DISCOS SPM course, CRC, Liège, 2009

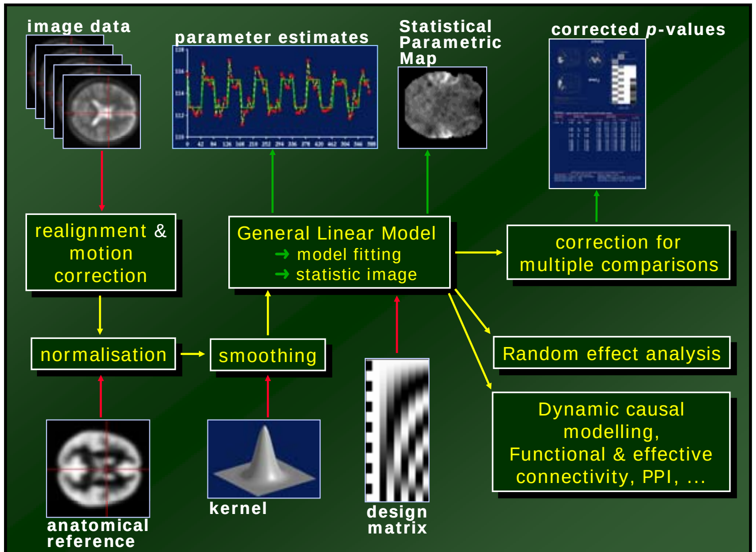
The General Linear Model, Part I

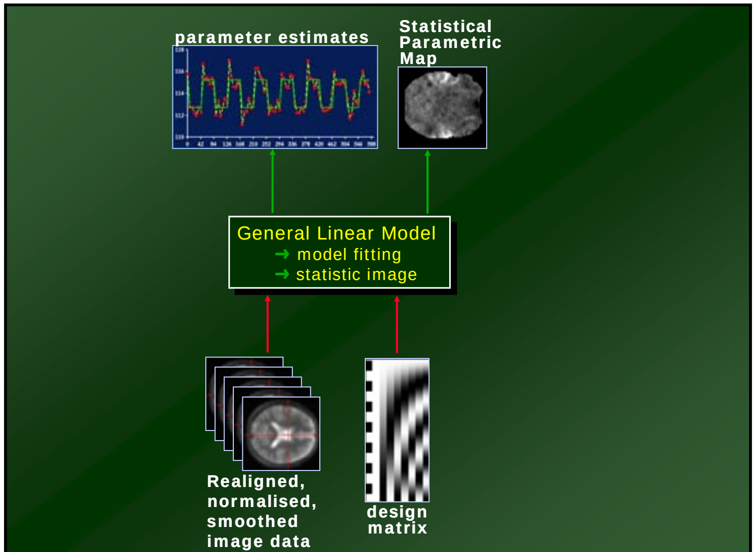
C. Phillips, Centre de Recherches du Cyclotron, ULg, Belgium

Based on slides from: JB. Poline

Contents

- **Introduction**
- The General Linear Model
  - Data & model
  - Design matrix
  - Parameter estimates & interpretation
  - Simple contrast
- « Take home » message





General Linear Model

What does it mean ?

General **Linear** **Model**

General Linear Model

What does it mean ?

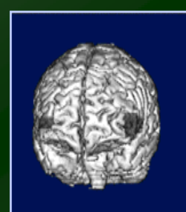
General
The model can be used to answer a wide variety of questions.

Linear
The model uses simple linear relationships between the variables.

Model
A set of equations are used to describe the data. Questions about the data can then be stated as mathematical expressions.

GLM is the basic model or general framework underlying the analysis of variance and multiple regression.

SPM key concepts…

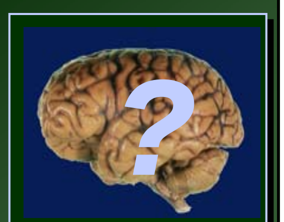


…a voxel by voxel hypothesis testing approach

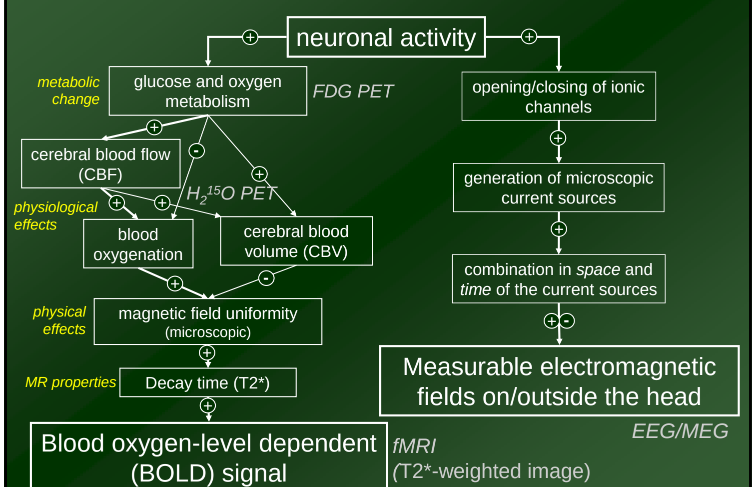
→ ***reliably identify regions showing a significant experimental effect of interest***

Key concepts

- Type I error
  - significance test at each voxel
- Parametric statistics
  - parametric model for voxel data, test model parameters
- No exact prior anatomical hypothesis
  - multiple comparisons

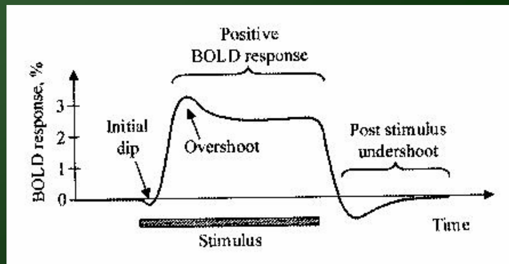


Functional neuroimaging signal



Source: Doug Noll's primer

Hemodynamic Response Function



% signal change
- = (point – baseline)/baseline
- usually 0.5-3%

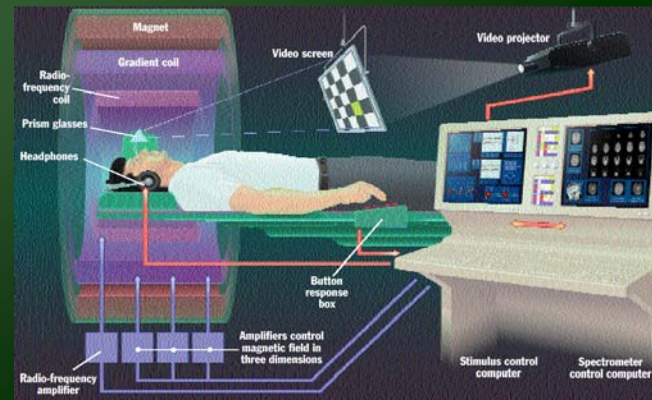
initial dip
- -more focal and potentially a better measure
- -somewhat elusive so far, not everyone can find it

time to rise
- signal begins to rise soon after stimulus start

time to peak
- signal peaks 4-6 sec after stimulus begins

post stimulus undershoot
- signal suppressed after stimulation ends

fMRI / PET Setup

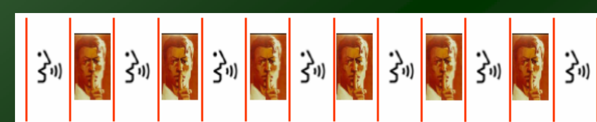


Contents

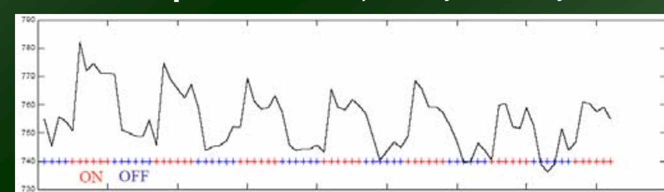
- Introduction
- **The General Linear Model**
  - Data & model
  - Design matrix
  - Parameter estimates & interpretation
  - Simple contrast
- « Take home » message

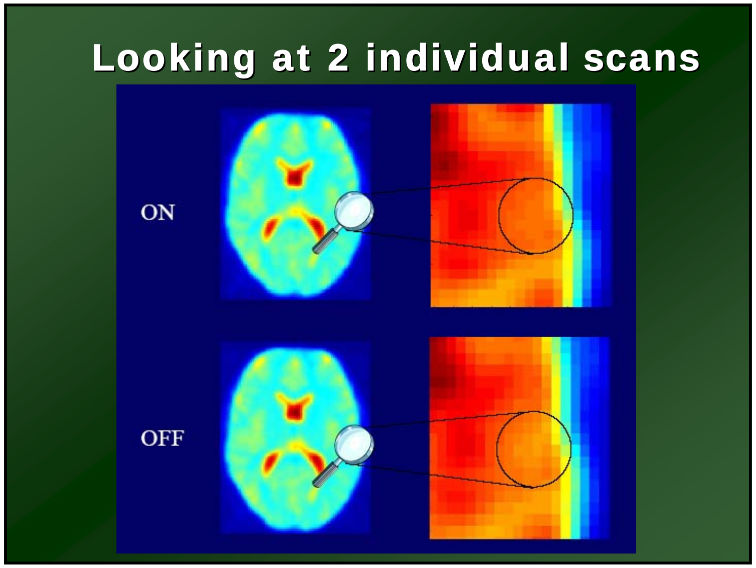
A simple experiment

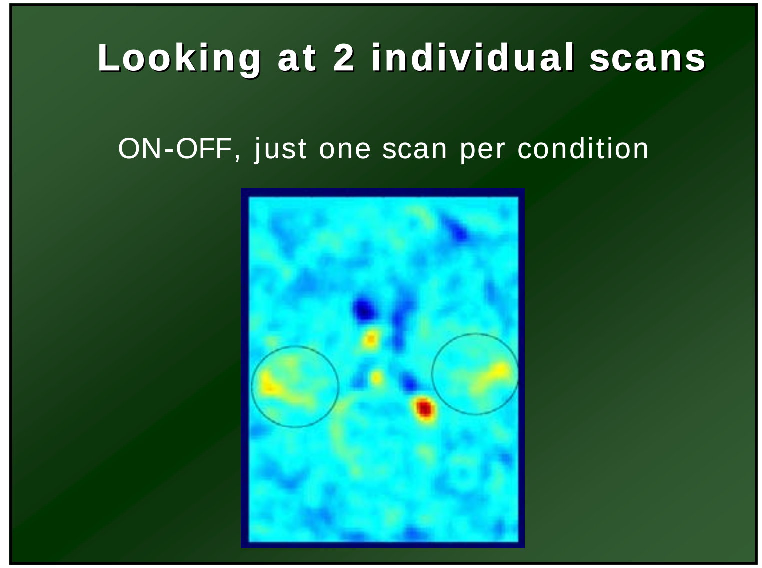
Stimuli: passive word listening versus rest

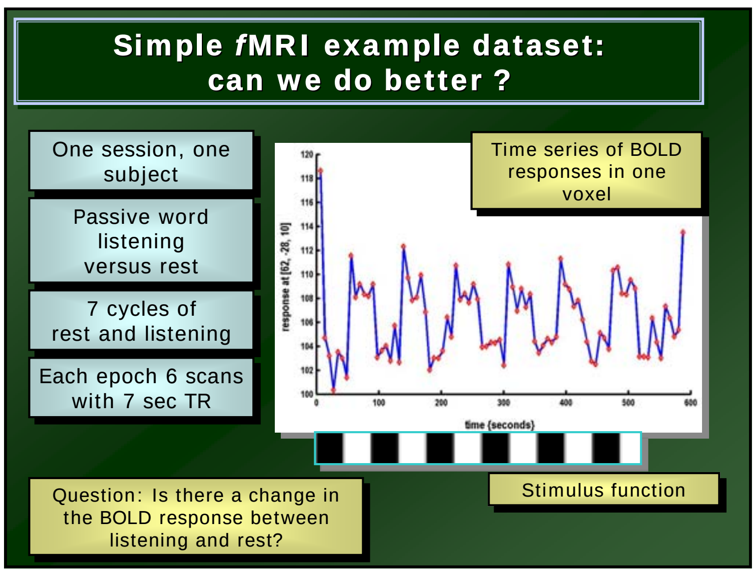


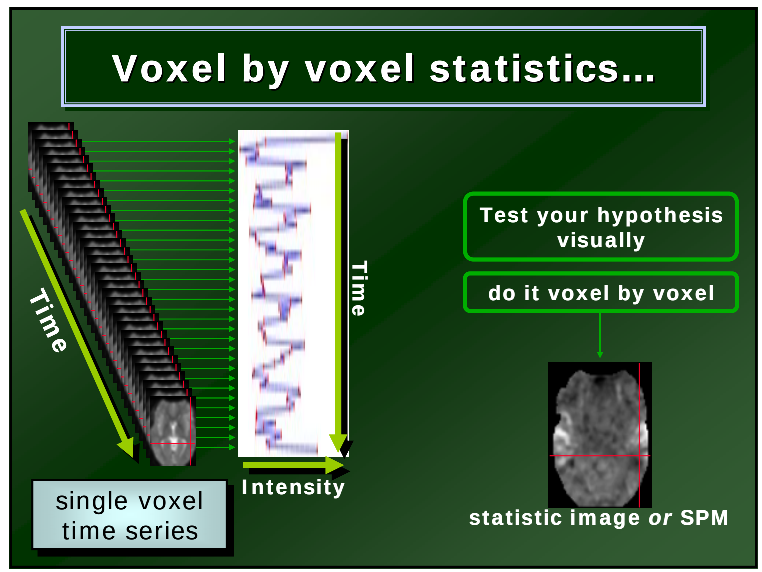
BOLD response in the primary auditory cortex




Looking at 2 individual scans
ON
OFF


Looking at 2 individual scans
ON-OFF, just one scan per condition

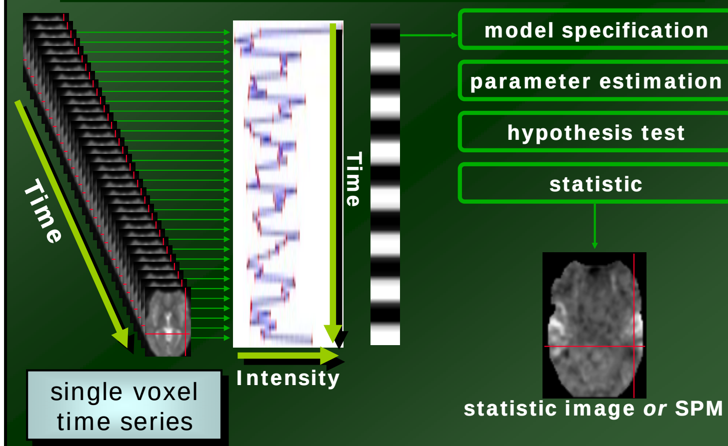

Simple *f*MRI example dataset:
can we do better ?
One session, one subject
Passive word listening versus rest
7 cycles of rest and listening
Each epoch 6 scans with 7 sec TR
Time series of BOLD responses in one voxel
response at [62, -28, 10]
time (seconds)
Stimulus function
Question: Is there a change in the BOLD response between listening and rest?


Voxel by voxel statistics…
Time
Time
Intensity
single voxel time series
Test your hypothesis visually
do it voxel by voxel
statistic image *or* SPM

Why do we need stats?

- Clearly voxel surfing isn't a viable option. We'd have to do it 1,000 of times and it would require a lot of subjective decisions about whether activation was real
- Statistics:
  - tell us where to look for activation that is related to our paradigm
  - help us decide how likely it is that activation is "real"

Voxel by voxel statistics…



Classical statistics…

- **parametric**
  - **one sample *t*-test**
  - **two sample *t*-test**
  - **paired *t*-test**
  - **Anova**
  - **AnCova**
  - **correlation**
  - **linear regression**
  - **multiple regression**
  - ***F*-tests**
  - **etc…**

all cases of the
General Linear Model
assume normality
to account for serial correlations:
Generalised Linear Model

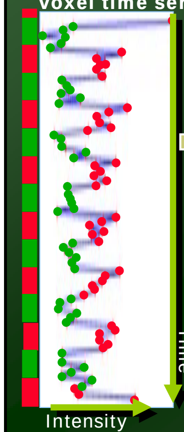
- **non-parametric?** → **SnPM**

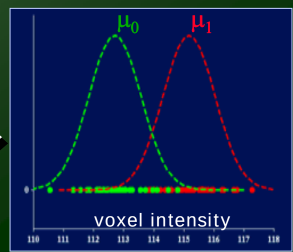
Statistics

- Formal statistics are just doing what your eyeball test of significance did
  - Estimate how likely it is that the signal is real given how noisy the data is
- "p value" = probability value of the "null hypothesis".
  - Null hypothesis = "No activation" !
  - confidence: how likely is it that the results could occur purely due to chance?
  - If "p = .03", that means there is a 3% chance that the results are bogus
- By convention, if the probability that a result could be due to chance is less than 5% (p < .05), we say that result is statistically significant
- Significance depends on
  - signal (differences between conditions)
  - noise (other variability)
  - sample size (more time points are more convincing)

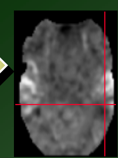
...e.g. two-sample *t*-test?

voxel time series





$$t = \frac{\mu_1 - \mu_0}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n_1} + \frac{1}{n_0}\right)}}$$



t-statistic image SPM{*t*}

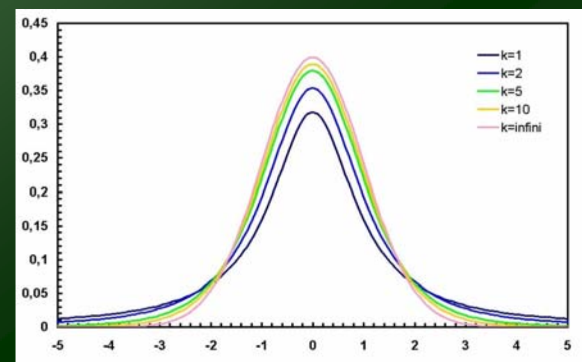
Question: Is there a change in the BOLD response between listening and rest?

compares size of effect to its error standard deviation

- *standard t-test assumes independence ⇒ ignores temporal autocorrelation!*

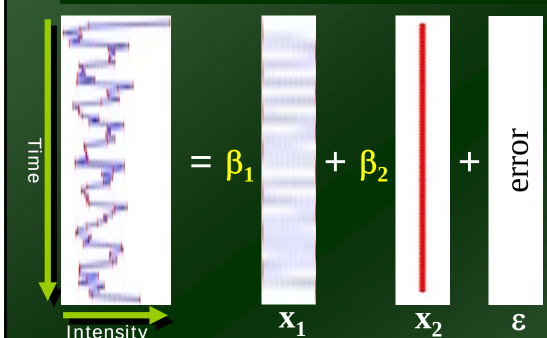
t distribution

Depends on 1 set of degrees of freedom !



ν	95%
1	6.314
5	2.015
10	1.812
15	1.753
20	1.725
30	1.697
50	1.676
100	1.660
Inf	1.645

Regression example...

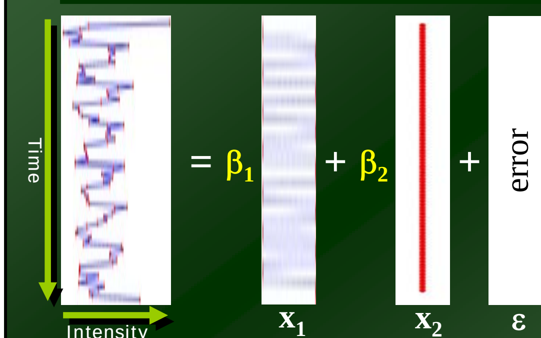


$\varepsilon \sim N(0, \sigma^2 I)$
(error is normal and independently and identically distributed)

Question: Is there a change in the BOLD response between listening and rest?

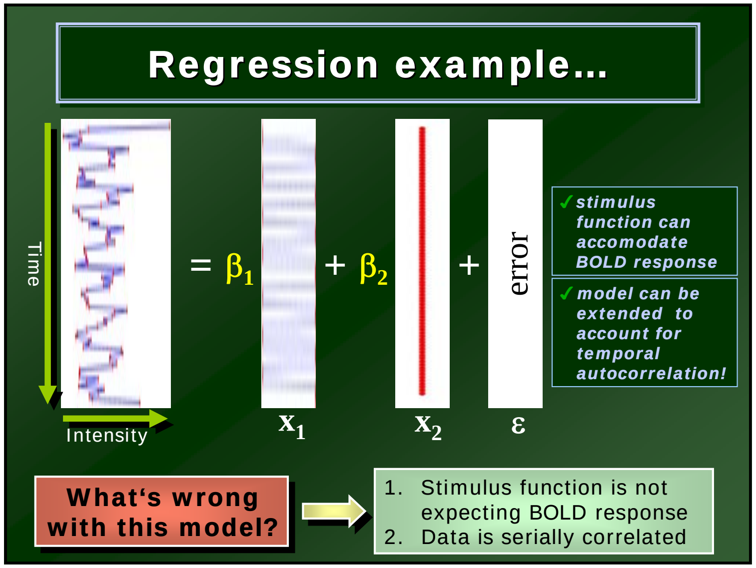
Hypothesis test: $\beta_1 = 0$? (using t-statistic)

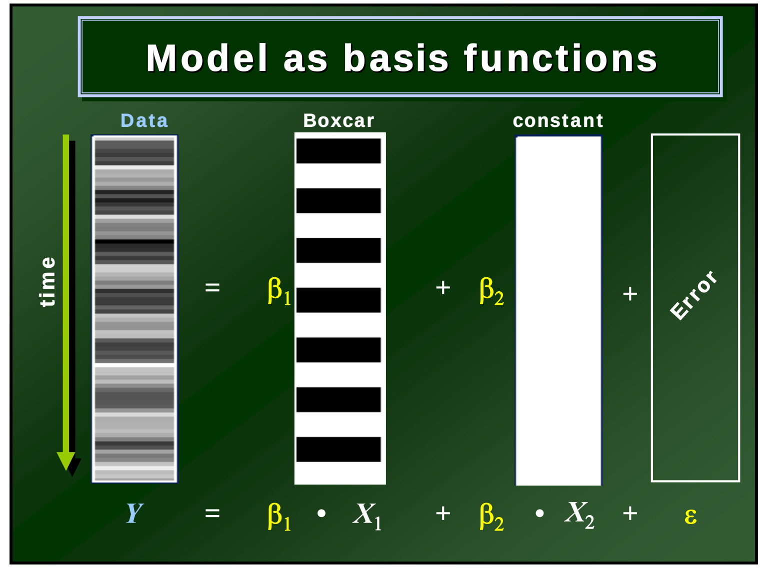
Regression example...

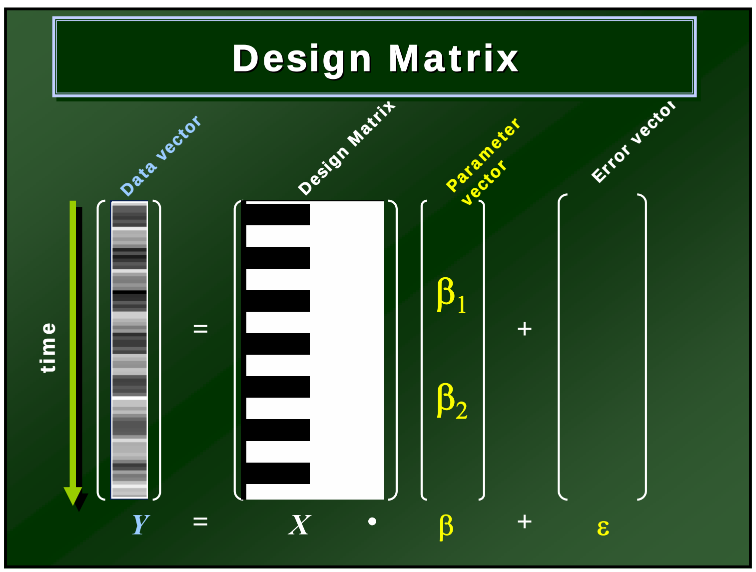


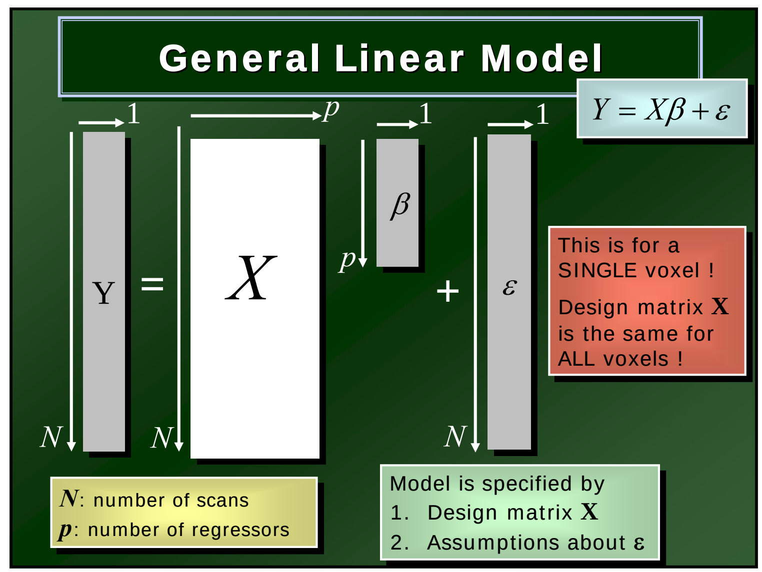
$\varepsilon \sim N(0, \sigma^2 I)$
(error is normal and independently and identically distributed)

- correlation: test H_0: $\rho = 0$ equivalent to test H_0: $\beta_1 = 0$
- two-sample *t*-test: test H_0: $\mu_0 = \mu_1$ equivalent to test H_0: $\beta_1 = 0$


Regression example…
Time
Intensity
= β1 + β2 + error
x1
x2
ε
✓ stimulus function can accomodate BOLD response
✓ model can be extended to account for temporal autocorrelation!
What's wrong with this model?
1. Stimulus function is not expecting BOLD response
2. Data is serially correlated


Model as basis functions
Data
Boxcar
constant
time
= β1 + β2 + Error
Y = β1 • X1 + β2 • X2 + ε


Design Matrix
Data vector
Design Matrix
Parameter vector
Error vector
time
= β1 β2 +
Y = X • β + ε


General Linear Model
Y = Xβ + ε
1
p
1
1
Y = X β + ε
N
N
p
N
This is for a SINGLE voxel !
Design matrix X is the same for ALL voxels !
N: number of scans
p: number of regressors
Model is specified by
1. Design matrix X
2. Assumptions about ε

General Linear Model

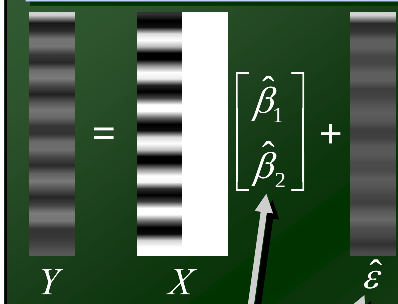
- ***f*MRI time series:** $Y_1, \ldots, Y_s, \ldots, Y_N$
  - acquired at times $t_1, \ldots, t_s, \ldots, t_N$
- **Model: Linear combination of basis functions**

  $Y_s = \beta_1 f^1(t_s) + \ldots + \beta_l f^l(t_s) + \ldots + \beta_L f^L(t_s) + \varepsilon_s$
- $f^l(.)$: ***basis* functions**
  - ***"reference waveforms"***
  - ***dummy variables***
- β_l: ***parameters*** *(fixed effects)*
  - **amplitudes of basis functions** *(regression slopes)*
- ε_s: ***residual errors*:** $\varepsilon_s \sim N(0,\sigma^2)$
  - **identically distributed**
  - **independent, or serially correlated**

    (Generalised Linear Model → GLM)

Parameter estimation

$Y = X\beta + \varepsilon$



$$\begin{bmatrix}\hat\beta_1 \\ \hat\beta_2\end{bmatrix} \qquad \hat\varepsilon$$

$\hat\varepsilon = Y - X\hat\beta$

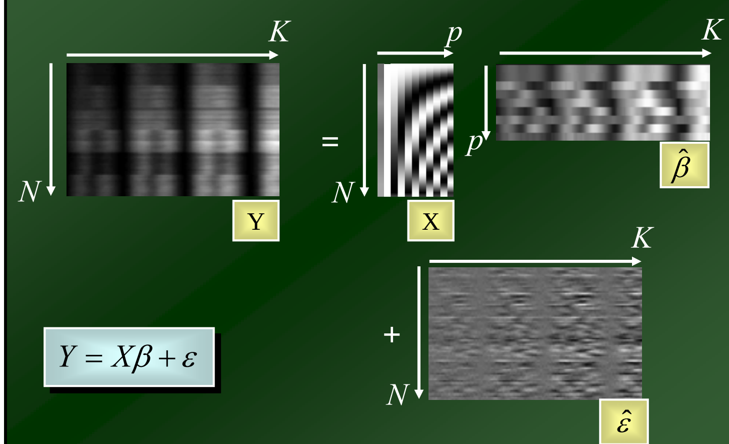
residuals

Assume iid error

Estimate parameters such that $\sum_{t=1}^{N} \hat\varepsilon_t^2$ minimal → **Least squares** parameter estimate

$\hat\beta = (X^T X)^{-1} X^T Y$

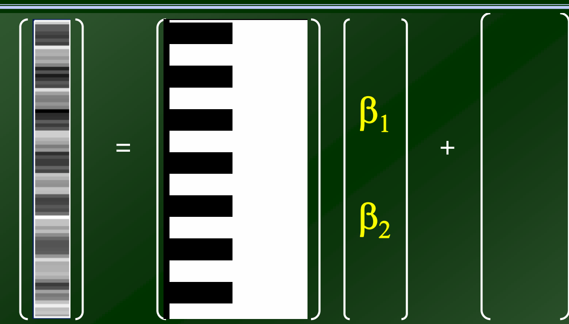
Mass univariate approach



$\hat\beta$ $\quad \hat\varepsilon$

$Y = X\beta + \varepsilon$

Design Matrix and contrast



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

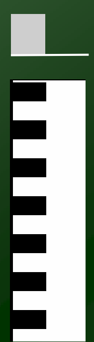
Question: Is there a change in the BOLD response between listening and rest?

Hypothesis test: $\beta_1 = 0$? (using t-statistic)

Model — contrasts — SPM{t}

Contrast **: specifies linear combination of parameter vector**: $c'\beta$

$c' = +1 \quad 0$



box-car amplitude > 0 ?

=

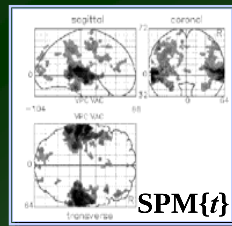
$\hat{\beta}_1 > 0$? ($\hat{\beta}_1$: estimation of β_1)

=

$1 \times \hat{\beta}_1 + 0 \times \hat{\beta}_2 > 0$?

=

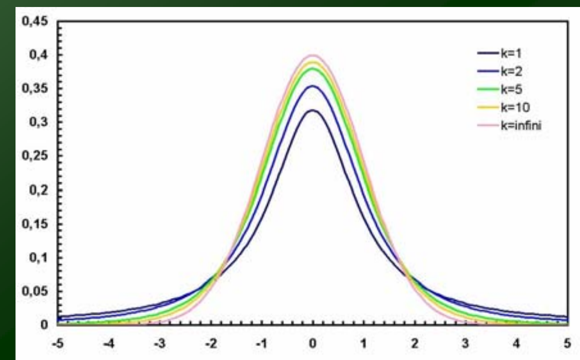
test H_0 : $c' \times \hat{\beta} > 0$?



$$T = \frac{\textit{contrast}\text{ of estimated parameters}}{\sqrt{\text{variance estimate}}} \longrightarrow T = \frac{c'\hat{\beta}}{\sqrt{s^2 c'(X'X)^{+}c}}$$

t distribution

Depends on 1 set of degrees of freedom !



ν	95%
1	6.314
5	2.015
10	1.812
15	1.753
20	1.725
30	1.697
50	1.676
100	1.660
Inf	1.645

How is this computed ? (t-test)

Estimation [Y, X] [b, s]

$Y = X\beta + \varepsilon$ $\qquad$ $\varepsilon \sim \sigma^2 N(0,I)$ (Y : at one position)

$\hat{\beta} = (X'X)^{+} X'Y$ $\qquad$ ($\hat{\beta}$ = estimation of β) → beta??? images

$e = Y - X\hat{\beta}$ $\qquad$ (e = estimation of ε)

$s^2 = (e'e/(n-p))$ $\qquad$ (s = estimation of σ, n: scans, p: parameters)

→ 1 image ResMS

Test [b, s^2, c] [$c'b$, t]

$\mathrm{Var}(c'\hat{\beta}) = s^2 c'(X'X)^{+}c$ $\qquad$ compute for each contrast c

$t = c'\hat{\beta} / \mathrm{sqrt}(s^2 c'(X'X)^{+}c)$ $\qquad$ $c'\hat{\beta}$ → images spm_con???

compute the t images → images spm_t???

under the null hypothesis H_0 : $t \sim Student(df)$ $\qquad$ $df = n - p$

Example: a line through 3 points...

simple linear regression

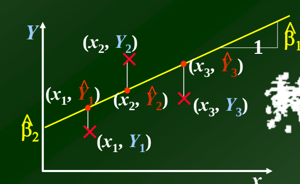
$Y_i = \beta_1 x_i + \beta_2 + \varepsilon_i \qquad i = 1,2,3$

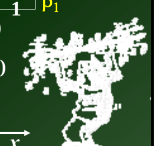
$Y_1 = \beta_1 x_1 + \beta_2 \times 1 + \varepsilon_1$

$Y_2 = \beta_1 x_2 + \beta_2 \times 1 + \varepsilon_2$

$Y_3 = \beta_1 x_3 + \beta_2 \times 1 + \varepsilon_3$

dummy variables





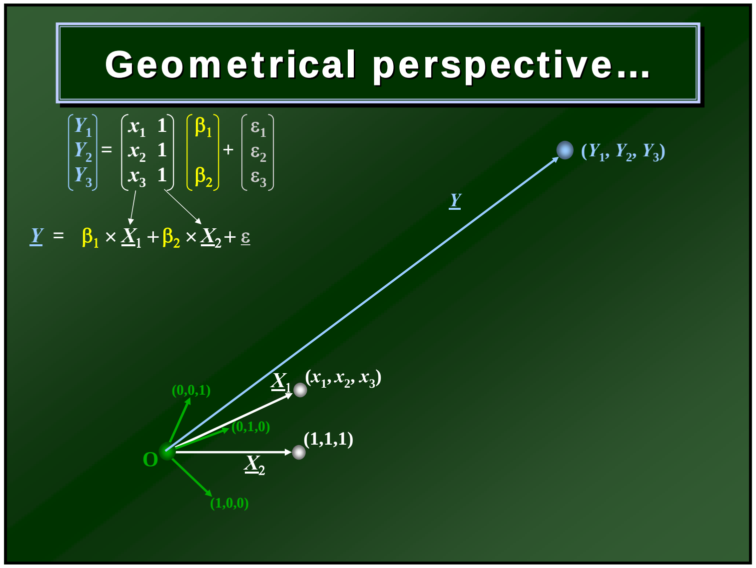
$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

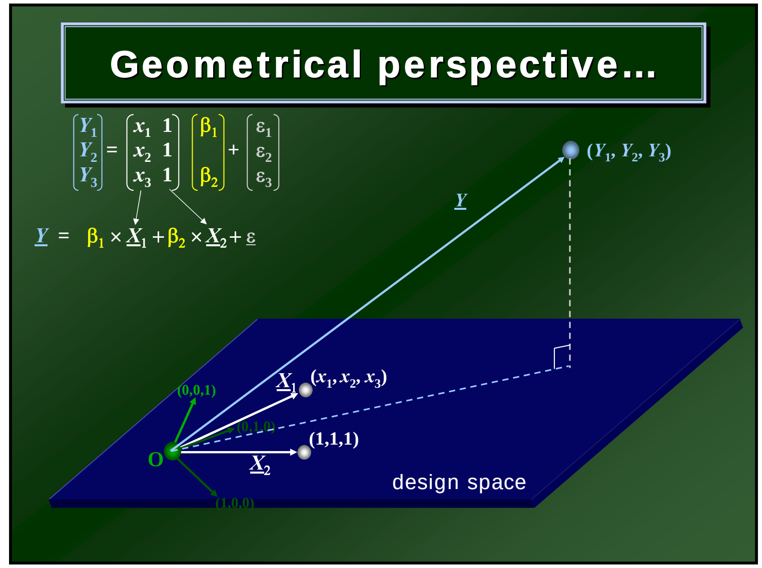
$\underline{Y} = \underline{X} \; \underline{\beta} + \underline{\varepsilon}$

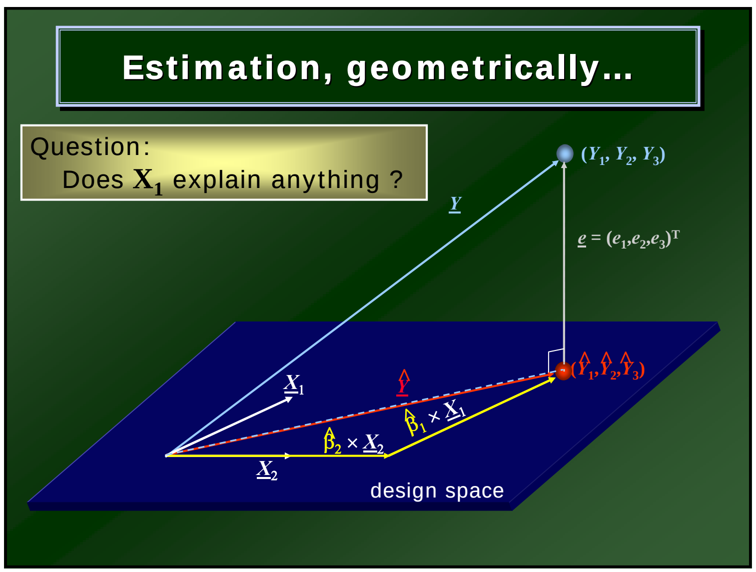
parameter estimates $\hat{\beta}_1$ & $\hat{\beta}_2$

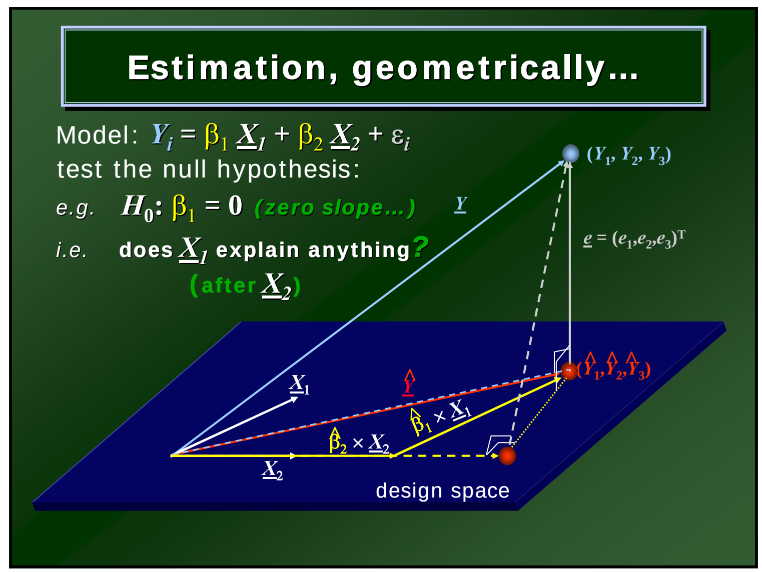
fitted values $\hat{Y}_1, \hat{Y}_2, \hat{Y}_3$

residuals e_1, e_2, e_3


Geometrical perspective…
Y = β1 × X1 + β2 × X2 + ε
(Y1, Y2, Y3)
Y
X1 (x1, x2, x3)
(0,0,1)
(0,1,0)
(1,1,1)
X2
O
(1,0,0)


Geometrical perspective…
Y = β1 × X1 + β2 × X2 + ε
(Y1, Y2, Y3)
Y
X1 (x1, x2, x3)
(0,0,1)
(1,1,1)
X2
O
design space


Estimation, geometrically…
Question:
Does X1 explain anything ?
(Y1, Y2, Y3)
Y
e = (e1,e2,e3)T
X1
X2
design space


Estimation, geometrically…
Model: Yi = β1 X1 + β2 X2 + εi
test the null hypothesis:
e.g. H0: β1 = 0 (zero slope…)
i.e. does X1 explain anything?
(after X2)
(Y1, Y2, Y3)
Y
e = (e1,e2,e3)T
X1
X2
design space

Contents

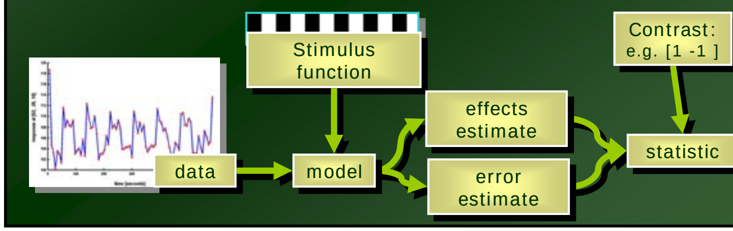
- Introduction
- The General Linear Model
  - Data & model
  - Design matrix
  - Parameter estimates & interpretation
  - Simple contrast
- **« Take home » message**

Why modelling?

Why? Make inferences about effects of interest

How?
1. Decompose data into effects and error
2. Form statistic using estimates of effects and error

Model? Use any available knowledge



Way to proceed

Prepare your questions.
ALL the questions !

Find a model which
- allows contrasts that translates these questions.
- takes into account ALL the effects (interaction, sessions,etc)

Devise task & stimulus presentation.

Acquire the data & analyse.

Not the other way round!!!