Module 15:
Inference

GLM Summary

model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \longrightarrow$$

estimate

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

fitted values

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta}$$

residuals

$$\begin{aligned}\mathbf{r} &= \mathbf{Y} - \hat{\mathbf{Y}} \\ &= (\mathbf{I} - (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})\mathbf{Y} \\ &= \mathbf{R}\mathbf{Y}\end{aligned}$$

Inference

- After fitting the GLM use the estimated parameters to determine whether there is significant activation present in the voxel.

- Inference is based on the fact that:

$$\hat{\beta} \sim N(\beta, (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1})$$

- Use t and F procedures to perform tests on effects of interest.

Contrasts

- It is often of interest to see whether a linear combination of the parameters are significant.

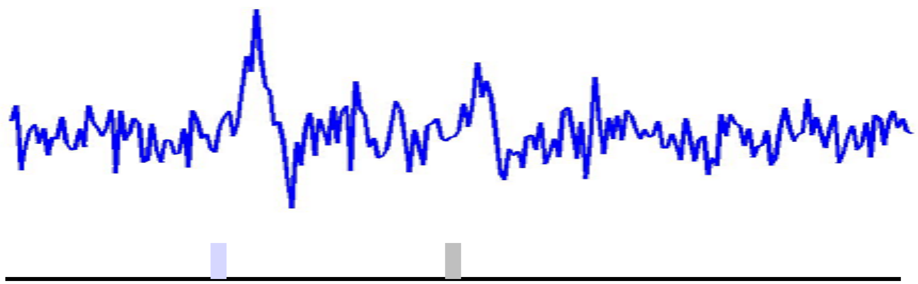
- The term $\mathbf{c}^T\boldsymbol{\beta}$ specifies a linear combination of the estimated parameters, i.e.

$$\mathbf{c}^T\boldsymbol{\beta} = c_1\beta_1 + c_2\beta_2 + \ldots + c_n\beta_n$$

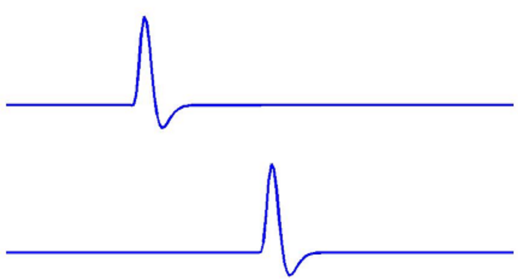
- Here **c** is called a contrast vector.

Example

Event-related experiment with two types of stimuli.



=



β_1

β_2

β_3

+ Noise

$$H_0 : \beta_2 = \beta_3$$

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

$$\mathbf{c}^T = (0,1,-1)$$

T-test

- To test

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

use the t-statistic:

$$T = \frac{\mathbf{c}^T \hat{\boldsymbol{\beta}}}{\sqrt{Var\left(\mathbf{c}^T \hat{\boldsymbol{\beta}}\right)}}$$

- Under H_0, T is approximately $t(\nu)$ with $\nu = \dfrac{tr(\mathbf{RV})^2}{tr((\mathbf{RV})^2)}$

Multiple Contrasts

- We often want to make simultaneous tests of several contrasts at once.
- $\mathbf{c}$ is now a contrast matrix.
- Suppose

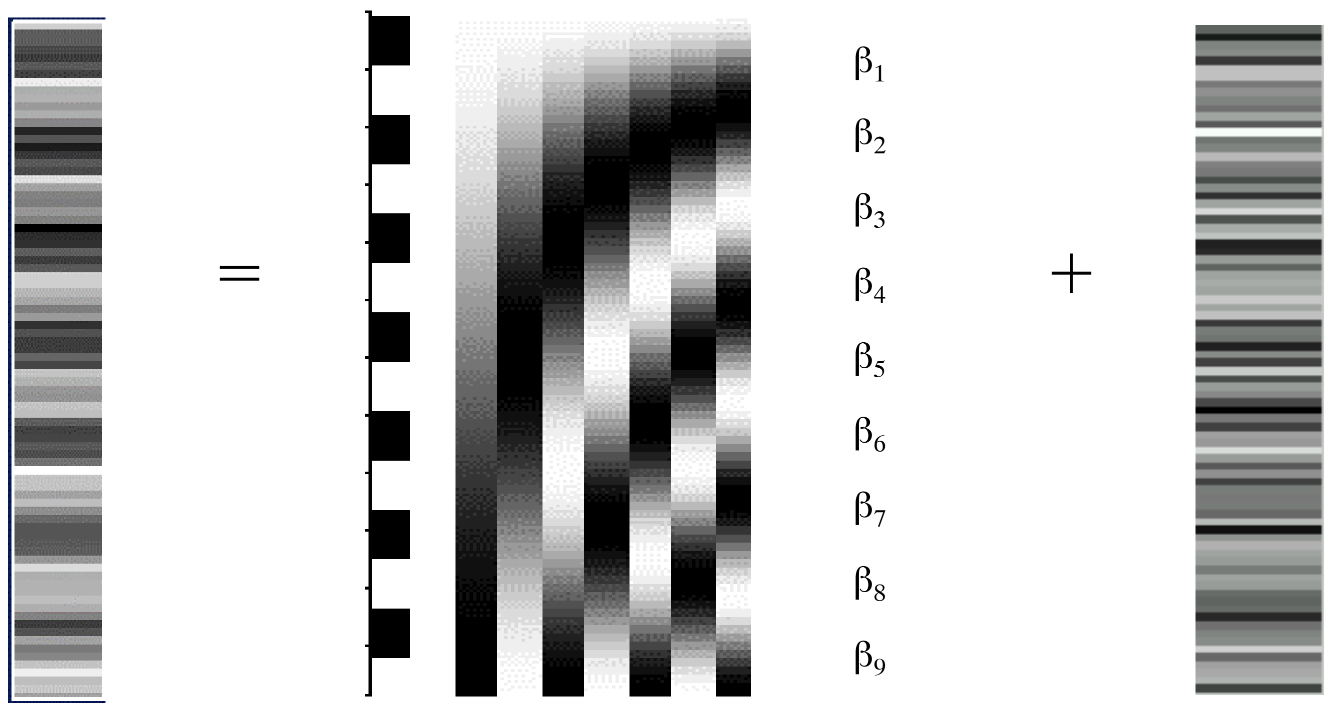
$$\mathbf{c} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

then

$$\mathbf{c}^T \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

Example

Consider a model with box-car shaped activation and drift modeled using the discrete cosine basis.



$$Y = X \times \beta + \varepsilon$$

Example

Do the drift components add anything to the model?

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

where

$$\mathbf{c} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Example

This is equivalent to testing:

$$H_0 : \begin{pmatrix} \beta_3 & \beta_4 & \beta_5 & \beta_6 & \beta_7 & \beta_8 & \beta_9 \end{pmatrix}^T = \mathbf{0}$$

To understand what this implies, we split the design matrix into two parts:

$$\begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{19} \\ 1 & X_{21} & X_{22} & \cdots & X_{29} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{n9} \end{bmatrix}$$

$$\underbrace{\qquad\qquad}_{\mathbf{X}_0} \qquad \underbrace{\qquad\qquad\qquad\qquad}_{\mathbf{X}_1}$$

Example

- Do the drift components add anything to the model?

- The $\mathbf{X_1}$ matrix explains the drift. Does it contribute in a significant way to the model?

- Compare the results using the full model, with design matrix $\mathbf{X}$, with those obtained using a reduced model, with design matrix $\mathbf{X}_0$.

F-test

- Test the hypothesis using the F-statistic:

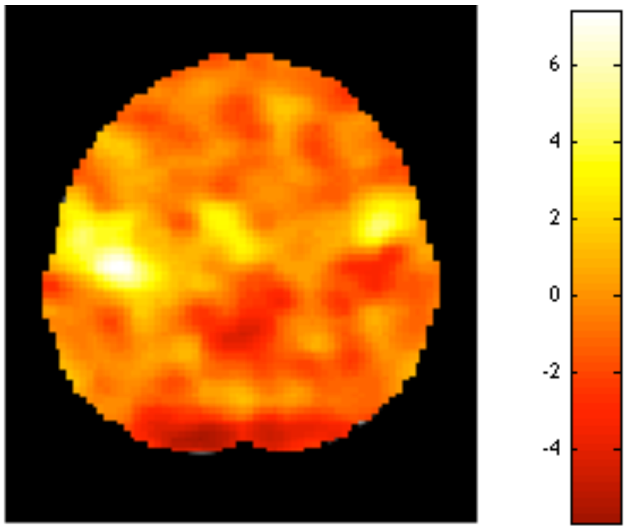
$$F = \frac{\left(\mathbf{r}_0^T \mathbf{r}_0 - \mathbf{r}^T \mathbf{r}\right)}{\hat{\sigma}^2 \left(tr((\mathbf{R} - \mathbf{R}_0)\mathbf{V}\right)}$$

- Assuming the errors are normally distributed, F has an approximate F-distribution with (ν_0, ν) degrees of freedom, where

$$\nu_0 = \frac{tr\left(\left[(\mathbf{R} - \mathbf{R}_0)\mathbf{V}\right]\right)^2}{tr\left(\left[(\mathbf{R} - \mathbf{R}_0)\mathbf{V}\right]^2\right)} \quad \text{and} \quad \nu = \frac{tr(\mathbf{RV})^2}{tr((\mathbf{RV})^2)}$$

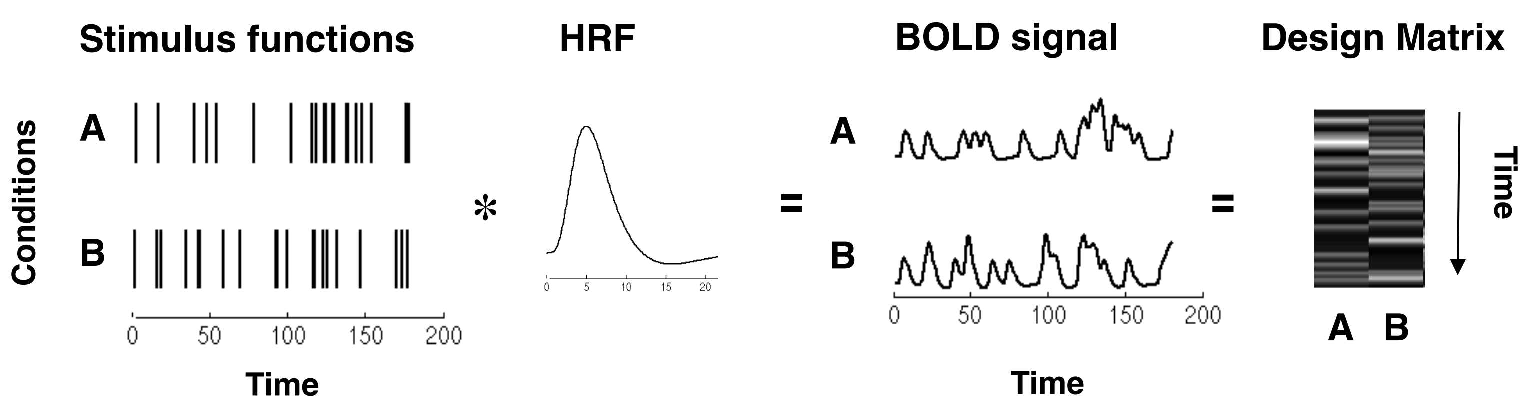
Statistical Images

- For each voxel a hypothesis test is performed. The statistic corresponding to that test is used to create a statistical image over all voxels.



Localizing Activation

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

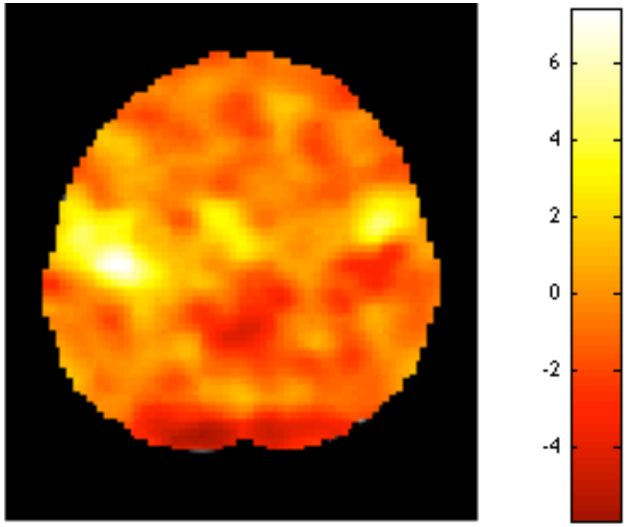


$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{V})$$

Localizing Activation

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

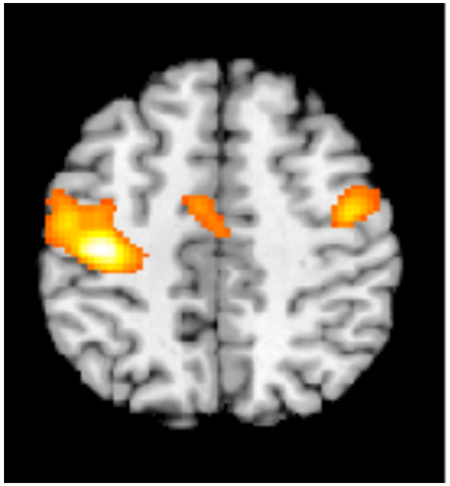
$$H_0 : \mathbf{c}^T \boldsymbol{\beta} = 0$$



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.

Statistical Images

How do we determine which voxels are actually active?

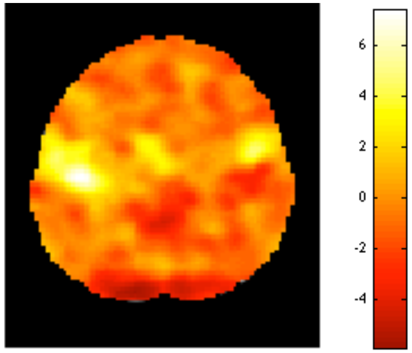
Problems:

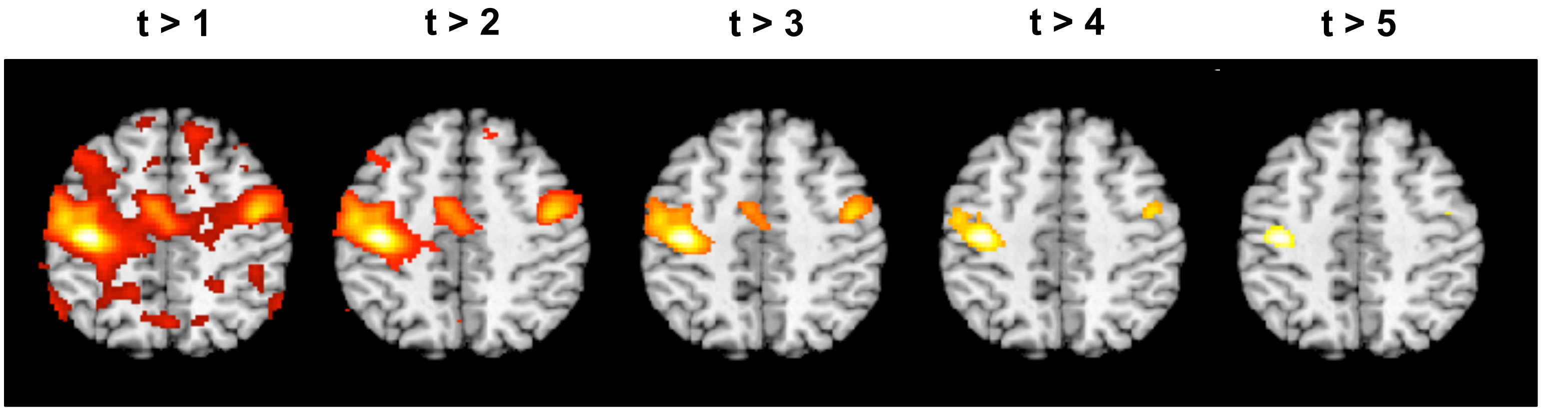
- The statistics are obtained by performing a large number of hypothesis tests.
- Many of the test statistics will be artificially inflated due to the noise.
- This leads to many false positives.

Multiple Comparisons

- Which of 100,000 voxels are significant?
  - α=0.05 $\Rightarrow$ 5,000 false positive voxels

- Choosing a threshold is a balance between sensitivity (true positive rate) and specificity (true negative rate).





End of Module