal

model **estimate**

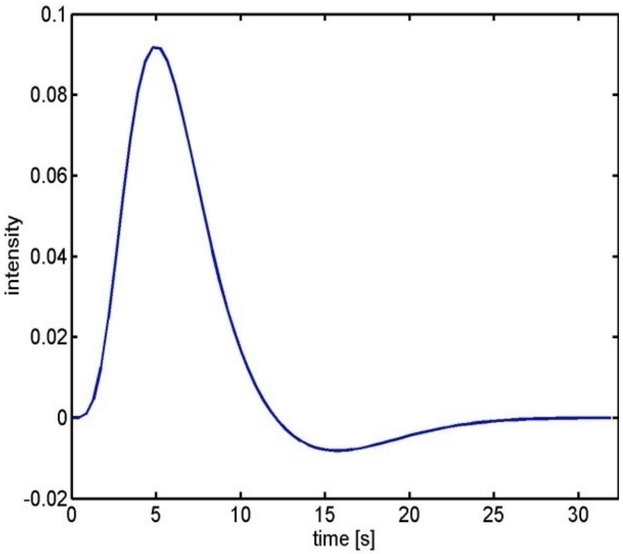
$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \longrightarrow \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$$

Model Refinement

- This model has a number of shortcomings.
- We want to use our understanding of the signal and noise properties of BOLD fMRI to aid us in constructing appropriate models.
- This includes deciding on an appropriate design matrix, as well as an appropriate noise model.

Issues

1. BOLD responses have a delayed and dispersed form.



2. The fMRI signal includes substantial amounts of low-frequency noise.

3. The data are serially correlated which needs to be considered in the model.

End of Module