Statistical analysis of fMRI data

Paul McCarthy

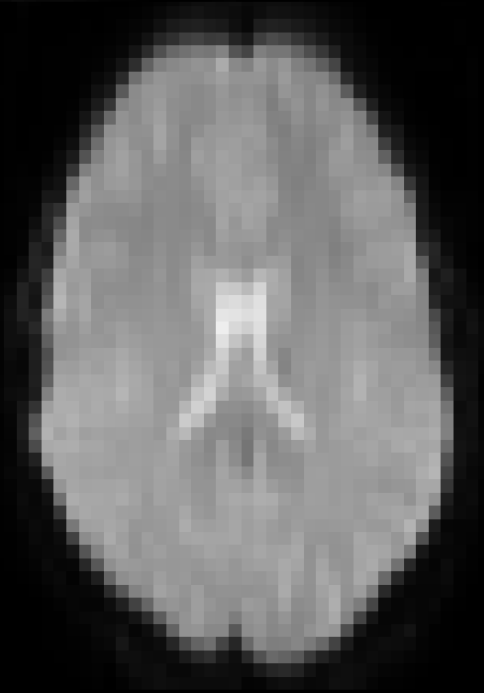
June 8, 2012

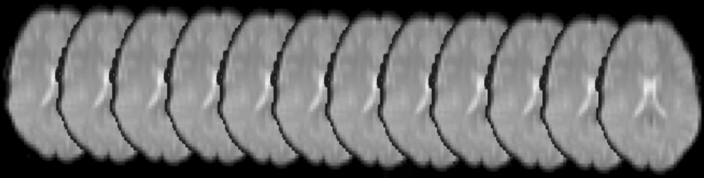
What is fMRI?

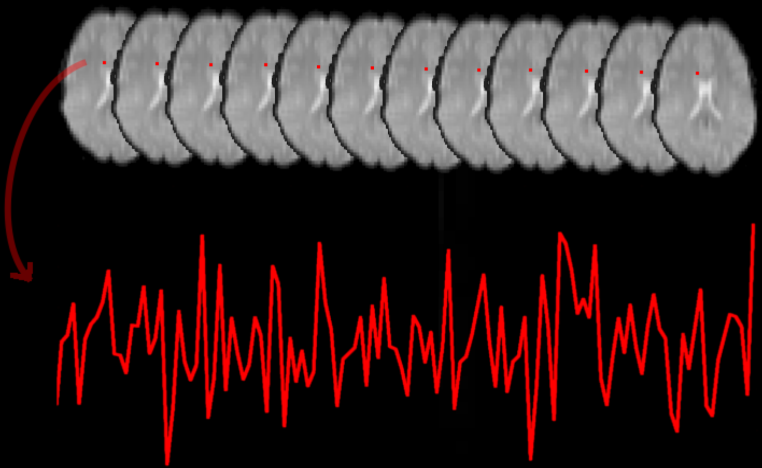
fMRI experiment design

The General Linear Model

Group inference





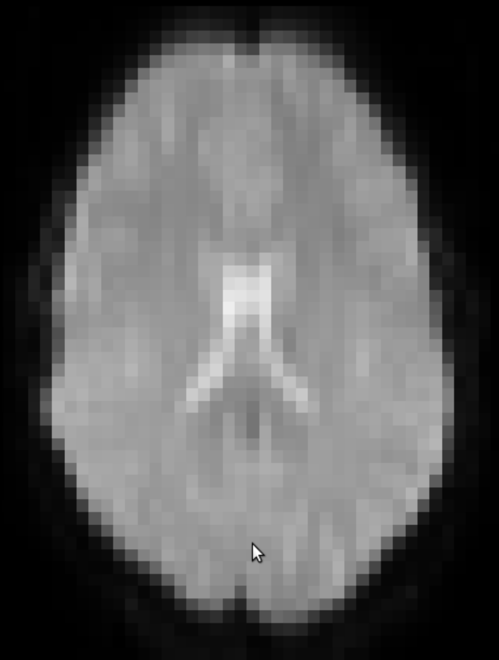


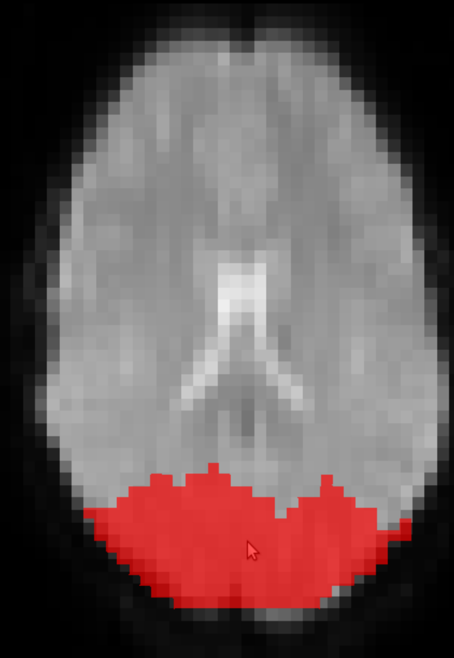
What is fMRI?

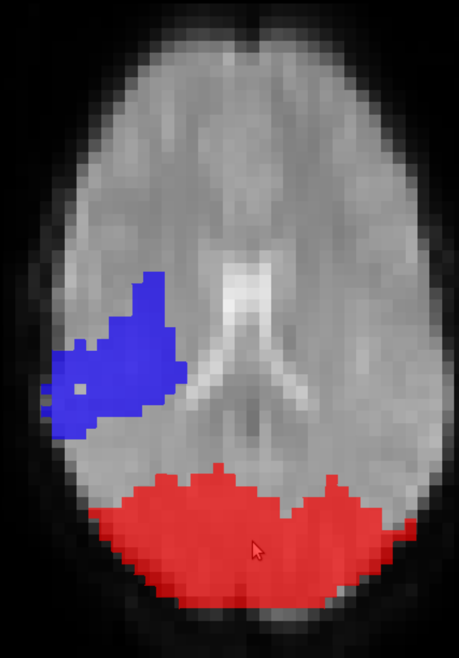
fMRI experiment design

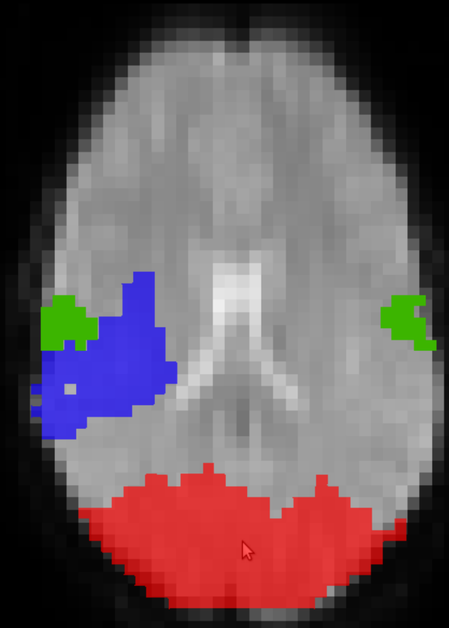
The General Linear Model

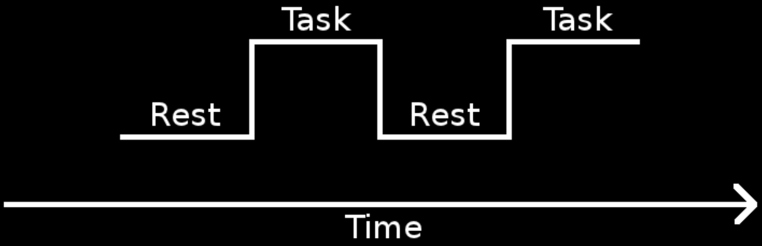
Group inference

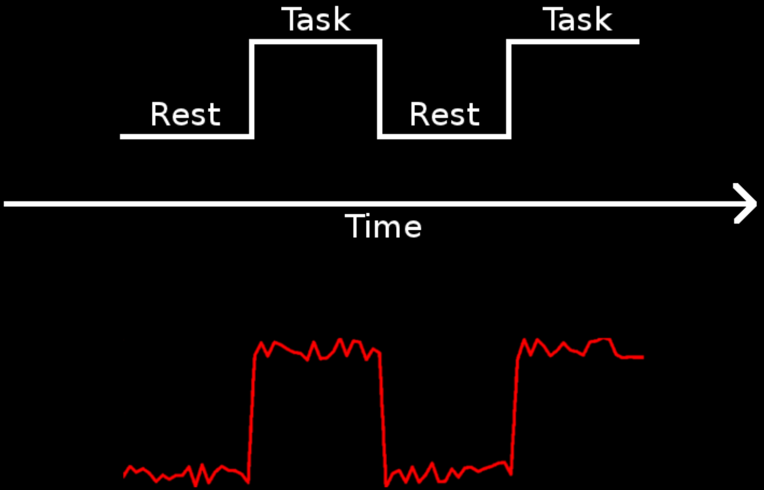


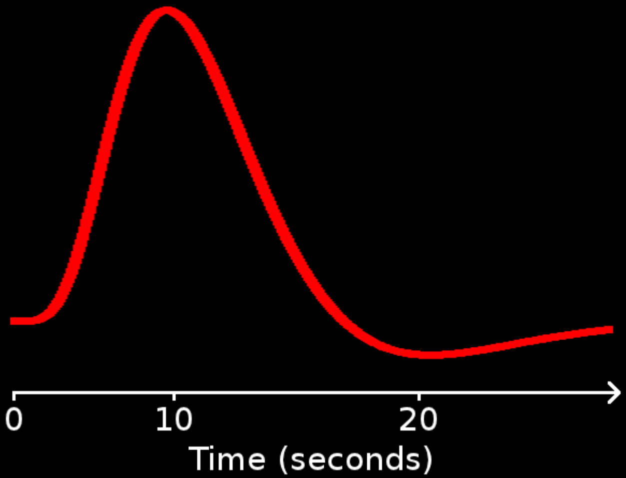


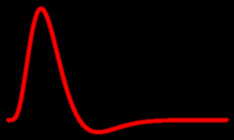


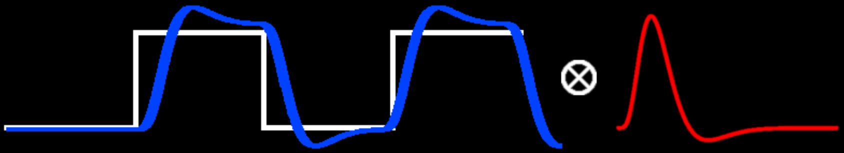



Task
Task
Rest
Rest
Time


Task
Task
Rest
Rest
Time


0
10
20
Time (seconds)





What is fMRI?

fMRI experiment design

The General Linear Model

Group inference

GLM (seriously simplified)

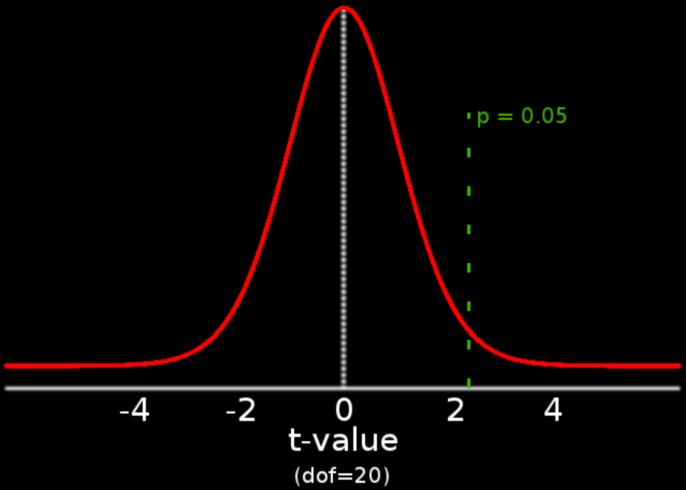
For a voxel	$v = (x, y, z) \in \mathbb{R}^3$
with an observed time course	$Y(v) = \{Y_1(v), Y_2(v), ..., Y_n(v)\}$
and expected time course (based upon the experiment design)	$X = \{X_1, X_2, ..., X_n\}$
Solve for $\beta(v)$:	$Y(v) = X\beta(v) + \epsilon(v)$

- $\beta(v)$ is the parameter estimate, which fits $Y(v)$ to X.
- $\epsilon(v)$ is the error at each time point.

$$t(v) \approx \frac{\beta(v)}{\epsilon(v)}$$

Dividing the parameter estimate $\beta(v)$ by its error $\epsilon(v)$ gives us a t value, which we can use to answer the question:

Was voxel v activated by our experimental task?

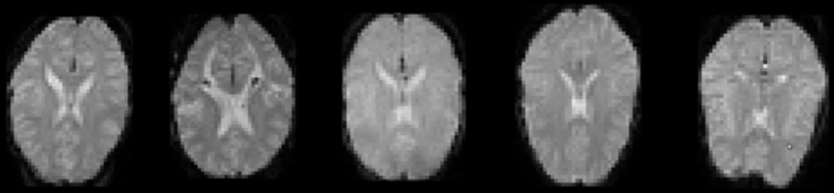

p = 0.05
-4
-2
0
2
4
t-value
(dof=20)

What is fMRI?

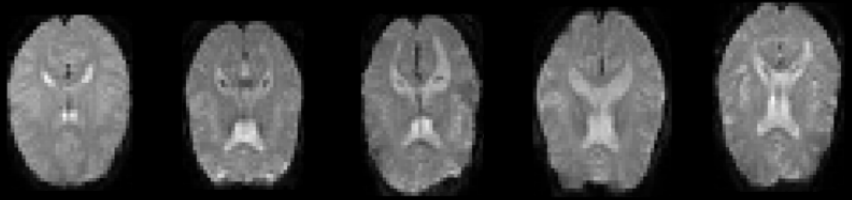
fMRI experiment design

The General Linear Model

Group inference



=

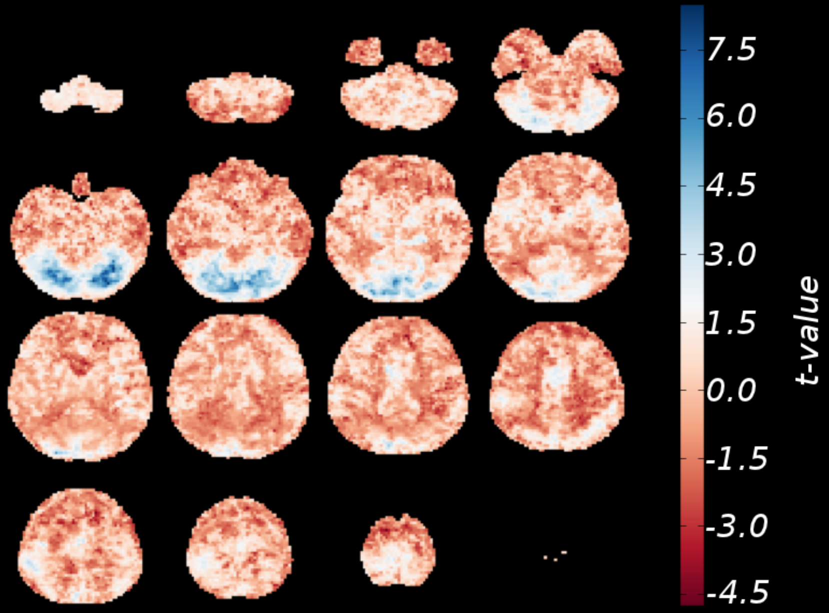


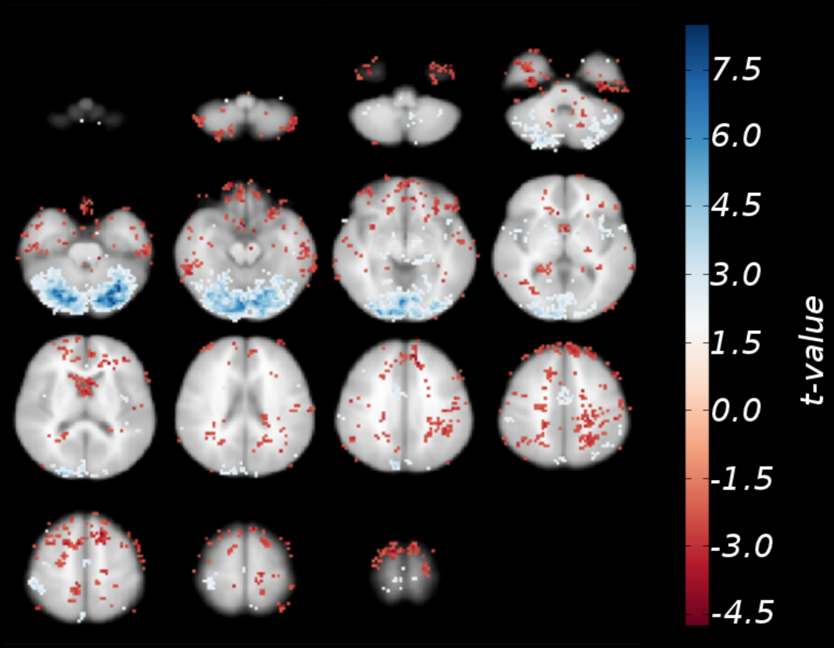
?

Two sample t-test

$$t(x, y) = \frac{\bar{x} - \bar{y}}{S_{xy}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$s_{xy} = \sqrt{\frac{(n_1 - 1)S_x^2 + (n_2 - 1)S_y^2}{n_1 + n_2 - 2}}$$


7.5
6.0
4.5
3.0
1.5
0.0
-1.5
-3.0
-4.5
t-value



7.5
6.0
4.5
3.0
1.5
0.0
-1.5
-3.0
-4.5
t-value

- **Null hypothesis**: There is no difference between the two groups at voxel X.
- If a single test at a voxel has a t value with 5% likelihood of occurrence, we can reject the null hypothesis with 95% confidence.
- But that means if we perform many tests, we will incorrectly reject the null hypothesis in 5% of those tests: **false positives**.
- We need to correct for multiple comparisons.

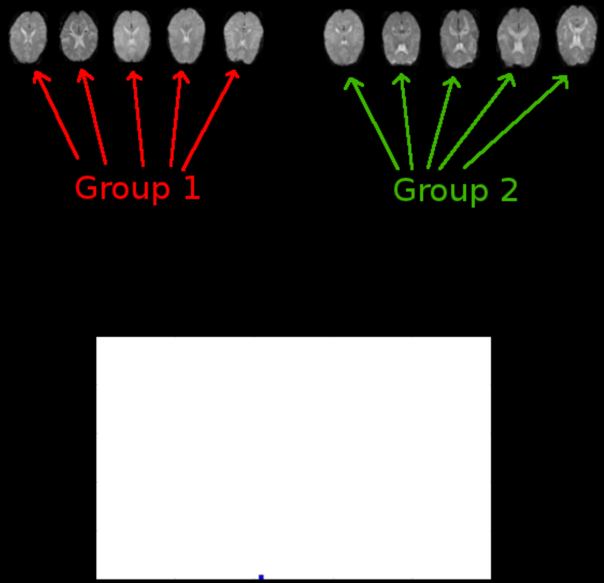
- Bonferroni correction
- False discovery rate
- Gaussian Random Field Theory
- **Random Permutation**
- **Cluster size thresholding**

Random permutation

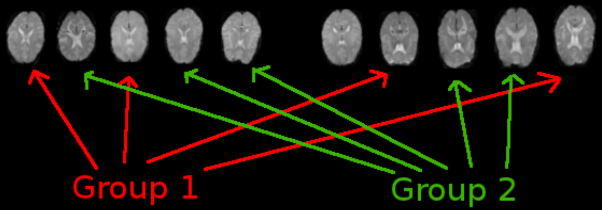
- *Parametric* statistical inferences makes assumptions about the data (e.g. comparing against an idealised, theoretical t-distribution).
- Why do this, when you can generate your own distribution from the data itself?
- Relabel your data set, and repeat the t-test for all possible relabellings (or for a random sample of them).

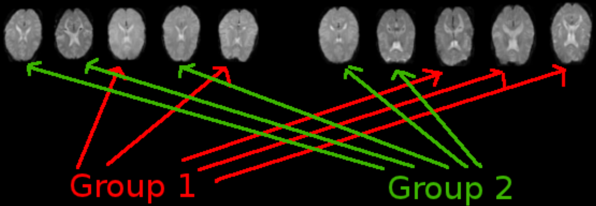
Random permutation

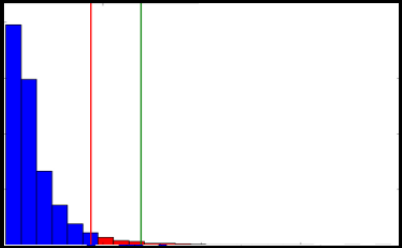
- For each relabelling of your groups, create a t image (results of a t test at every voxel).
- Save the maximum t value which occurs in this t image, and the t images for every relabelling. In this way, a *Maximal t distribution* is generated.
- Use the t value at the 95th percentile of this distribution to threshold your observed t image - all voxels with a t value greater than this threshold are significant.



Group 1
Group 2

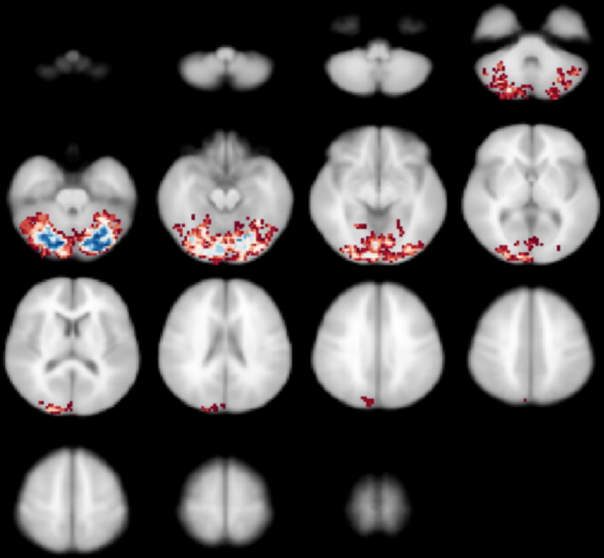

Group 1
Group 2


Group 1
Group 2



Cluster size thresholding

- If a voxel with a t-value of (say) 3.1 has a 5% likelihood, what is the likelihood that a cluster of 100 adjacent voxels all have t-values greater than (say) 2.3?
- Build a null-distribution of the likelihood of a cluster of size X occuring, using the random permutation techinque.
- Use the cluster size at the 95th percentile of the null distribution as your threshold - all clusters larger than this threshold are declared as significant.



The End

Questions?