Edinburgh course – Avril 2010
Linear Models – Contrasts – Variance components

Jean-Baptiste Poline

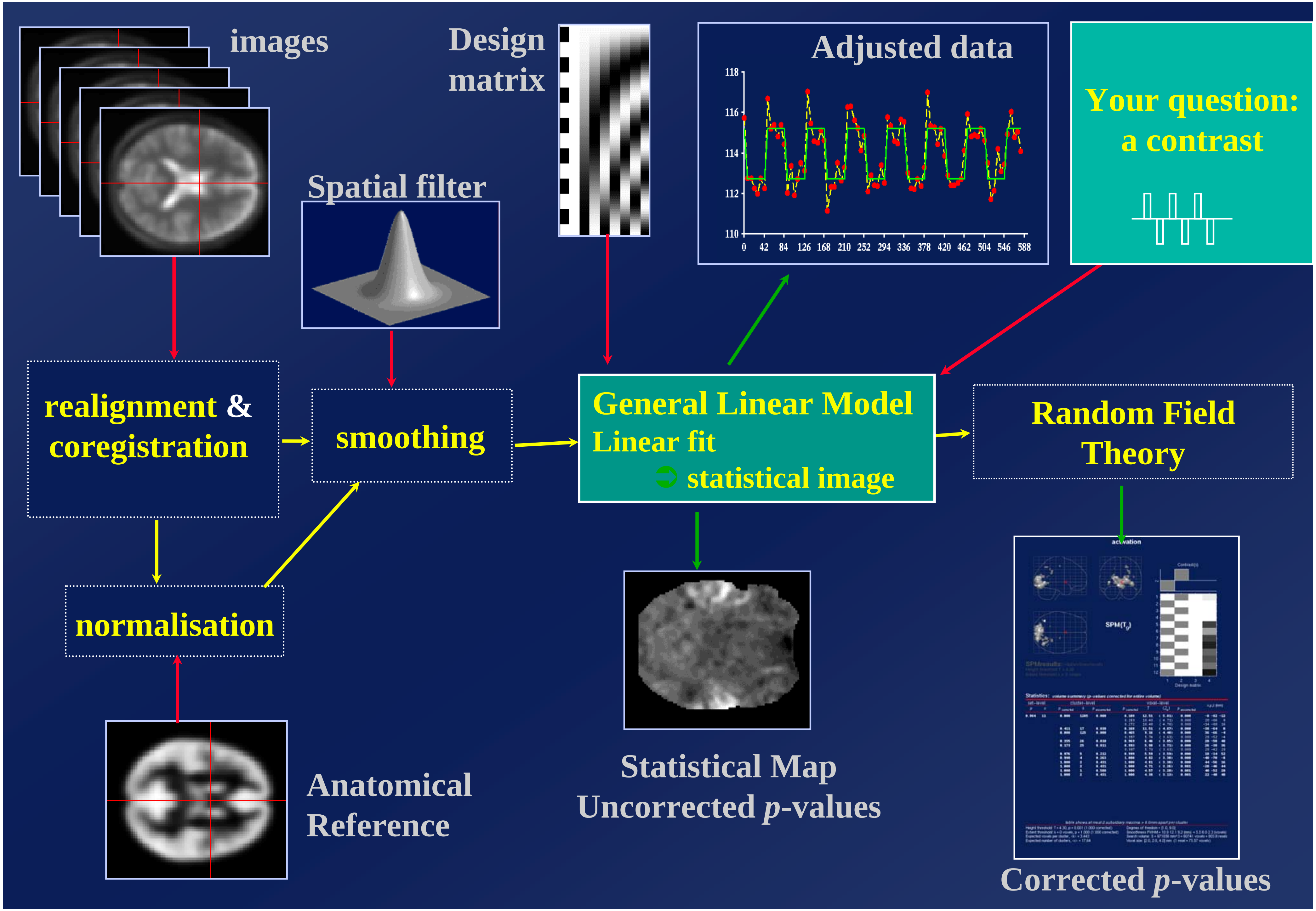
Neurospin, I2BM, CEA Saclay, France

Credits: Will Penny, G. Flandin, SPM course authors

Outline

- Part I: Linear model and contrast: going through it again and going further
- Part II: Variance component and group analyses*

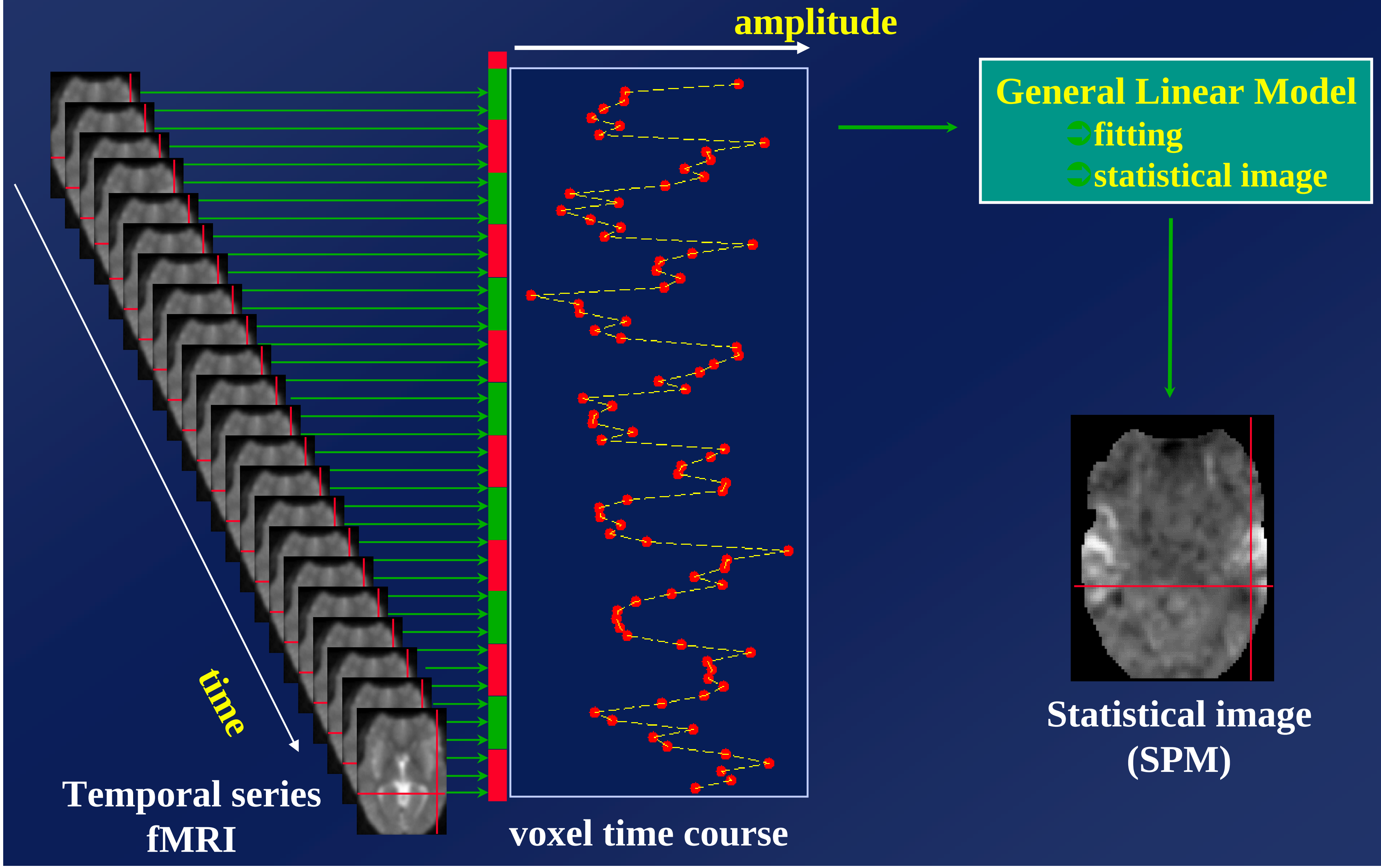
* (shamelessly stolen from Will Penny SPM course)


images
Design
matrix
Adjusted data
Your question:
a contrast
Spatial filter
realignment &
coregistration
smoothing
General Linear Model
Linear fit
statistical image
Random Field
Theory
normalisation
Anatomical
Reference
Statistical Map
Uncorrected p-values
Corrected p-values

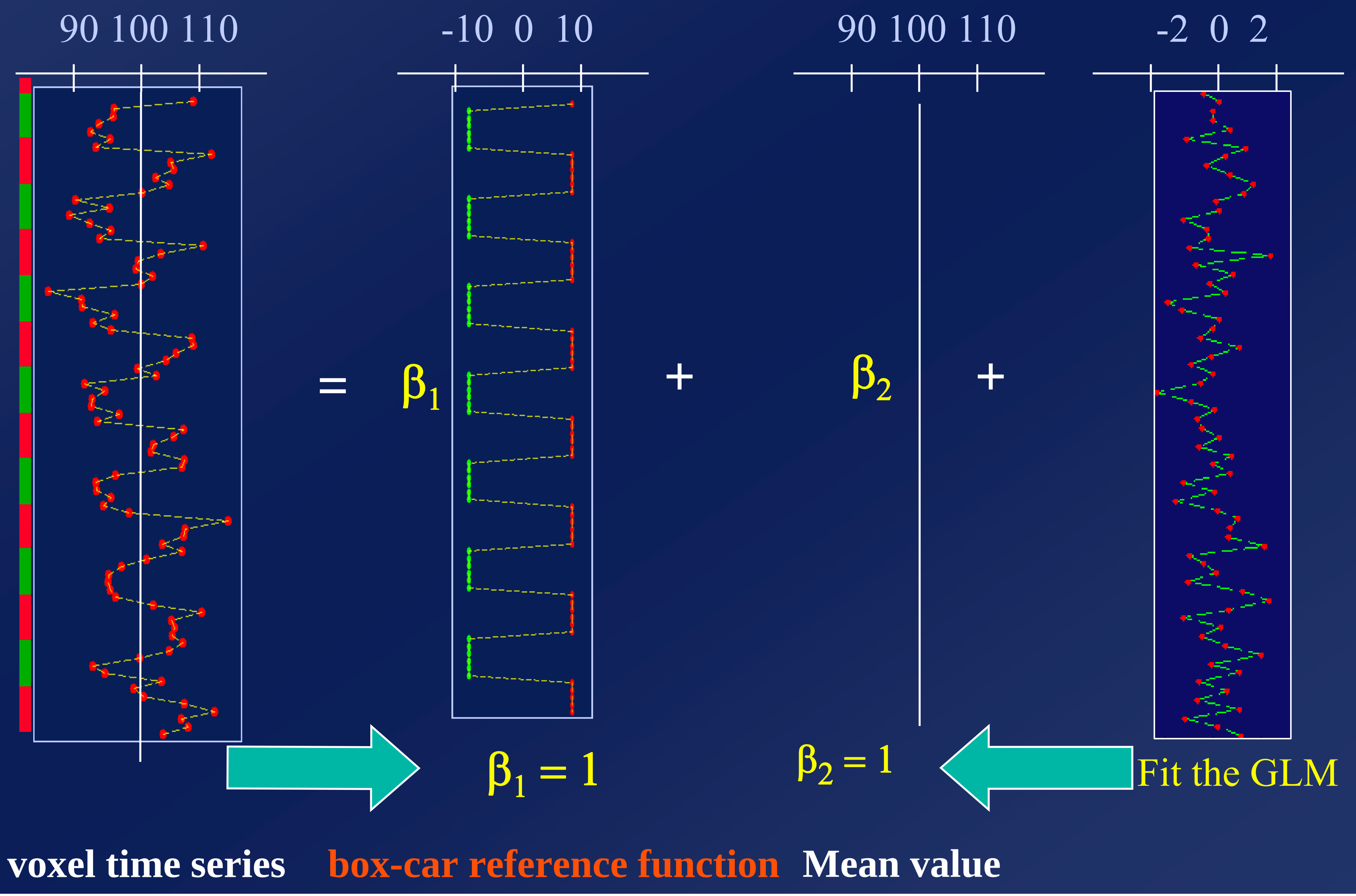
Plan

- *REPEAT: model and fitting the data with a Linear Model*
- *Make sure we understand the testing procedures : t- and F-tests*
- *But what do we test exactly ?*
- *Examples – almost real*

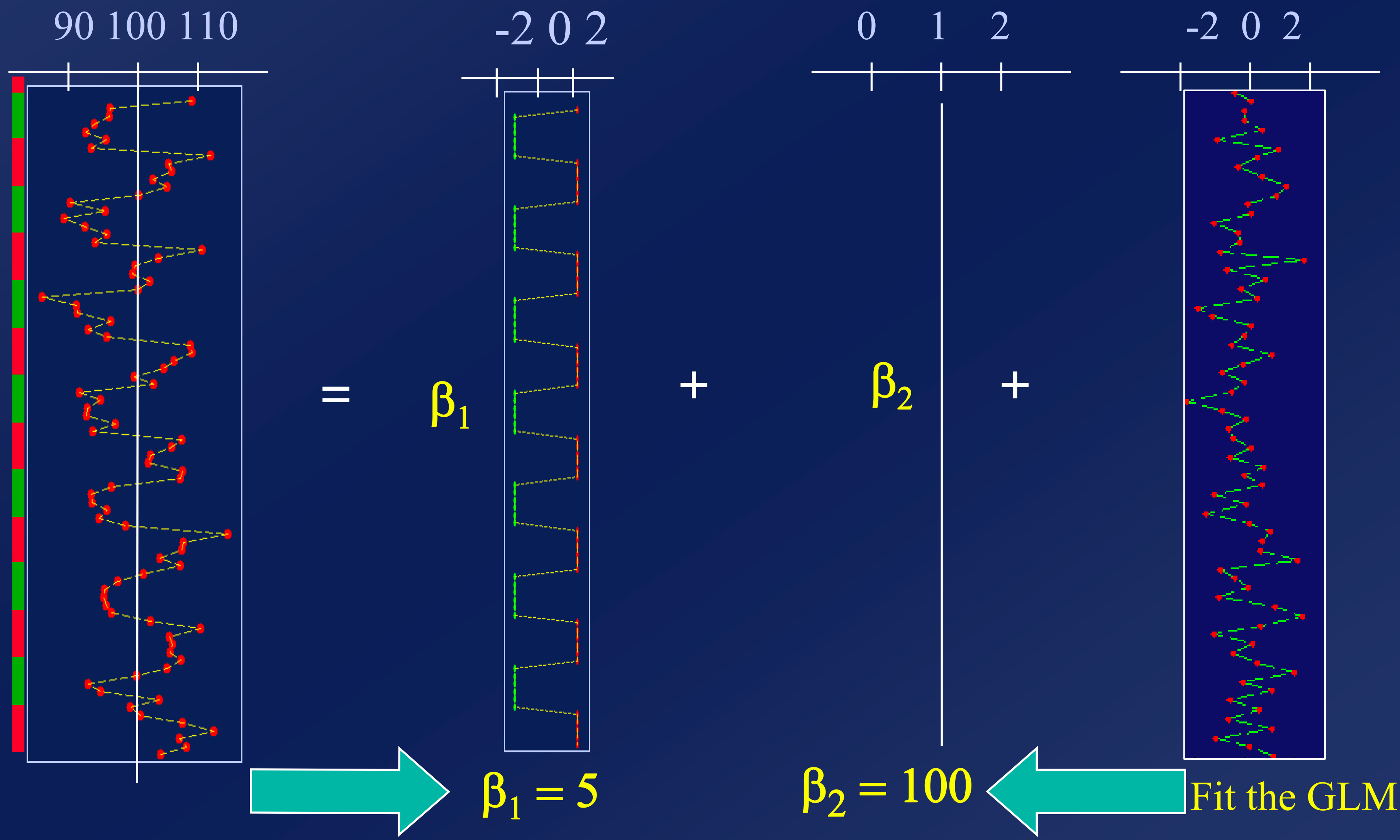
One voxel = One test (t, F, ...)



Regression example…

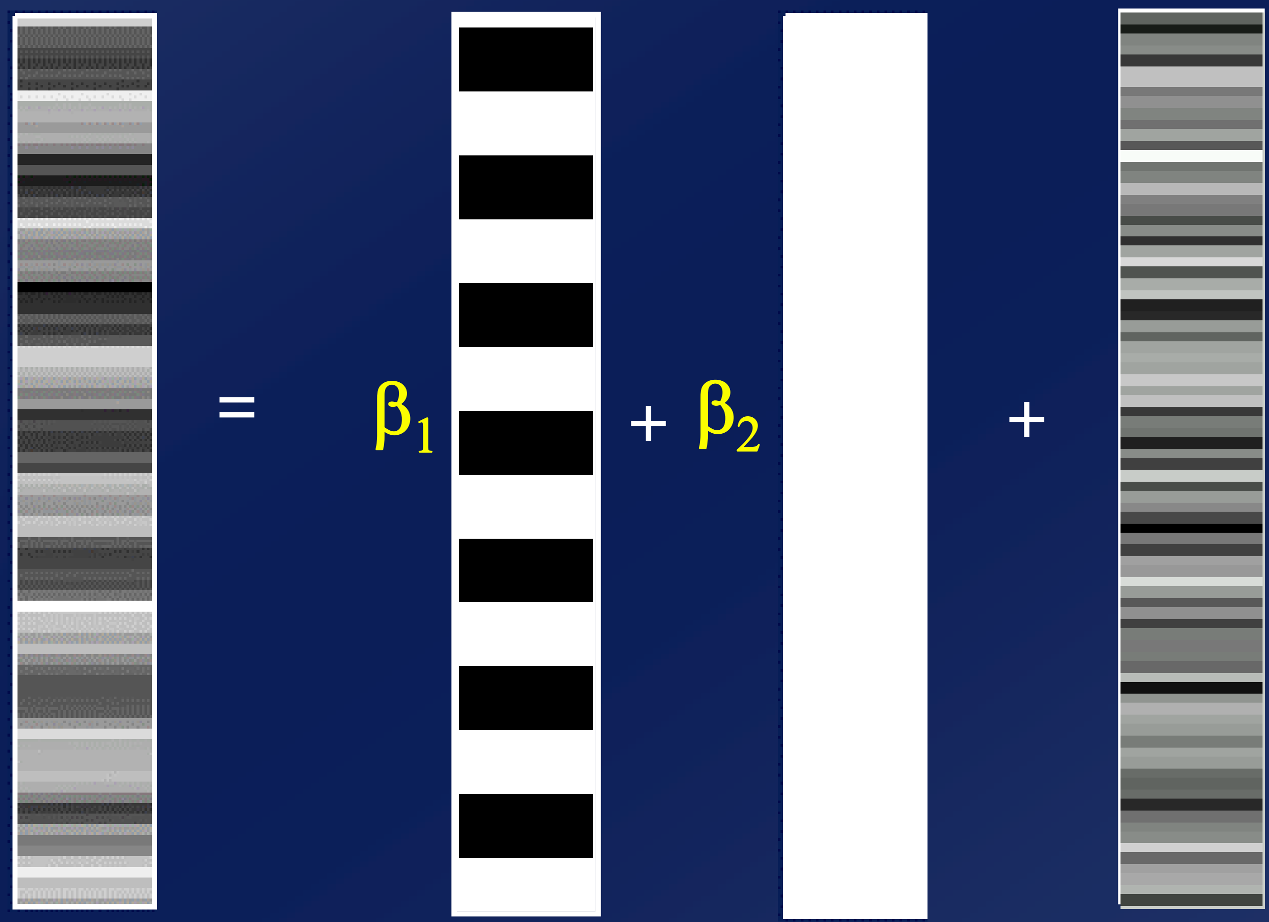


Regression example…



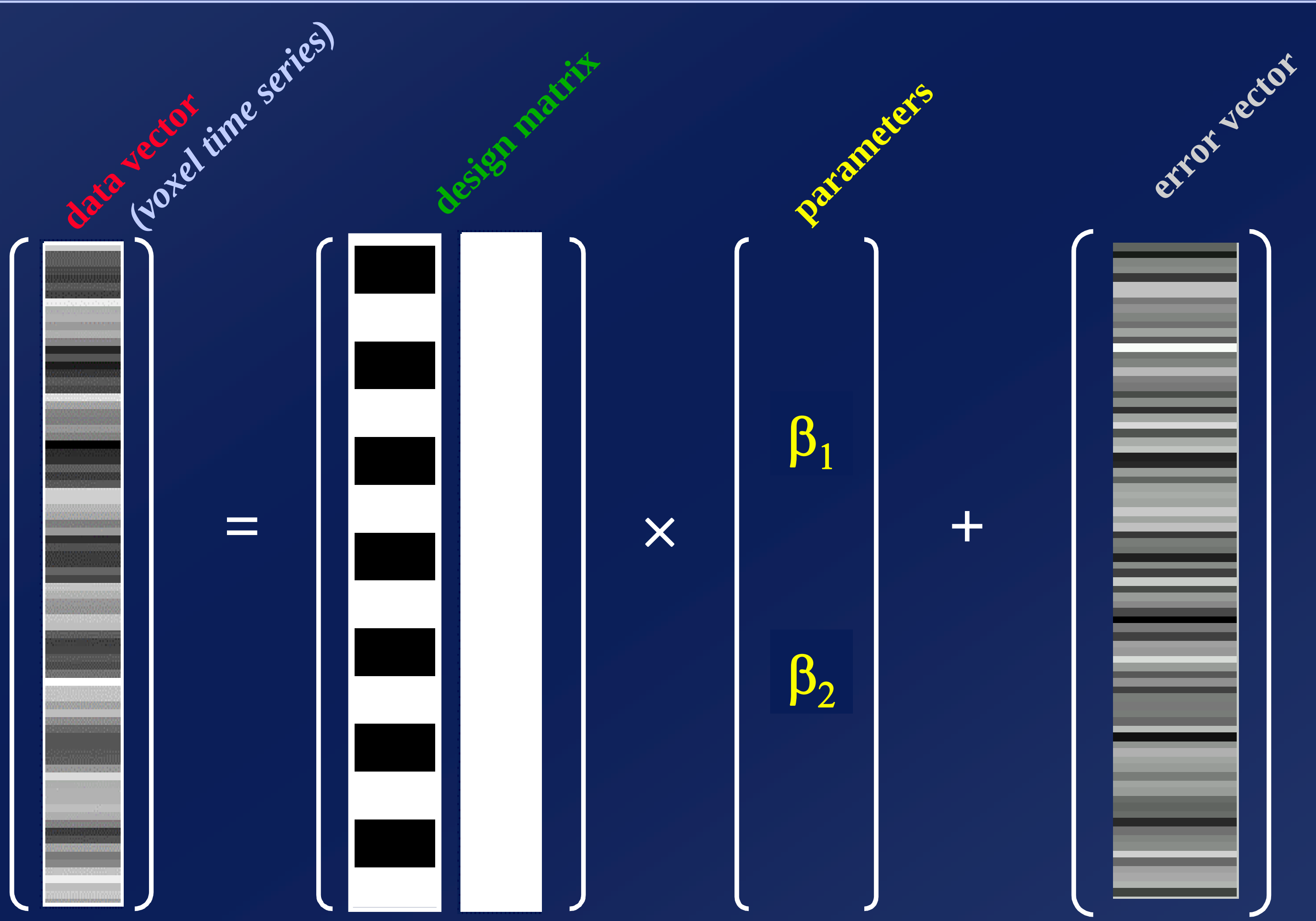
voxel time series **box-car reference function** **Mean value**

…revisited : matrix form



$$Y = \beta_1 \times f(t) + \beta_2 \times 1 + \varepsilon$$

Box car regression: design matrix…



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

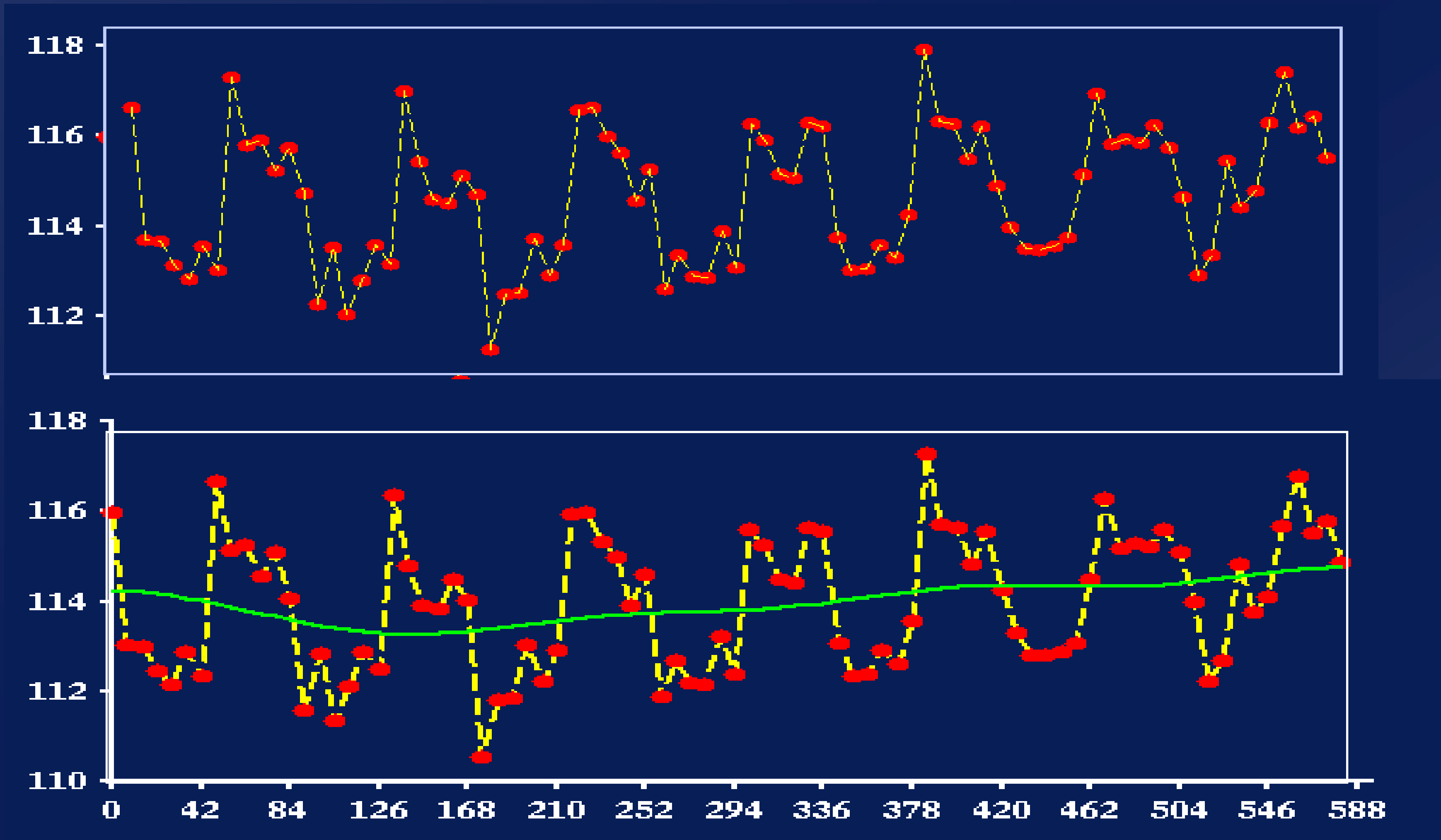
Fact: model parameters depend on regressors scaling

Q: When do I care ?

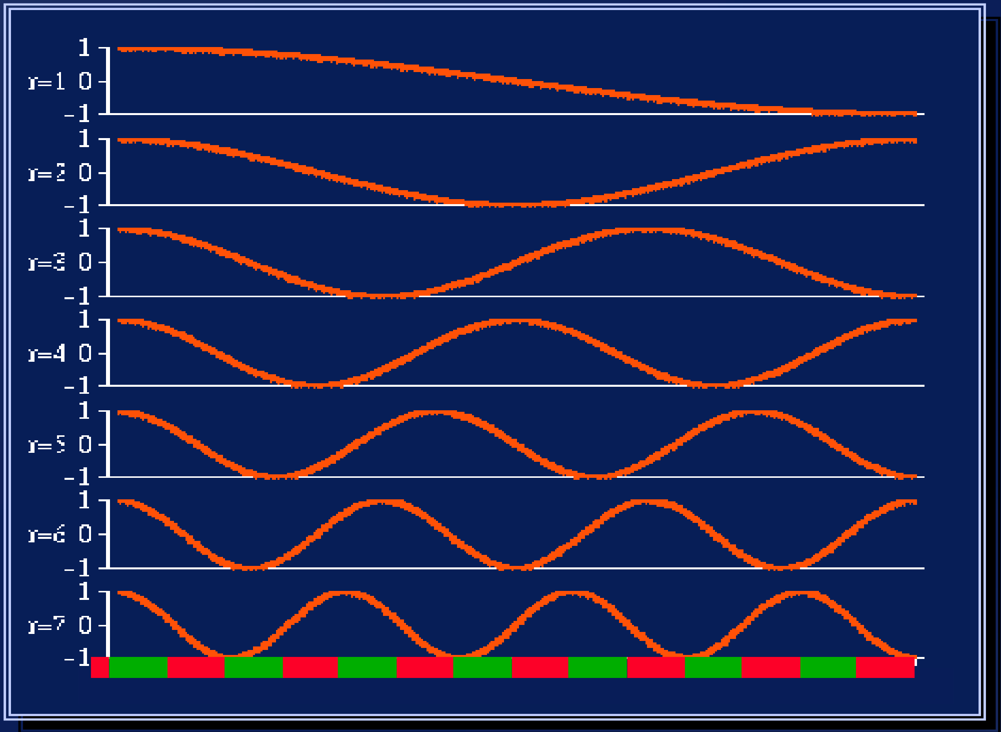
A: ONLY when comparing manually entered regressors (say you would like to compare two scores)

What if two conditions A and B are not of the same duration before convolution HRF?

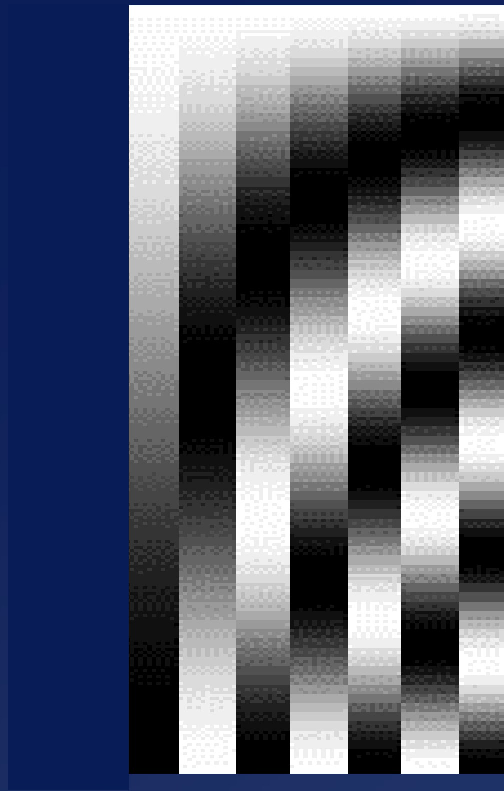
What if we believe that there are drifts?



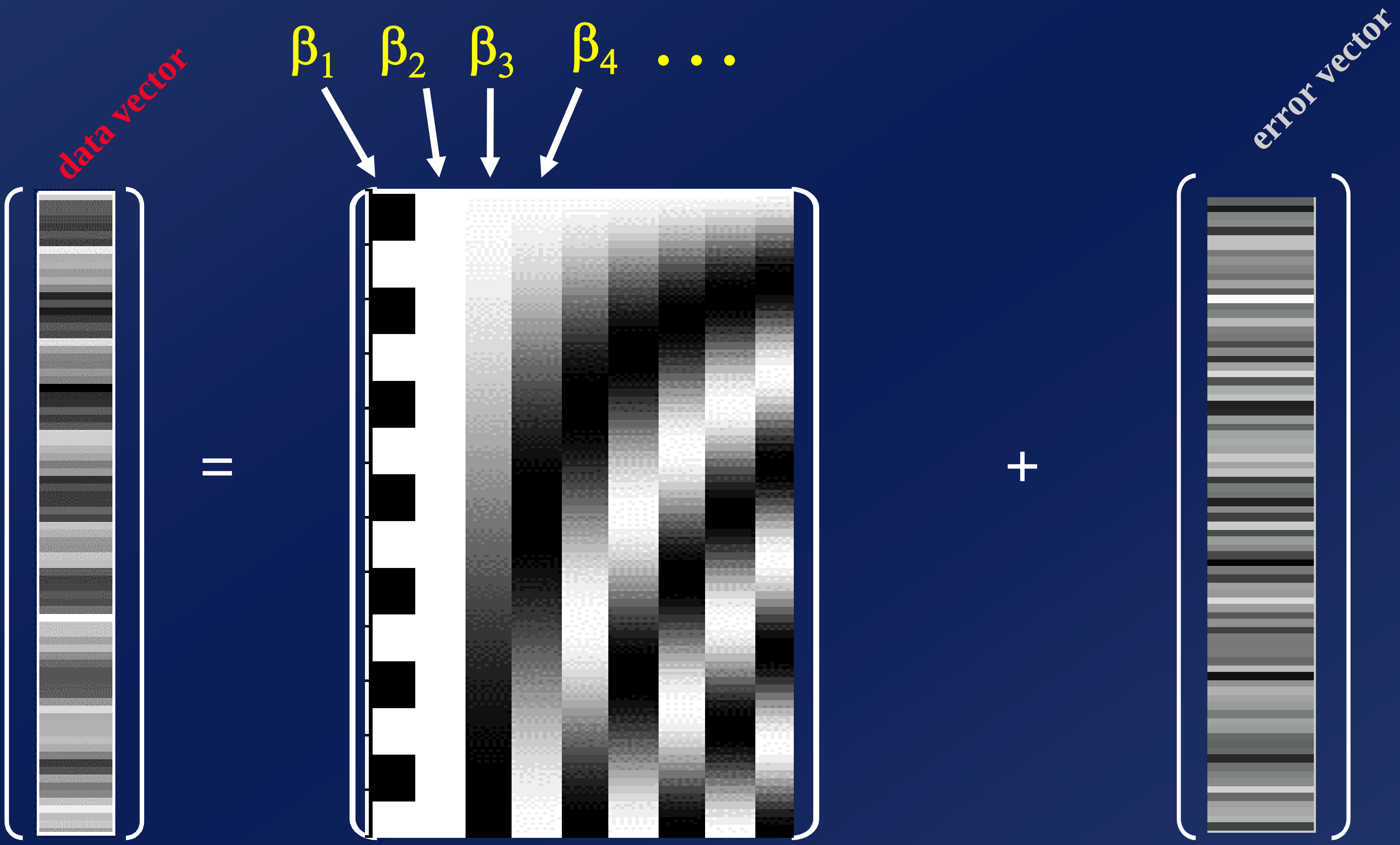
Add more reference functions / covariates ...



Discrete cosine transform basis functions

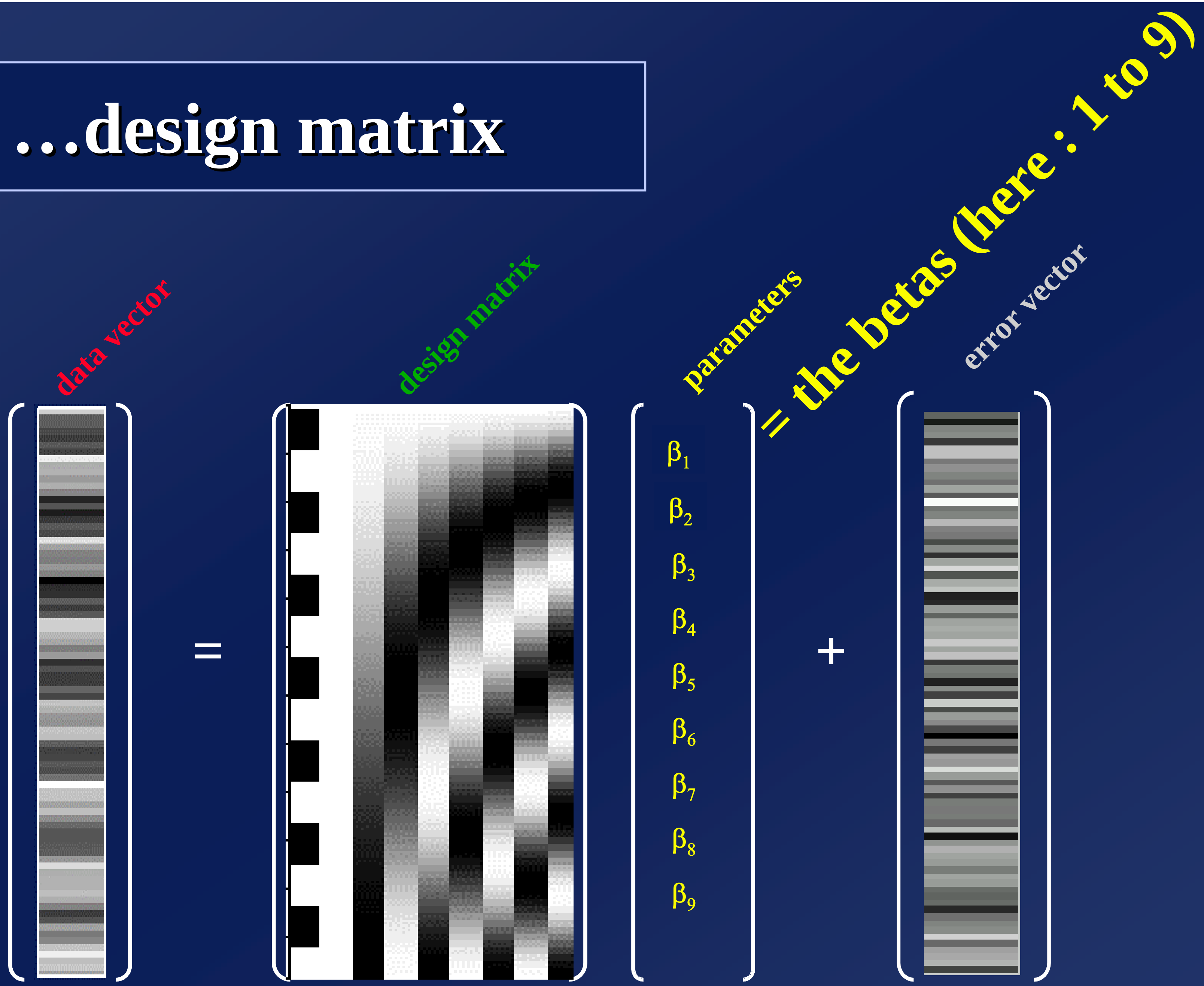


…design matrix



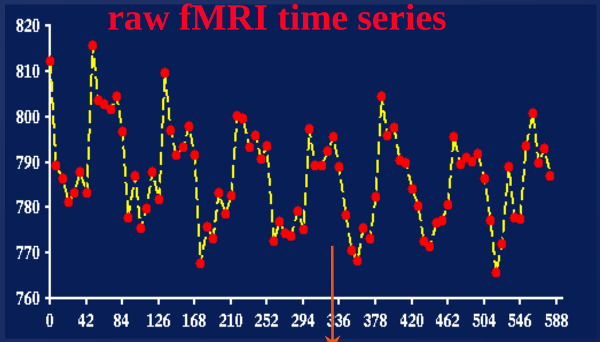
$$Y = X \times \beta + \varepsilon$$

…design matrix

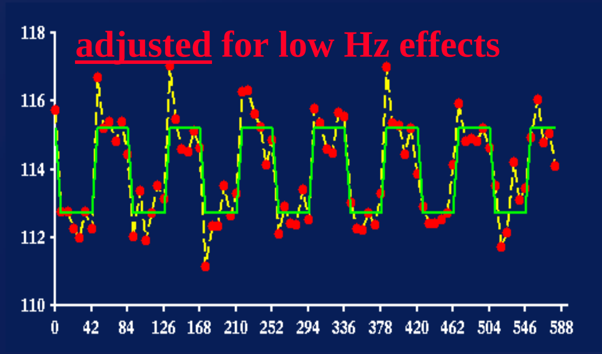


$$Y = X \times \beta + \varepsilon$$

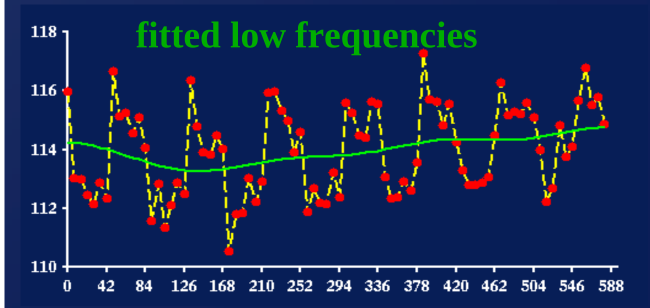
Fitting the model = finding some estimate of the betas



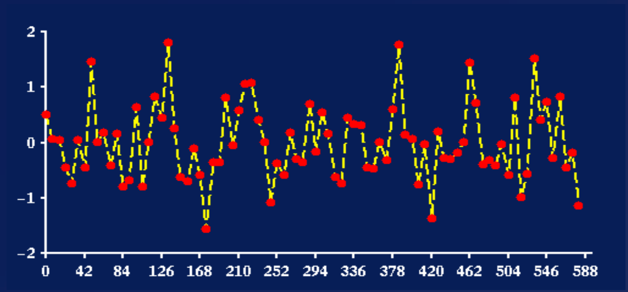
Raw data



fitted signal



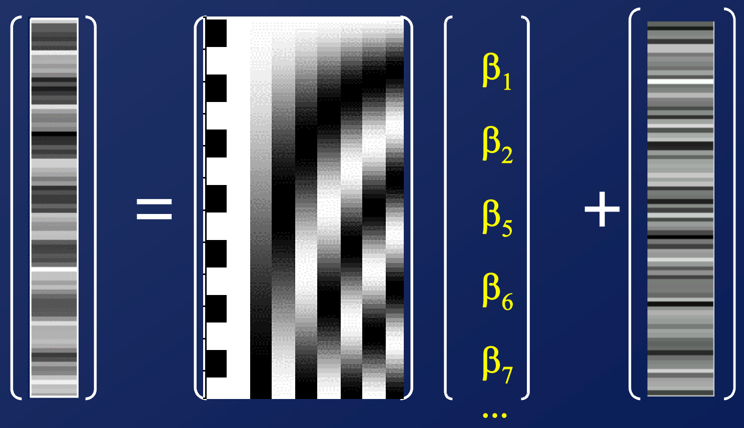
fitted drift



residuals

How do we find the betas estimates? By minimizing the residual variance

Fitting the model = finding some estimate of the betas



$$Y \;=\; X\,\beta \;+\; \varepsilon$$

finding the betas = minimising the sum of square of the residuals

$$\| \, Y - X \square \, \|^{2} = \Sigma_i \, [\, y_i - \square X \square \, \square_i \,]^{2}$$

when β are estimated: let's call them b

when ε is estimated : let's call it e

estimated SD of ε : let's call it s

Take home ...

- *We put in our model regressors (or covariates) that represent how we think the signal is varying (of interest and of no interest alike)*
    - ***WHICH ONE TO INCLUDE ?***
    - ***What if we have too many?***

- *Coefficients (= parameters) are estimated by minimizing the fluctuations, - variability – variance – of estimated noise – the residuals.*

- *Because the parameters depend on the scaling of the regressors included in the model, one should be careful in comparing manually entered regressors, or conditions of different durations*

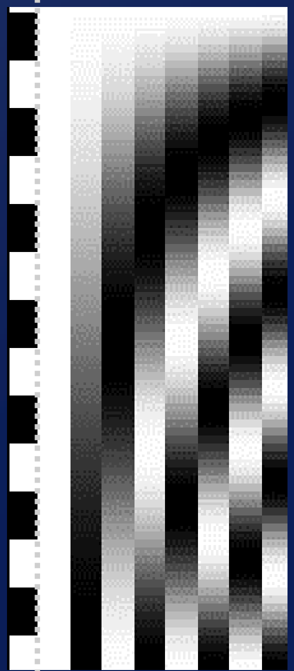
Plan

- *Make sure we all know about the estimation (fitting) part ....*
- *Make sure we understand t and F tests*
- *But what do we test exactly ?*
- *An example – almost real*

T test - one dimensional contrasts - SPM{*t*}

$c' = 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0$

$b_1\ b_2\ b_3\ b_4\ b_5$ ....



A contrast = a weighted sum of parameters: $c' \times b$

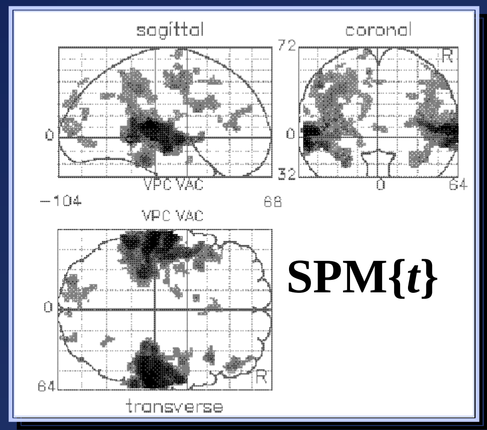
$b_1 > 0$?

Compute $1 \mathrm{x} b_1 + 0 \mathrm{x} b_2 + 0 \mathrm{x} b_3 + 0 \mathrm{x} b_4 + 0 \mathrm{x} b_5 + \ldots$

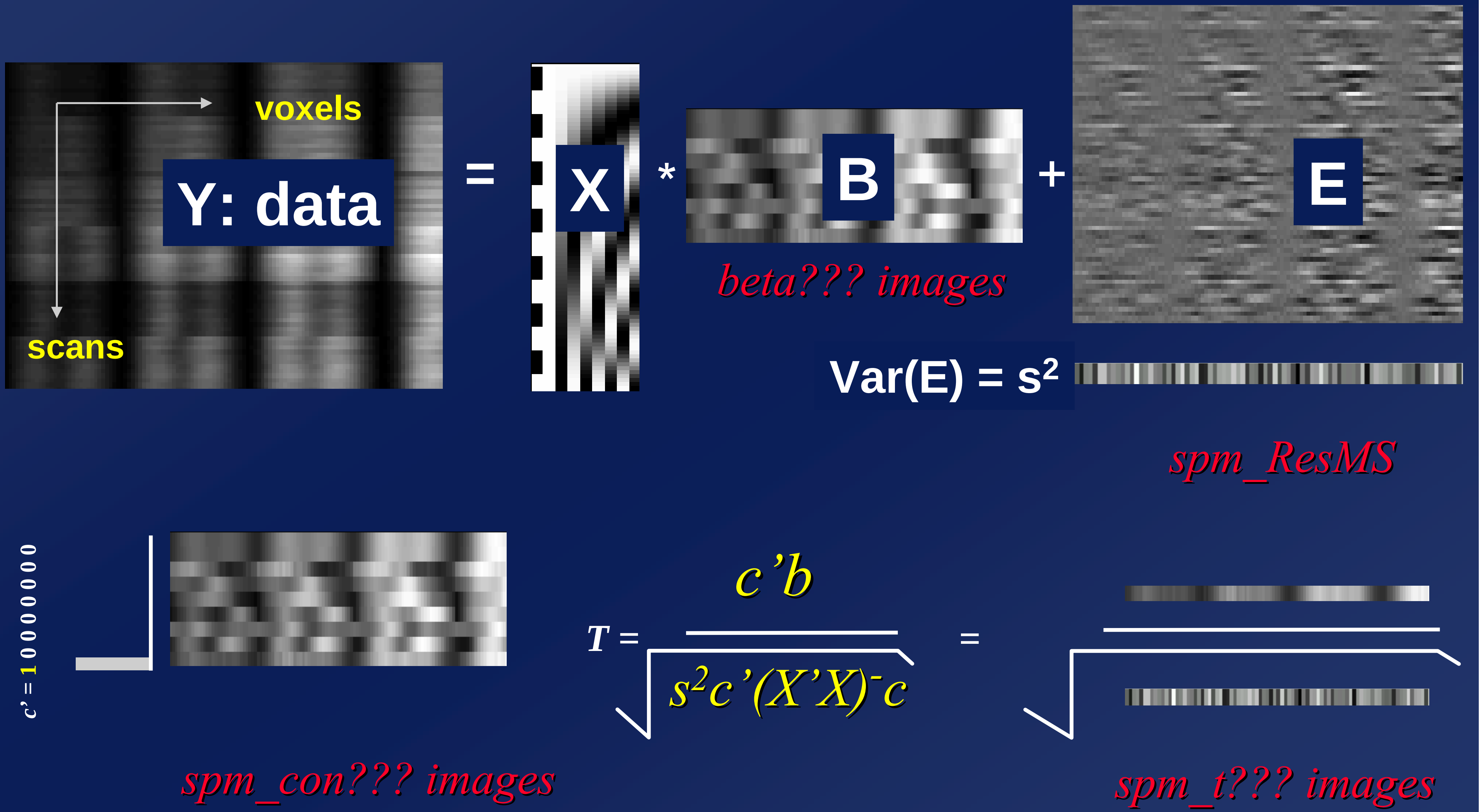
divide by estimated standard deviation of b_1

$$T = \frac{\textit{contrast} \text{ of estimated parameters}}{\sqrt{\text{variance estimate}}}$$

$$T = \frac{c'b}{\sqrt{s^2 c'(X'X)^- c}}$$



From one time series to an image



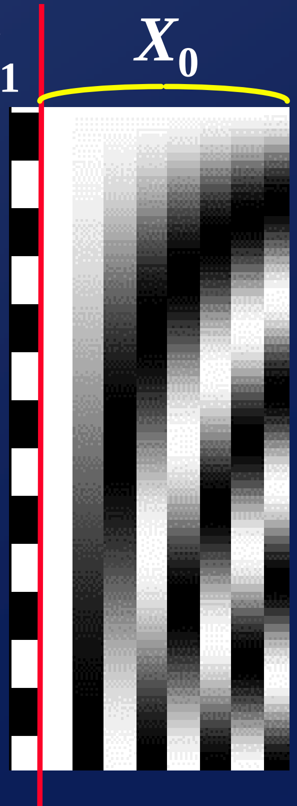
$$T = \frac{c'b}{\sqrt{s^2 c'(X'X)^{-}c}}$$

F-test : a reduced model

H_0: True model is X_0

H_0: $\beta_1 = 0$

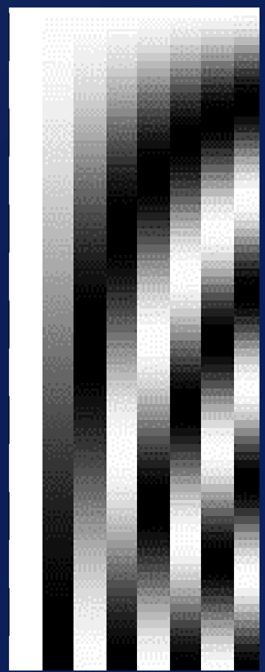
X_1 X_0



$\rightarrow S^2$

This (full) model ?

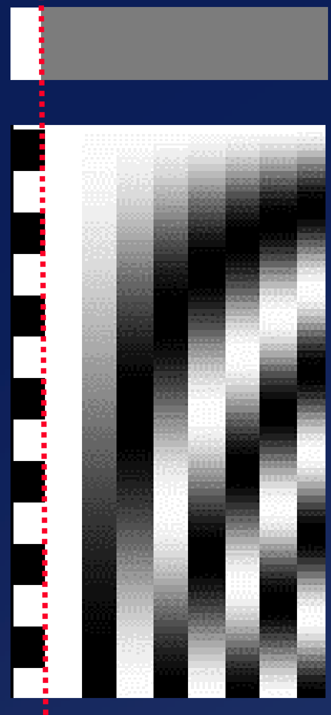
X_0



$\rightarrow S_0^2$

Or this one?

$c' = 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0$



$F \sim (S_0^2 - S^2) / S^2$

T values become F values. $F = T^2$

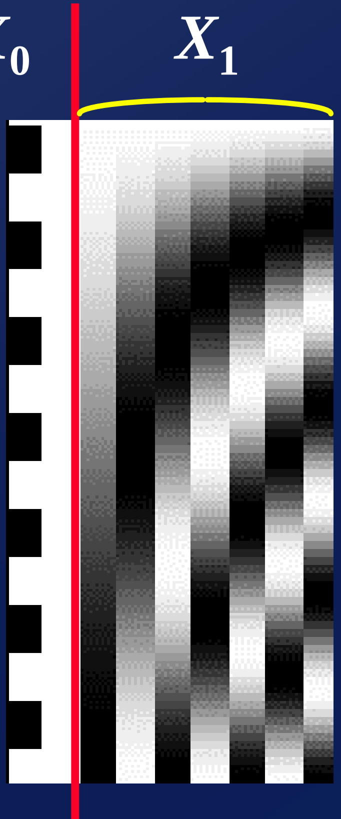
Both “activation” and “deactivations” are tested. Voxel wise p-values are halved.

F-test : a reduced model or ...

Tests multiple linear hypotheses : Does X1 model anything ?

H$_0$: True (reduced) model is X_0

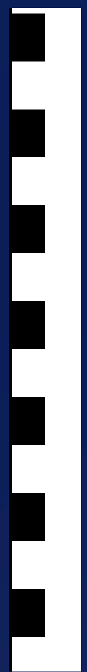
X_0 X_1



$\longrightarrow S^2$

This (full) model ?

X_0



$\longrightarrow S_0^2$

Or this one?

$$F = \frac{\text{additional variance accounted for by tested effects}}{\text{error variance estimate}}$$

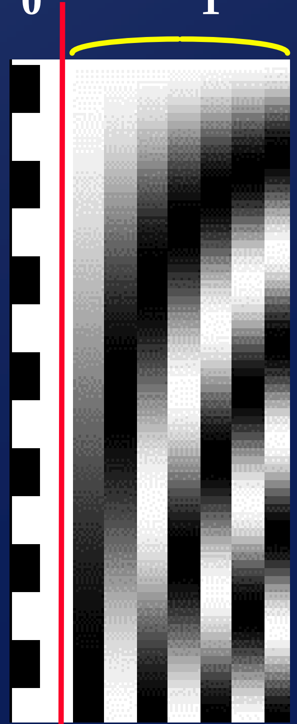
$$F \sim \left(S_0^2 - S^2 \right) / S^2$$

F-test : a reduced model or ... multi-dimensional contrasts ?

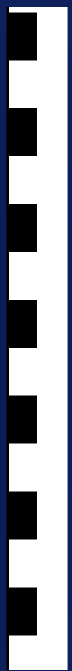
tests multiple linear hypotheses. Ex : does drift functions model anything?

H$_0$: True model is X_0

X_0 X_1



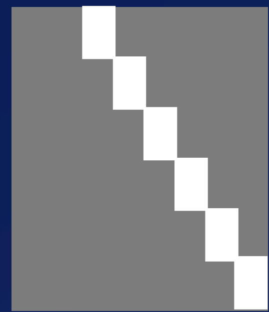
X_0

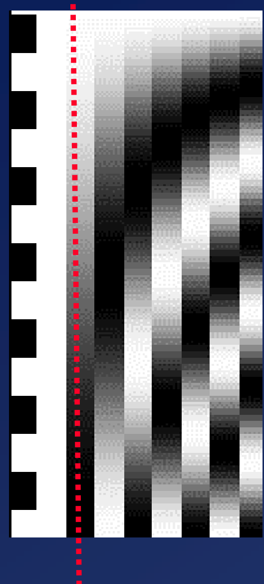


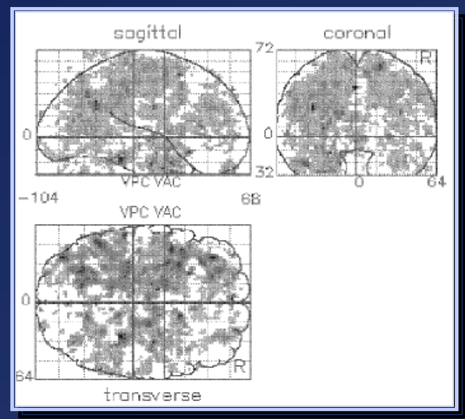
This (full) model ? Or this one?

H$_0$: β_{3-9} = (0 0 0 0 ...)

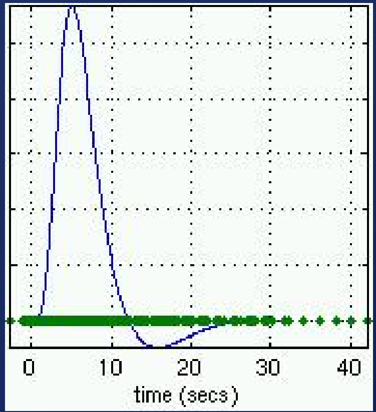
$$c' = \begin{matrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{matrix}$$

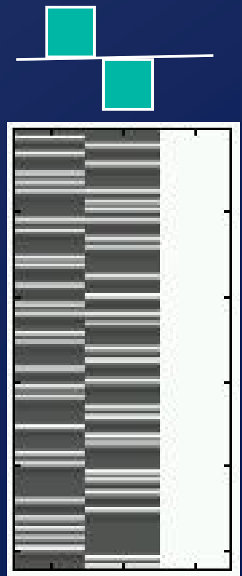


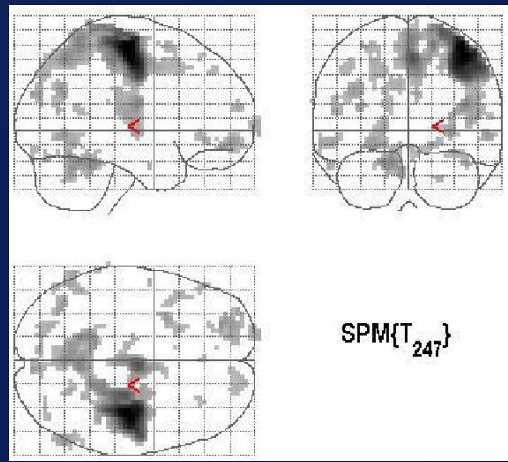


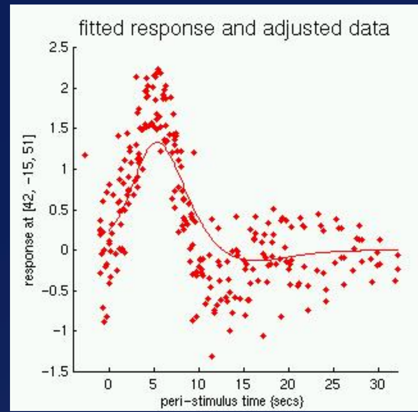


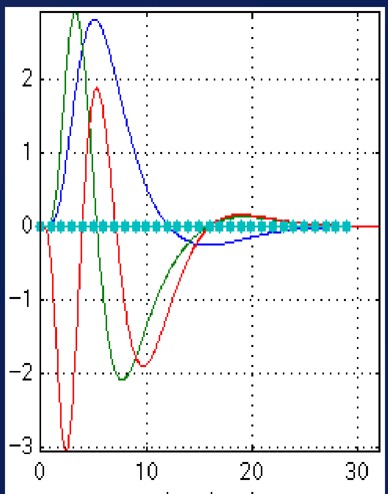
Convolution model	Design and contrast	SPM(t) or SPM(F)	Fitted and adjusted data

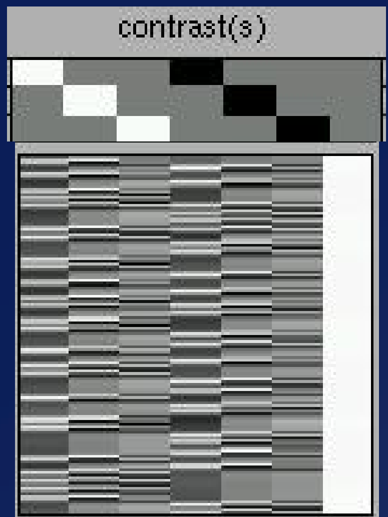


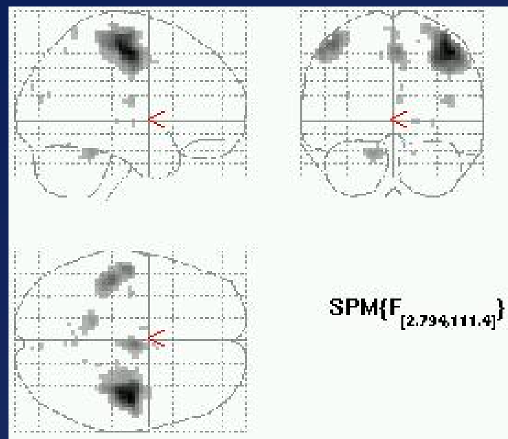


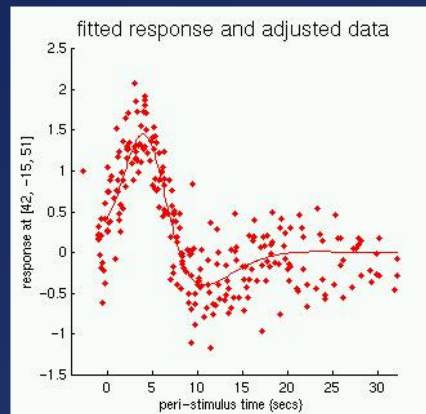












T and F test: take home ...

- *T tests are simple combinations of the betas; they are either positive or negative (b1 – b2 is different from b2 – b1)*

- *F tests can be viewed as testing for the additional variance explained by a larger model wrt a simpler model, or*

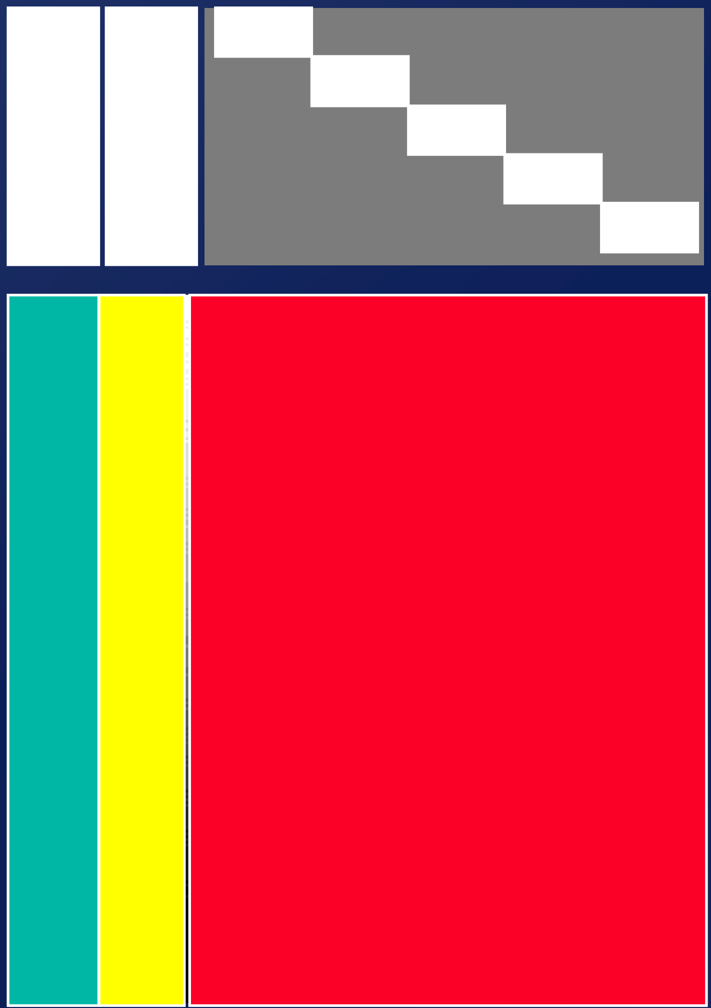
- *F tests the sum of the squares of one or several combinations of the betas*

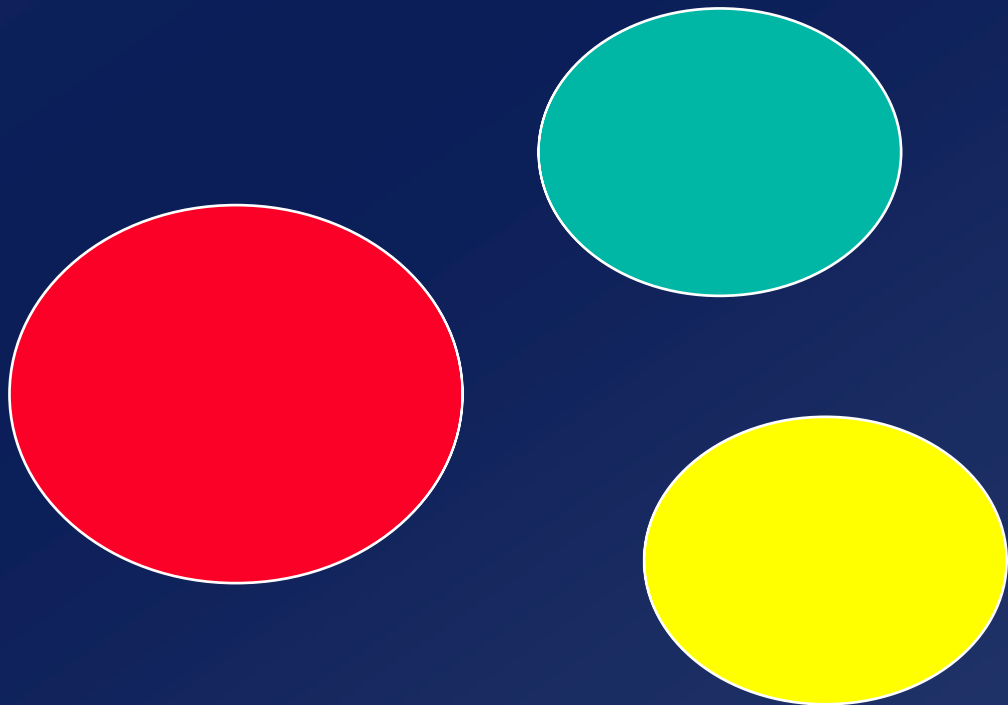
- *in testing “single contrast” with an F test, for ex. b1 – b2, the result will be the same as testing b2 – b1. It will be exactly the square of the t-test, testing for both positive and negative effects.*

Plan

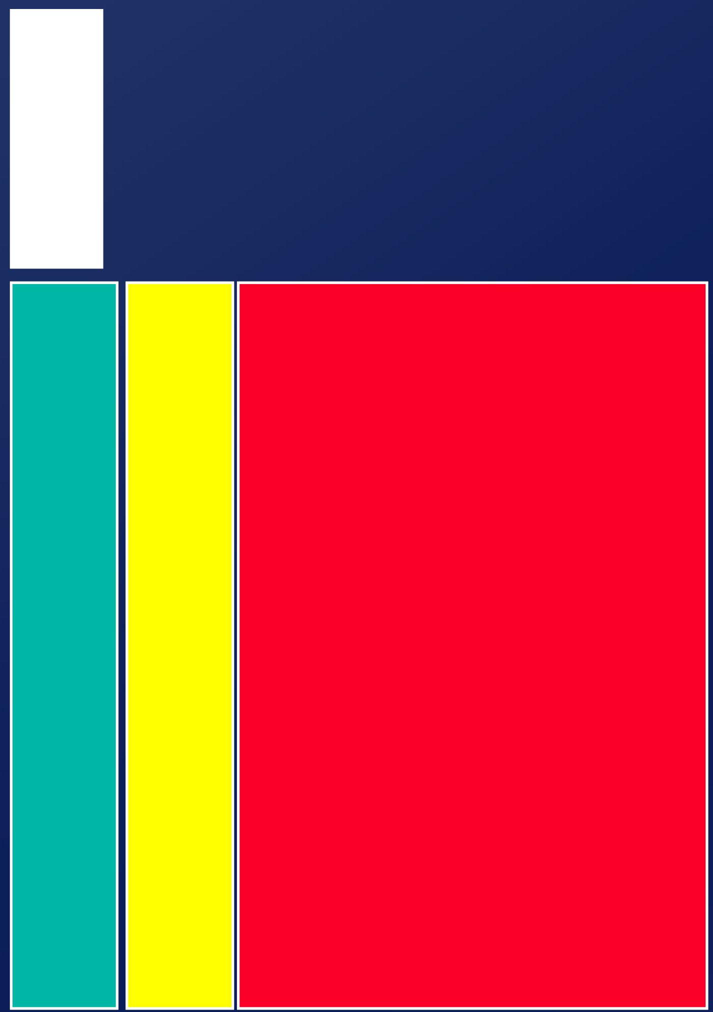
- *Make sure we all know about the estimation (fitting) part ....*
- *Make sure we understand t and F tests*
- *But what do we test exactly ? Correlation between regressors*
- *An example – almost real*

« Additional variance » : Again

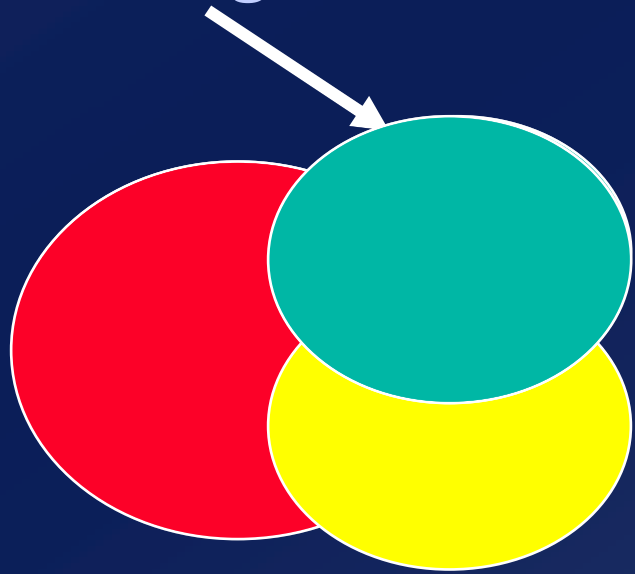




No correlation between green red and yellow



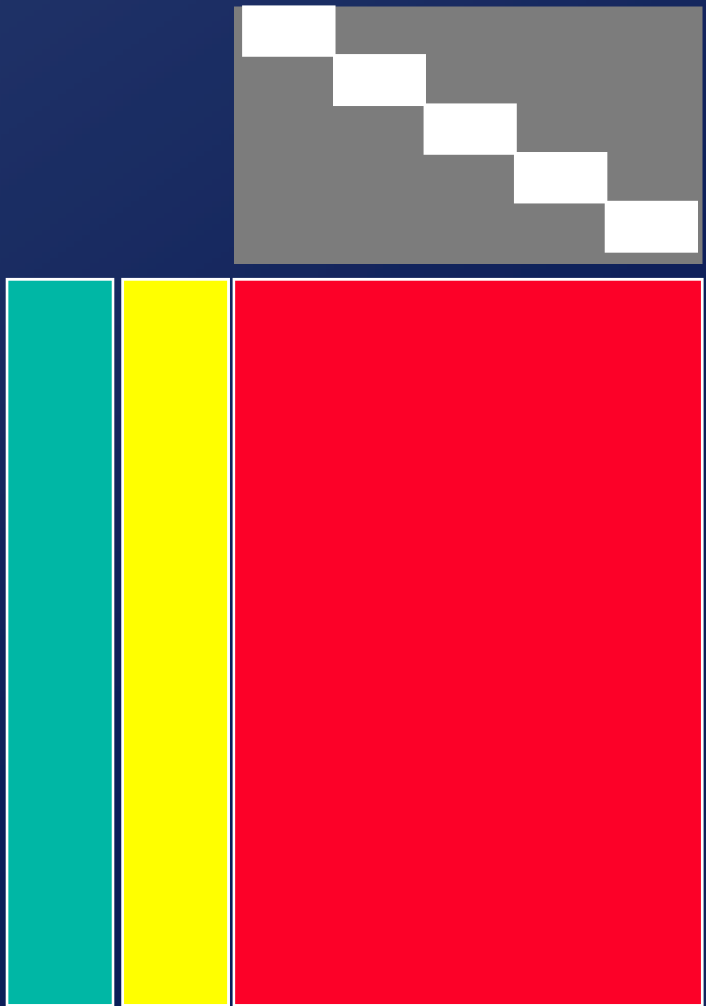
Testing for the green



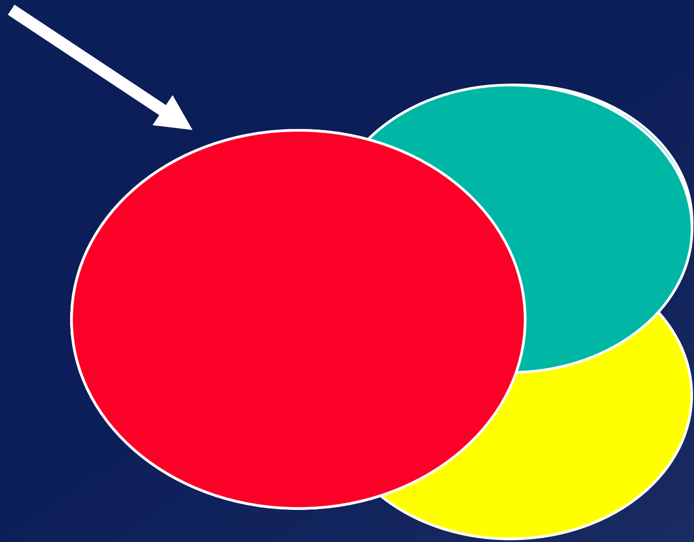
correlated regressors, for example

green: subject age

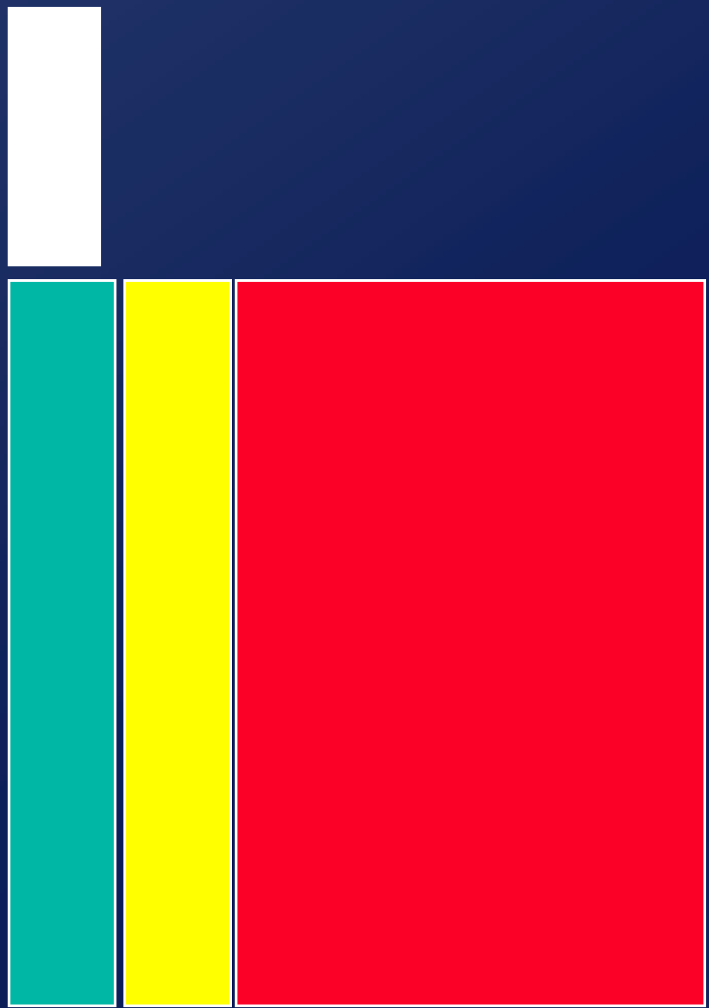
yellow: subject score



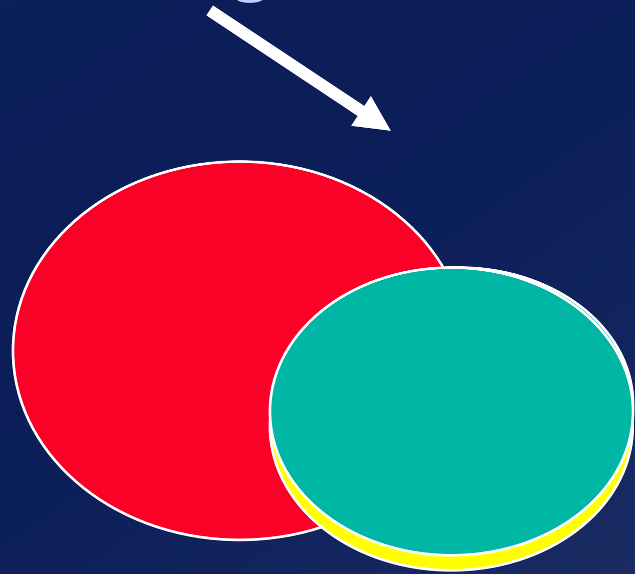
Testing for the red



correlated contrasts

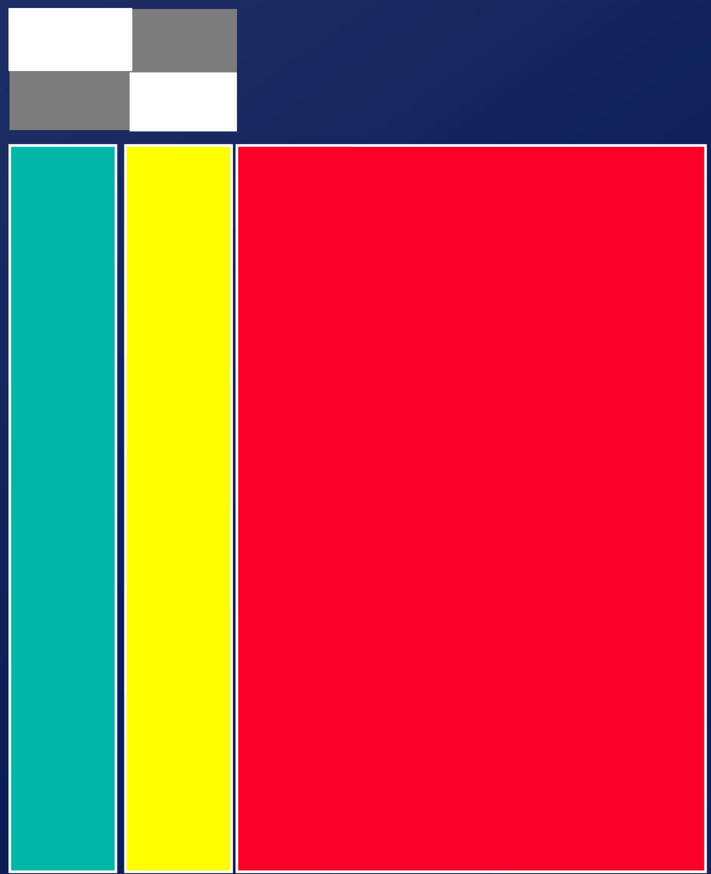


Testing for the green

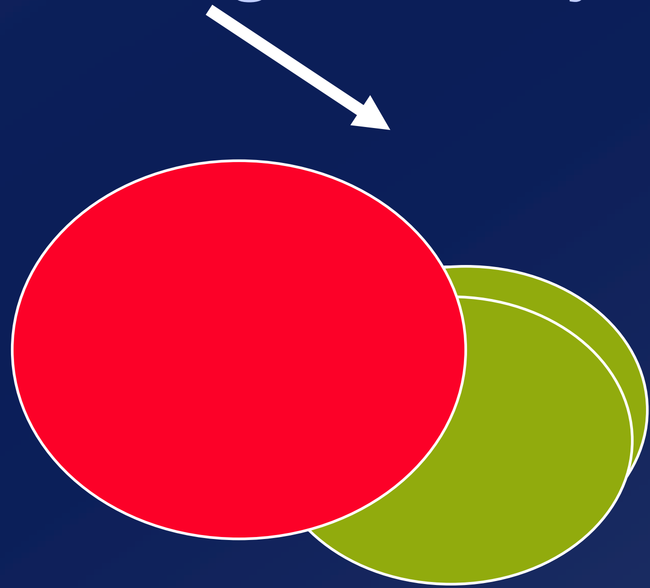


Very correlated regressors ?

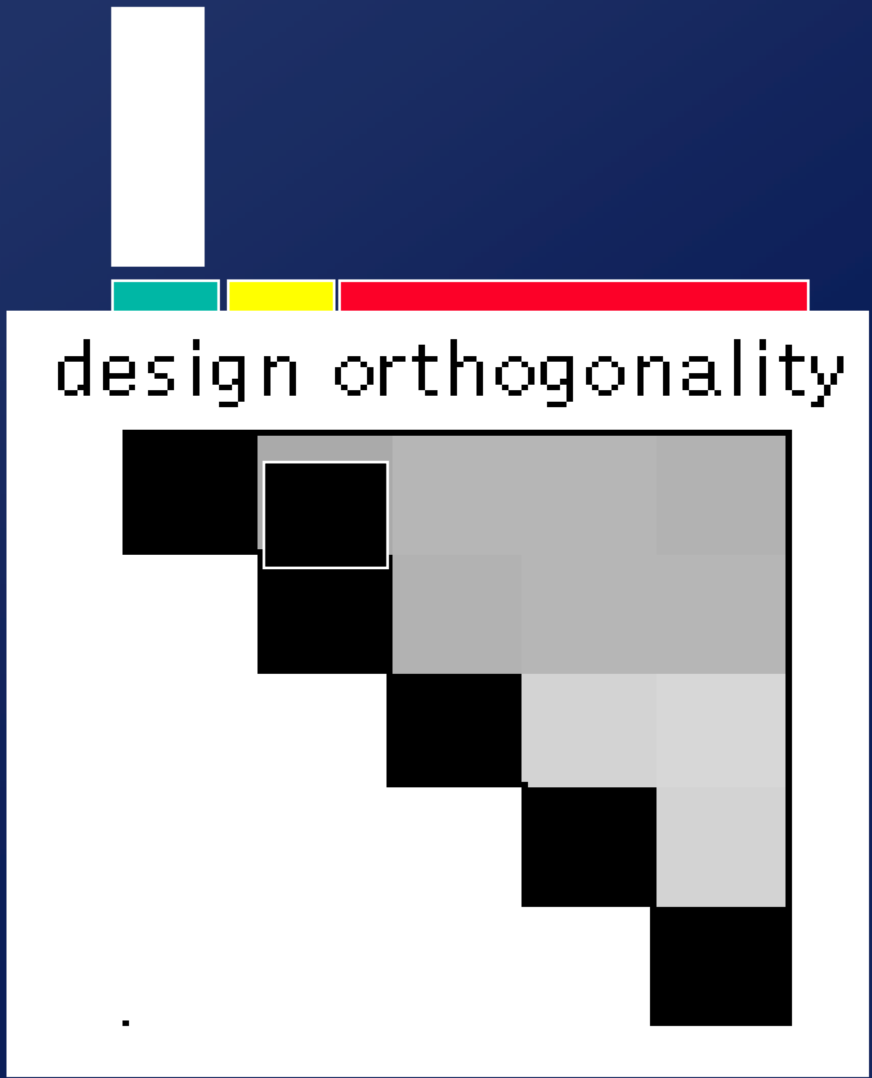
Dangerous !



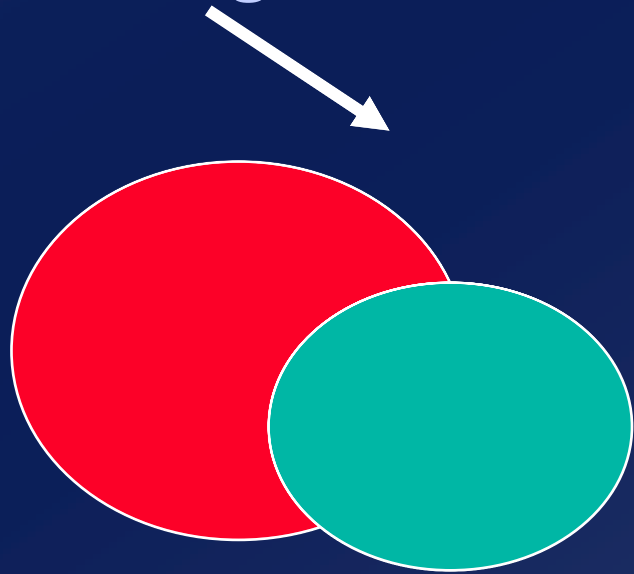
Testing for the green and yellow



If significant ? Could be G or Y !

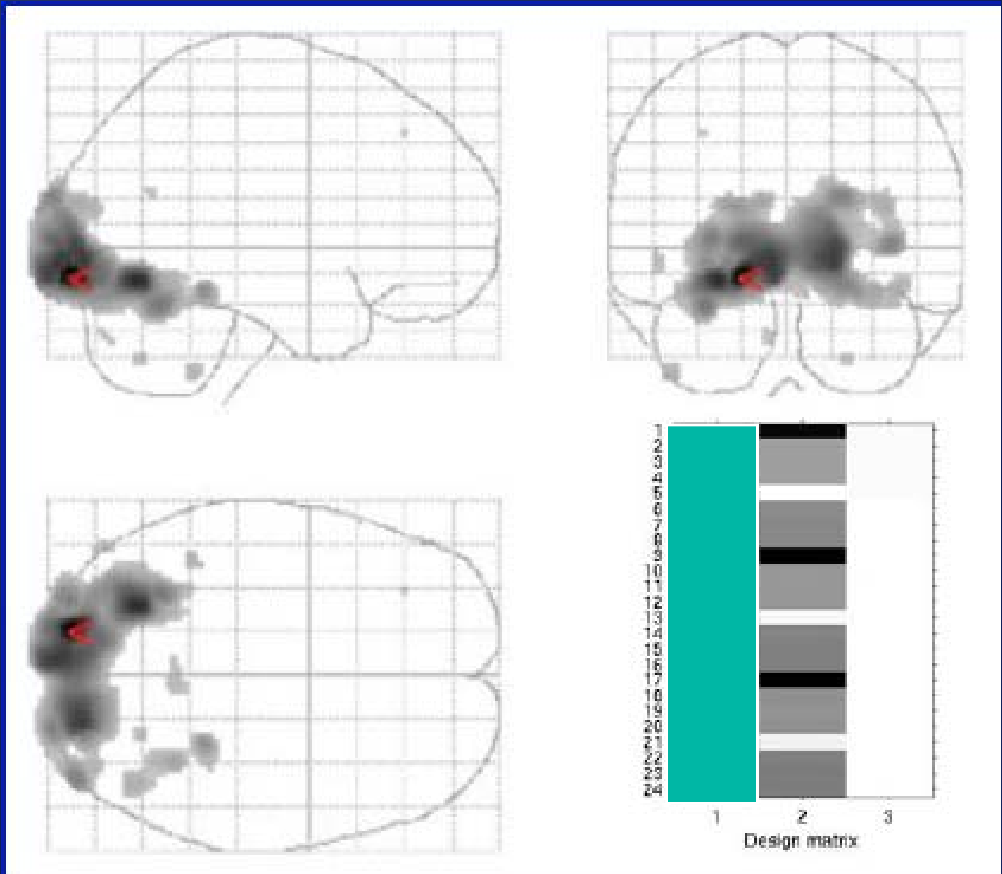


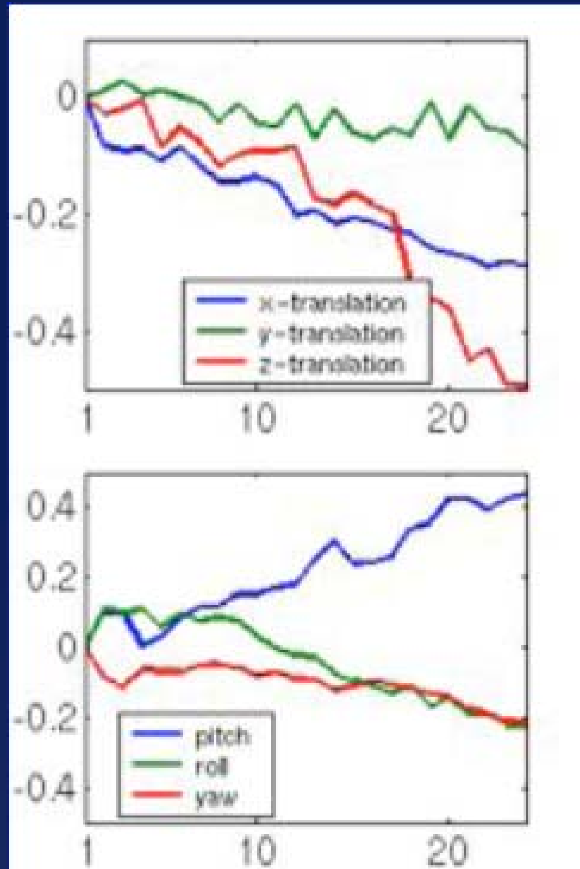
Testing for the green



Completely correlated regressors ?
Impossible to test ! (not estimable)

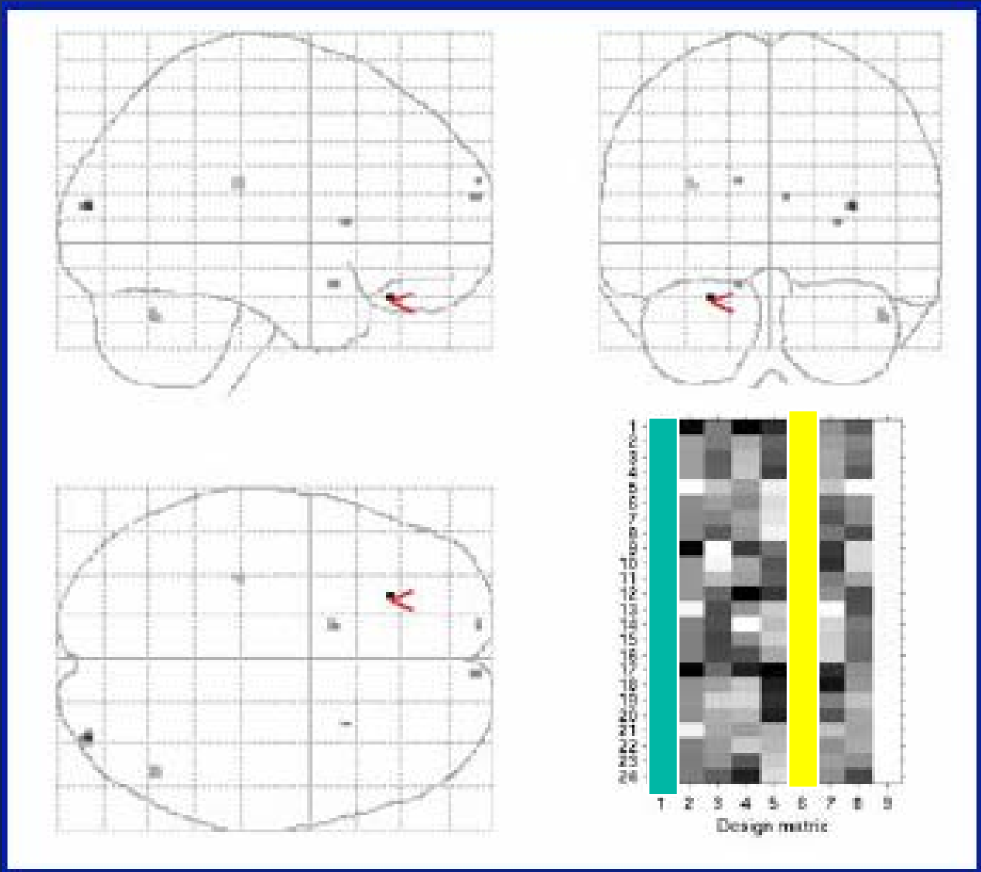
An example: real

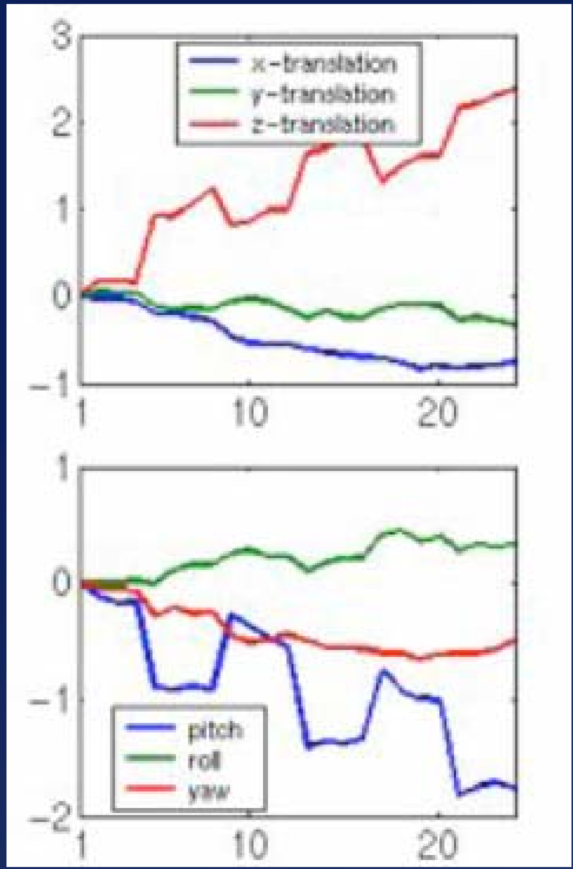




Testing for first regressor: T max = 9.8

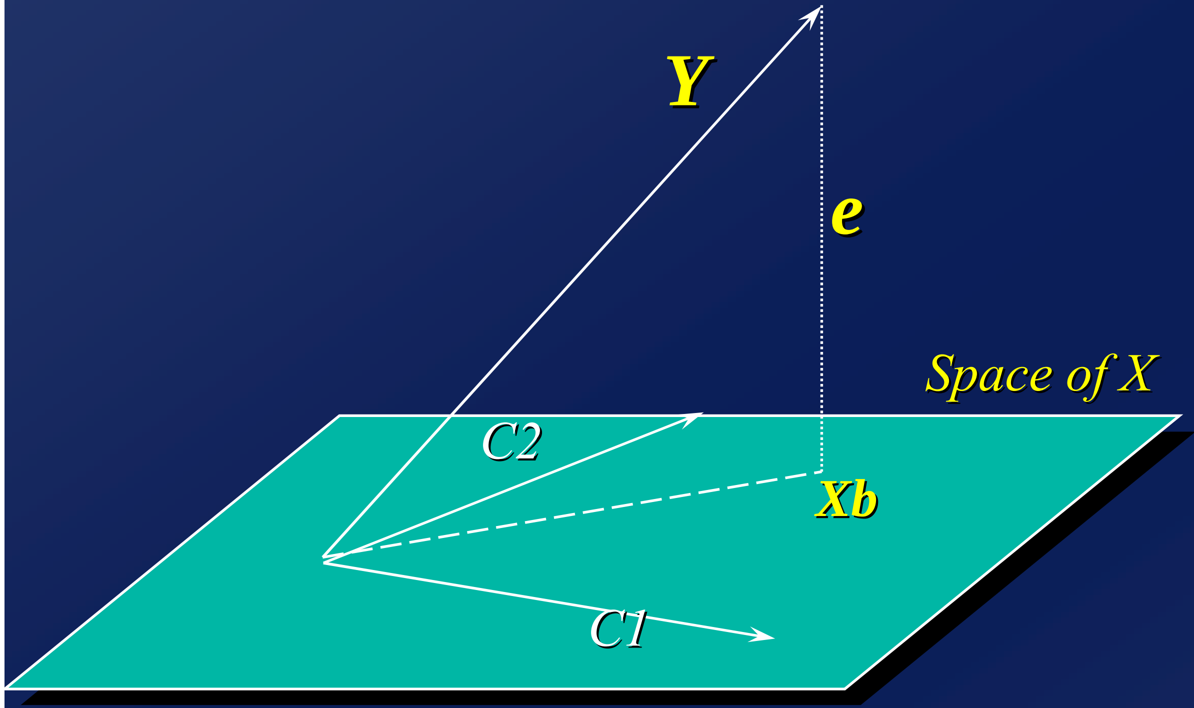
Including the movement parameters in the model

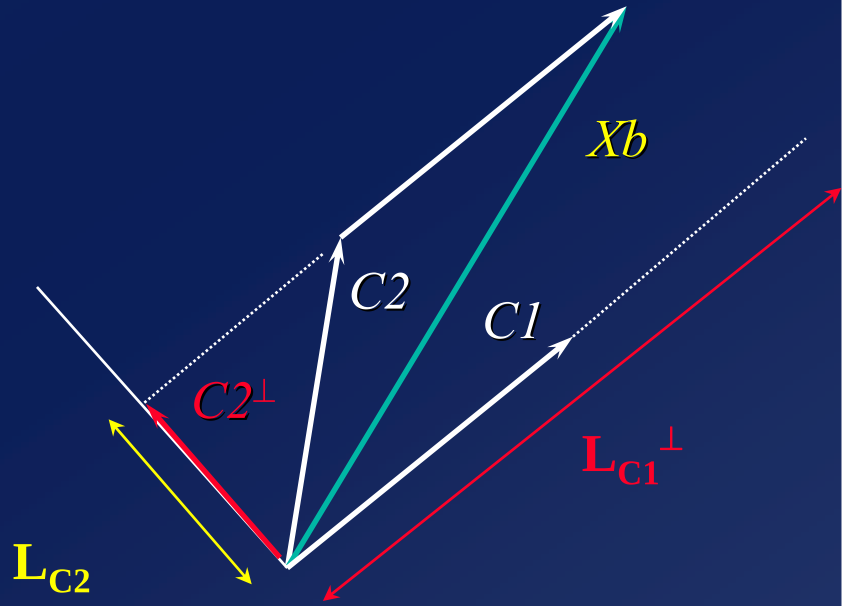




Testing for first regressor: activation is gone !

Implicit or explicit (⊥) decorrelation (or orthogonalisation)





This generalises when testing several regressors (F tests)

cf Andrade et al., NeuroImage, 1999

L_{C2} : test of C2 in the implicit ⊥ model

$L_{C1}^{\perp}$: test of C1 in the explicit ⊥ model

Correlation between regressors: take home ...

- *Do we care about correlation in the design ? Yes, always*

- *Start with the experimental design : conditions should be as uncorrelated as possible*

- *use F tests to test for the overall variance explained by several (correlated) regressors*

Plan

- *Make sure we all know about the estimation (fitting) part ....*
- *Make sure we understand t and F tests*
- *But what do we test exactly ? Correlation between regressors*
- *An example – almost real*

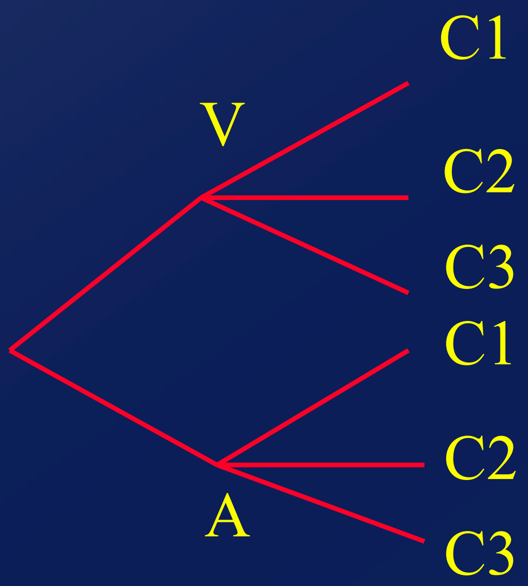
A real example (almost !)

Experimental Design → Design Matrix

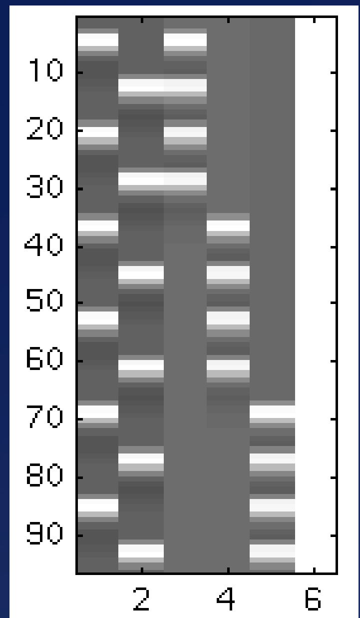
Factorial design with 2 factors : modality and category

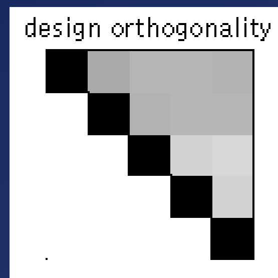
2 levels for modality (eg Visual/Auditory)

3 levels for category (eg 3 categories of words)

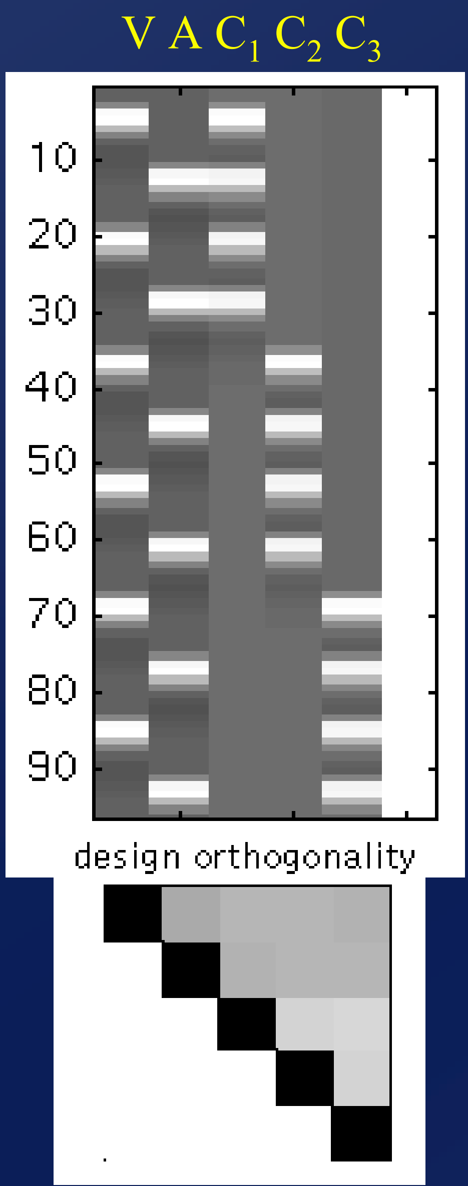


V A C_1 C_2 C_3





Asking ourselves some questions ...



Test C1 > C2 : c = [0 0 1 -1 0 0]

Test V > A : c = [1 -1 0 0 0 0]

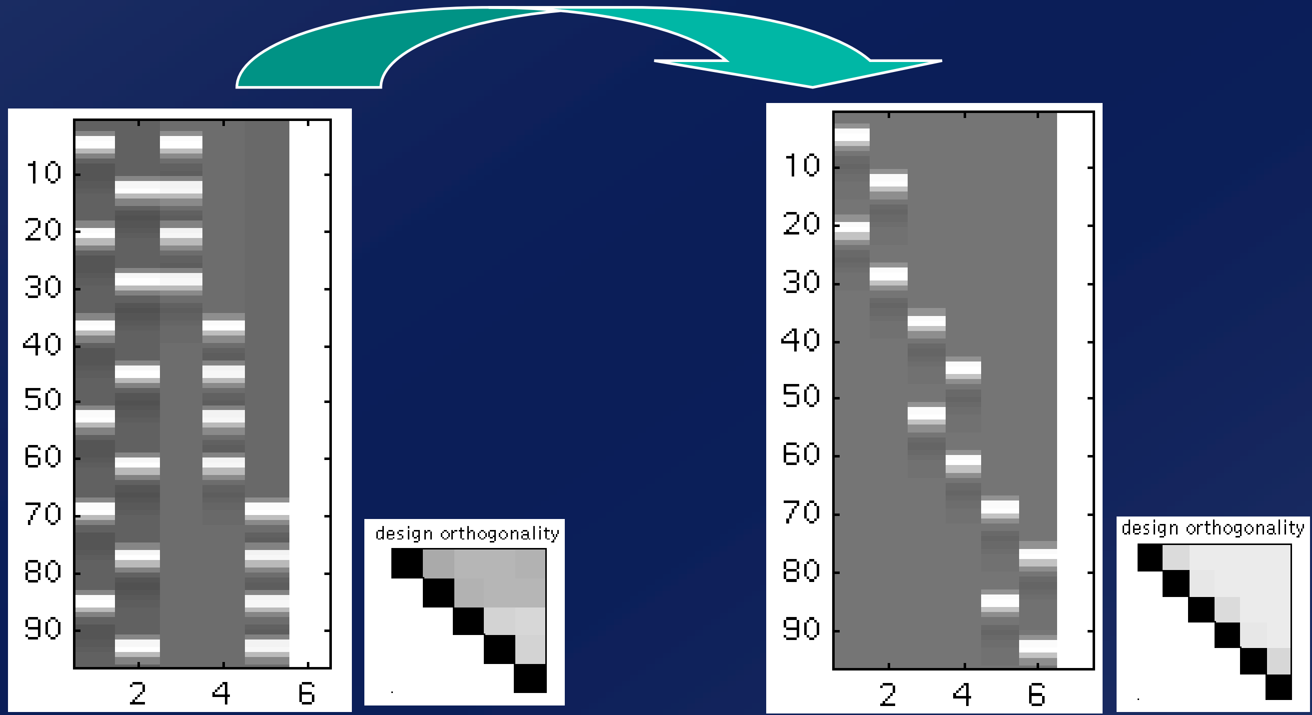
Test C1,C2,C3 ? (F)

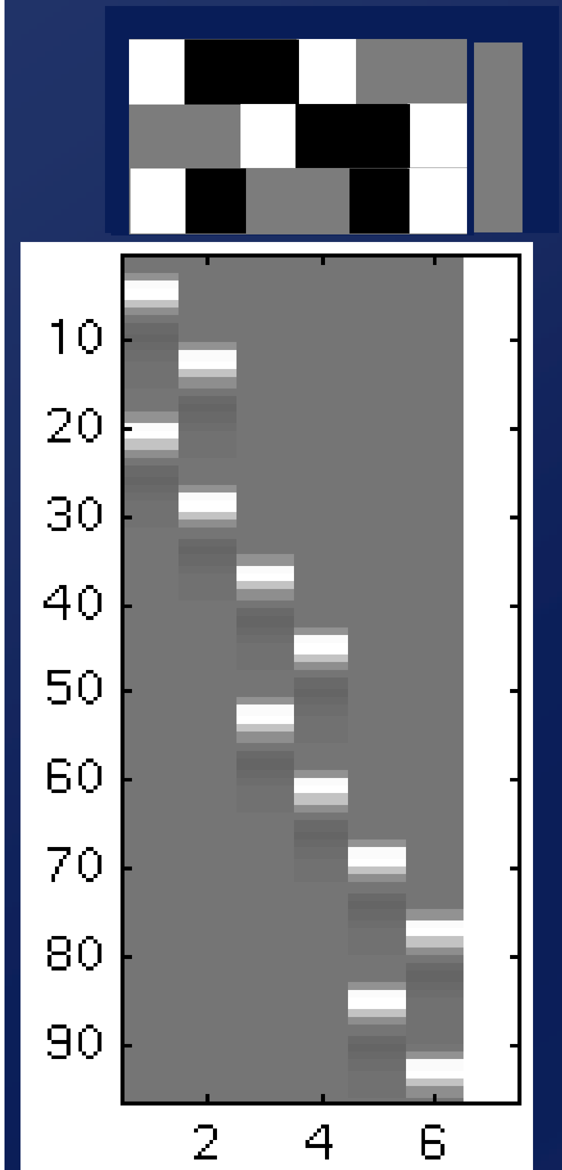
$$c = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Test the interaction MxC ?

- Design Matrix not orthogonal
- Many contrasts are non estimable
- Interactions MxC are not modelled

Modelling the interactions





Test C1 > C2 : c = [1 1 -1 -1 0 0 0]

Test V > A : c = [1 -1 1 -1 1 -1 0]

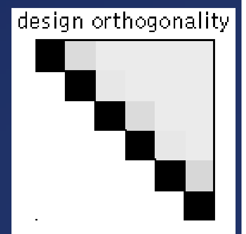
Test the category effect :

$$c = \begin{bmatrix} 1 & 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & -1 & -1 & 0 \\ 1 & 1 & 0 & 0 & -1 & -1 & 0 \end{bmatrix}$$

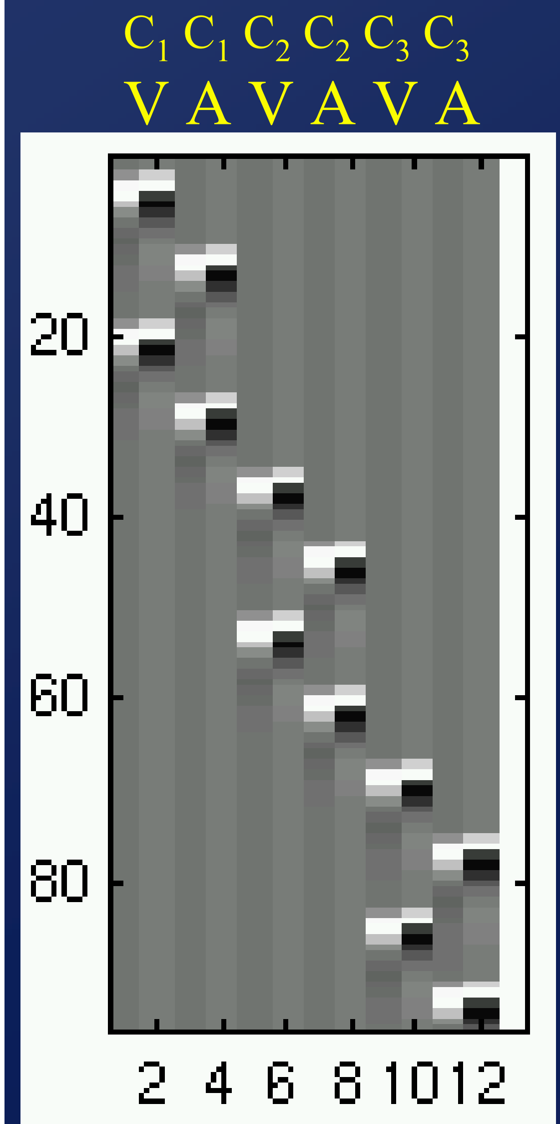
Test the interaction MxC :

$$c = \begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 1 & 0 \\ 1 & -1 & 0 & 0 & -1 & 1 & 0 \end{bmatrix}$$

- Design Matrix orthogonal
- All contrasts are estimable
- Interactions MxC modelled
- If no interaction ... ? Model is too “big” !



With a more flexible model



Test C1 > C2 ?
Test C1 different from C2 ?
from

c = [1 1 -1 -1 0 0 0]

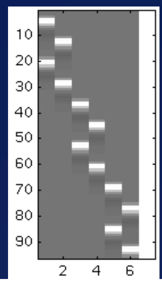
to

c = [1 0 1 0 -1 0 -1 0 0 0 0 0 0]
[0 1 0 1 0 -1 0 -1 0 0 0 0 0]

becomes an F test!

What if we use only:

c = [1 0 1 0 -1 0 -1 0 0 0 0 0 0]



OK only if the regressors coding for the delay are all equal

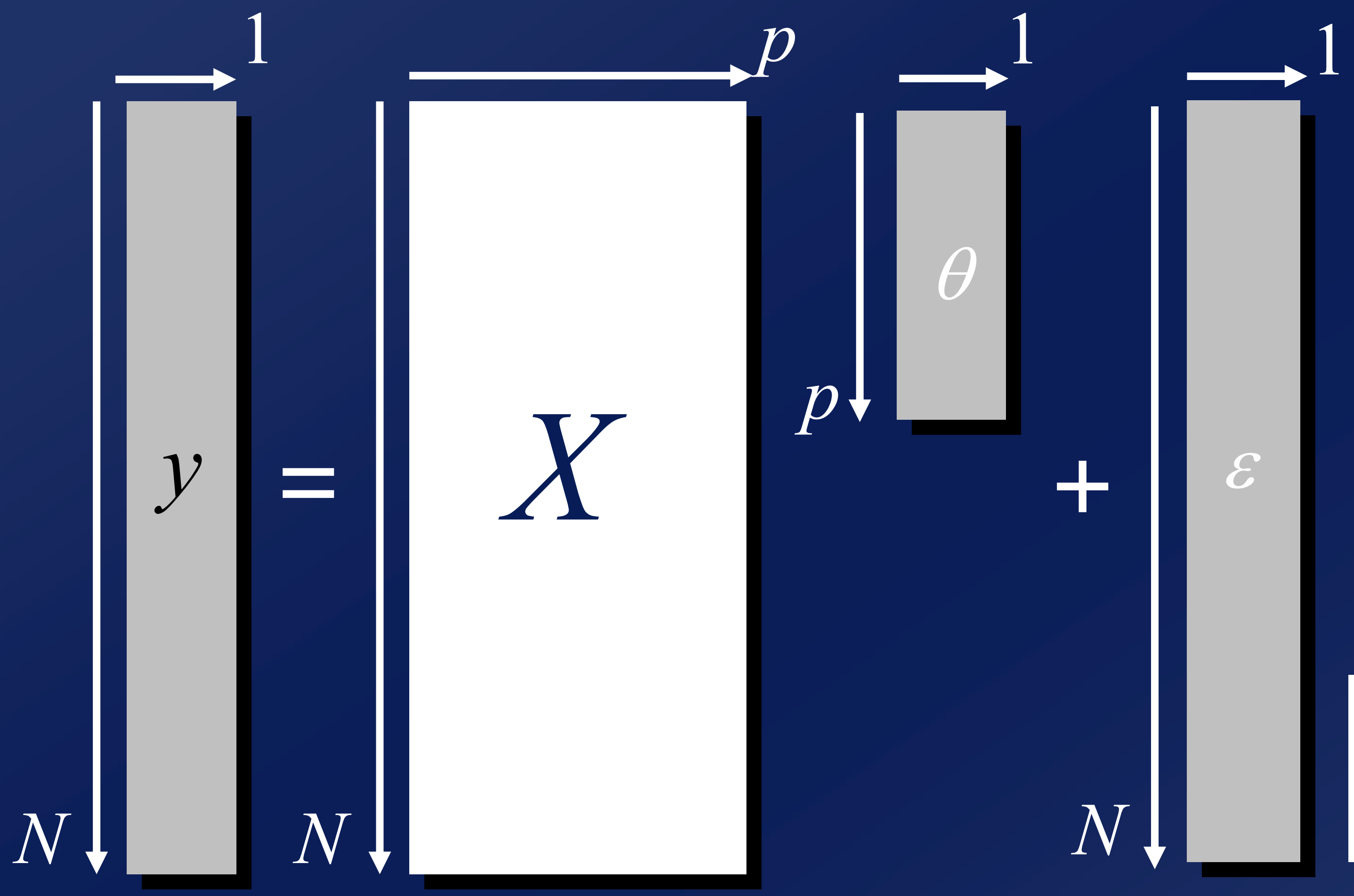
Toy example: take home ...

- *use F tests when*
  - Test for >0 and <0 effects
  - Test for more than 2 levels in factorial designs
  - Conditions are modelled with more than one regressor

- *Check post hoc*

General Linear Model

$$y = X\theta + \varepsilon$$



Error Covariance

$$C_\varepsilon = \sum_k \lambda_k Q_k$$

N: number of scans
p: number of regressors

Model is specified by
1. Design matrix X
2. Assumptions about ε

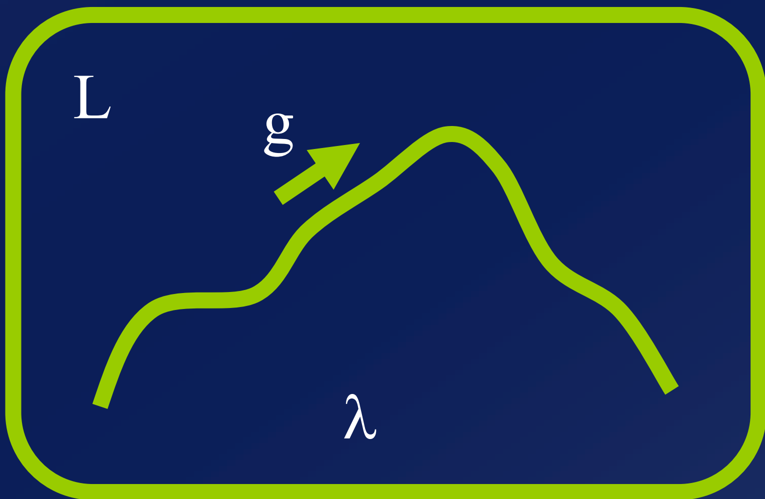
Estimation

$$\underset{N\times 1}{y} = \underset{N\times p}{X} \; \underset{p\times 1}{\theta} + \underset{N\times 1}{\varepsilon}$$

1. ReML-algorithm

$$C_\varepsilon = \sum_k \lambda_k Q_k$$

Maximise $\quad L = \ln p(y \mid \lambda) = \ln \int p(y \mid \theta, \lambda) d\theta$



$$g = \frac{dL}{d\lambda}$$

$$J = \frac{d^2 L}{d\lambda^2}$$

$$\lambda = \lambda + J^{-1} g$$

2. Weighted Least Squares

$$\theta = (X^T C_e^{-1} X^T) X^T C_e^{-1} y$$

Friston et al. 2002, Neuroimage

Hierarchical model

Hierarchical model

$$
\begin{aligned}
y &= X^{(1)}\theta^{(1)} + \varepsilon^{(1)} \\
\theta^{(1)} &= X^{(2)}\theta^{(2)} + \varepsilon^{(2)} \\
&\vdots \\
\theta^{(n-1)} &= X^{(n)}\theta^{(n)} + \varepsilon^{(n)}
\end{aligned}
$$

Multiple variance components at each level

$$C_{\varepsilon}^{(i)} = \sum_k \lambda_k^{(i)} Q_k^{(i)}$$

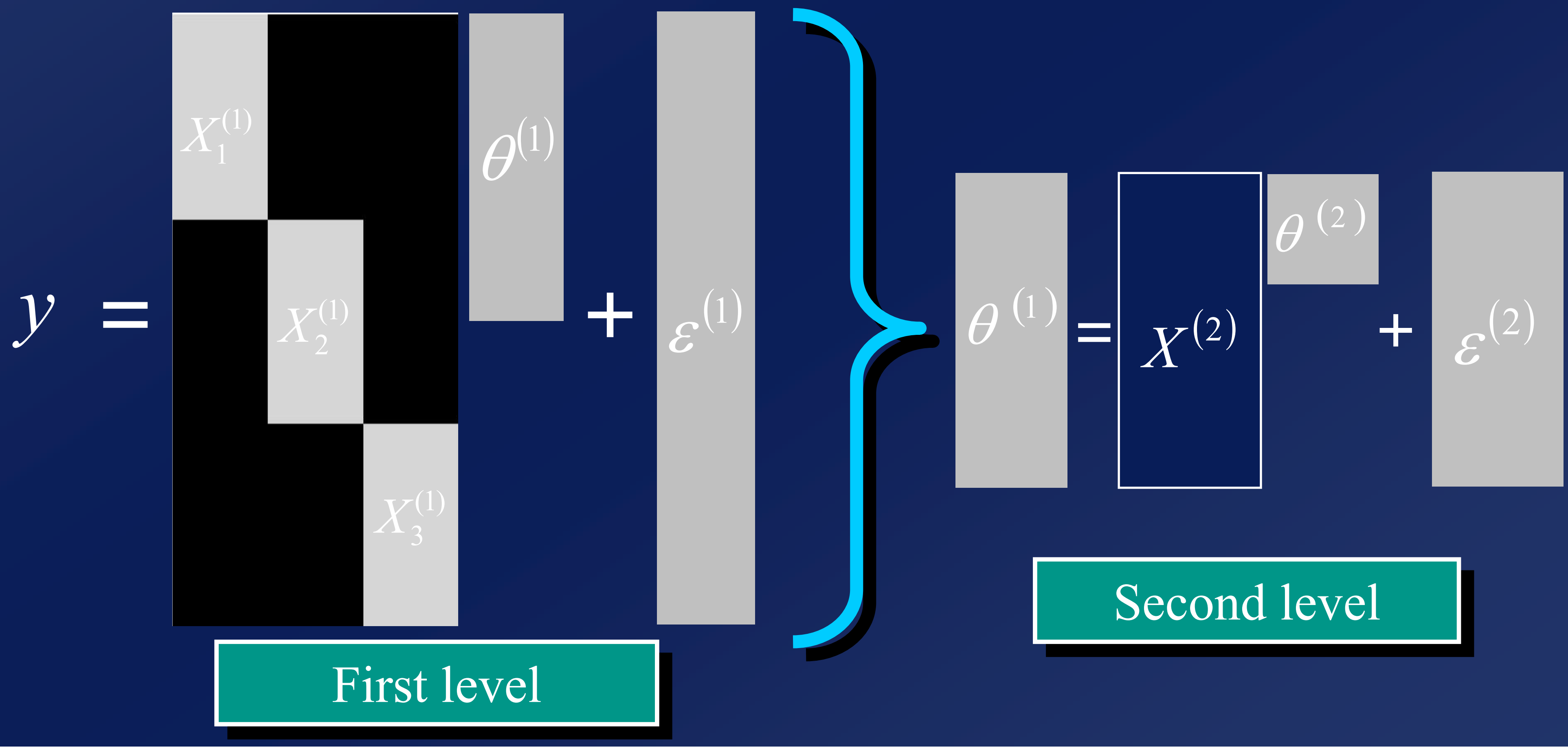
At each level, distribution of parameters is given by level above.

What we don't know: distribution of parameters and variance parameters.

Example: Two level model

$$y = X^{(1)}\theta^{(1)} + \varepsilon^{(1)}$$

$$\theta^{(1)} = X^{(2)}\theta^{(2)} + \varepsilon^{(2)}$$



Estimation

Hierarchical model

$$y = X^{(1)}\theta^{(1)} + \varepsilon^{(1)}$$
$$\theta^{(1)} = X^{(2)}\theta^{(2)} + \varepsilon^{(2)}$$
$$\vdots$$
$$\theta^{(n-1)} = X^{(n)}\theta^{(n)} + \varepsilon^{(n)}$$

Single-level model

$$\begin{aligned} y \quad = \quad & \varepsilon^{(1)} + X^{(1)}\varepsilon^{(2)} + \\ & \ldots + \\ & X^{(1)} \ldots X^{(n-1)}\varepsilon^{(n)} + \\ & X^{(1)} \ldots X^{(n)}\theta^{(n)} \\ & = X\theta + e \end{aligned}$$

Group analysis in practice

Many 2-level models are just too big to compute.

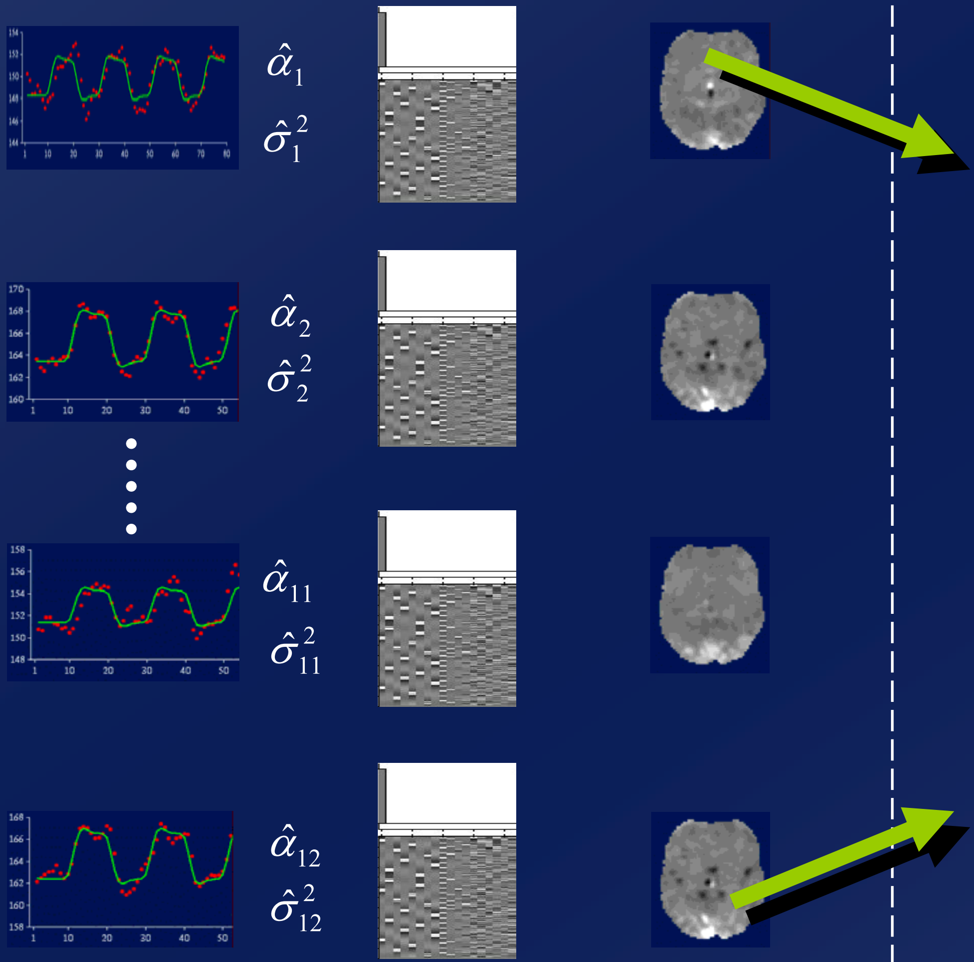
And even if, it takes a long time!

Is there a fast approximation?

Summary Statistics approach

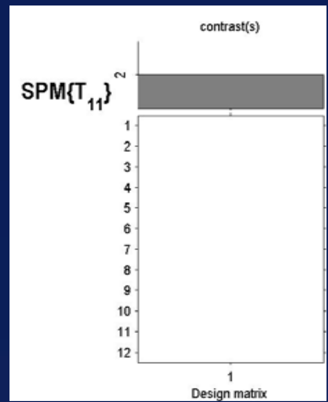
First level

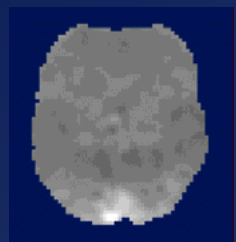
Data | Design Matrix | Contrast Images



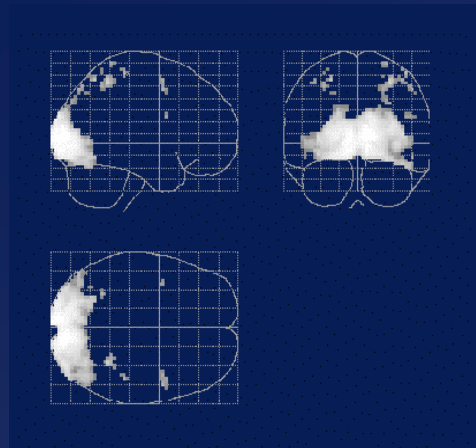
Second level

$$t = \frac{c^T \hat{\alpha}}{\sqrt{V\hat{a}r(c^T \hat{\alpha})}}$$





SPM(t)



One-sample t-test @ 2nd level

Validity of approach

The summary stats approach is exact if for each session/subject:

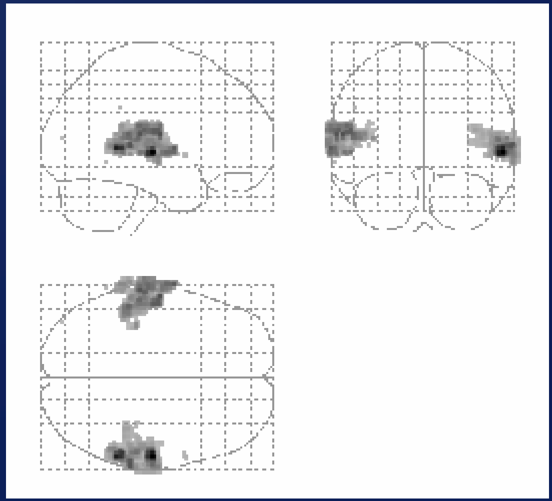
Within-session covariance the same

First-level design the same

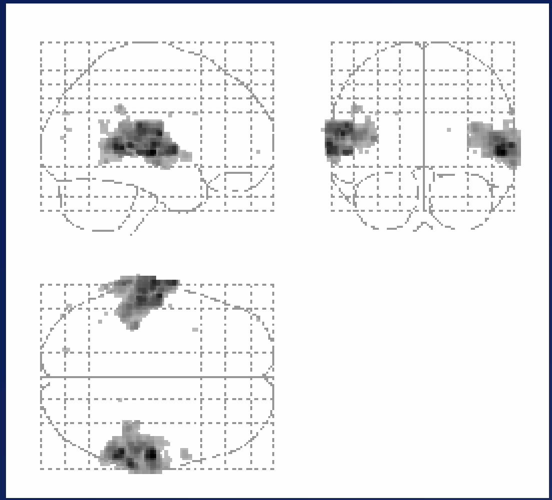
All other cases: Summary stats approach seems to be robust against typical violations.

Auditory Data

Summary statistics



Hierarchical Model



Friston et al. (2004) Mixed effects and fMRI studies, Neuroimage

Multiple contrasts per subject

Stimuli: Auditory Presentation (SOA = 4 secs) of words

Motion	Sound	Visual	Action
"jump"	"click"	"pink"	"turn"

Subjects: (i) 12 control subjects

Scanning: fMRI, 250 scans per subject, block design

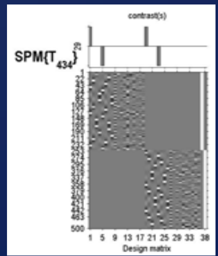
Question: What regions are affected by the semantic content of the words?

U. Noppeney et al.

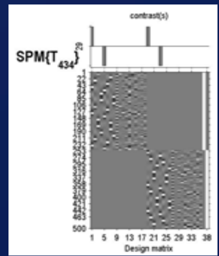
ANOVA

1st level:

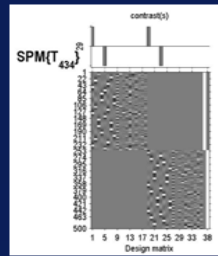
1.Motion | 2.Sound | 3.Visual | 4.Action



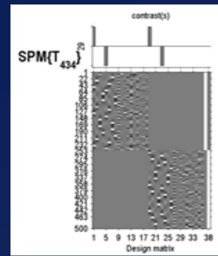
$\overset{?}{=}$



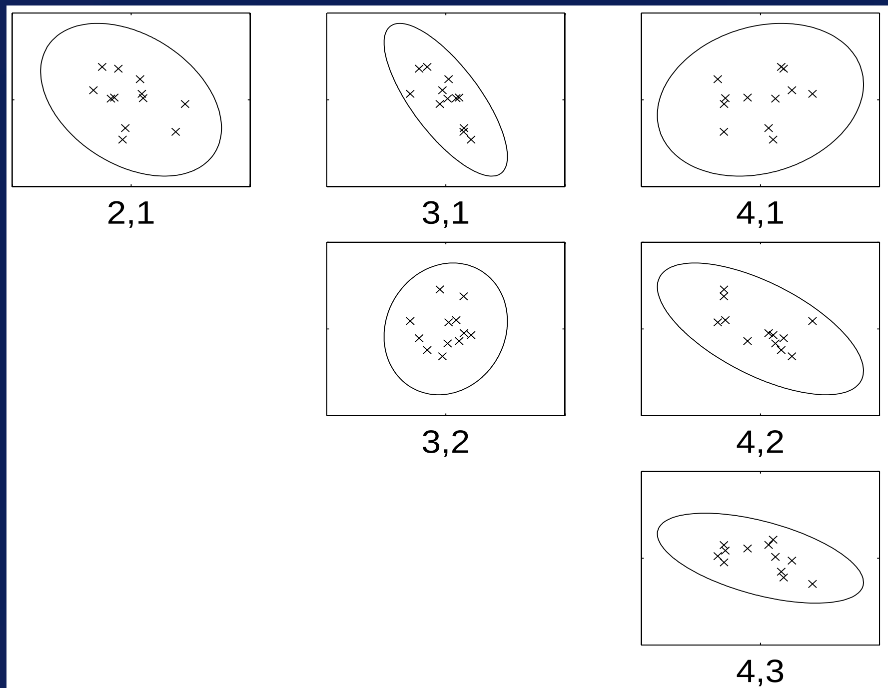
$\overset{?}{=}$



$\overset{?}{=}$



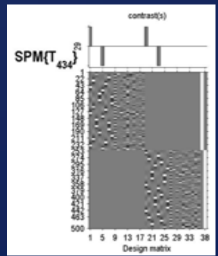
2nd level:

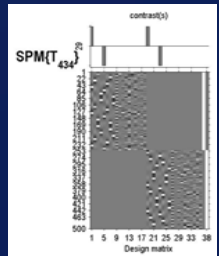


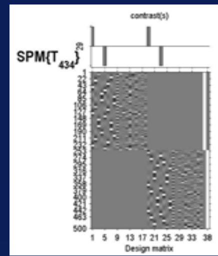
ANOVA

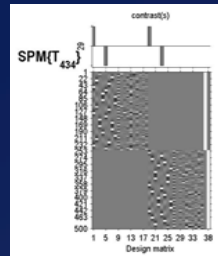
1st level:

Motion $\overset{?}{=}$ Sound $\overset{?}{=}$ Visual $\overset{?}{=}$ Action

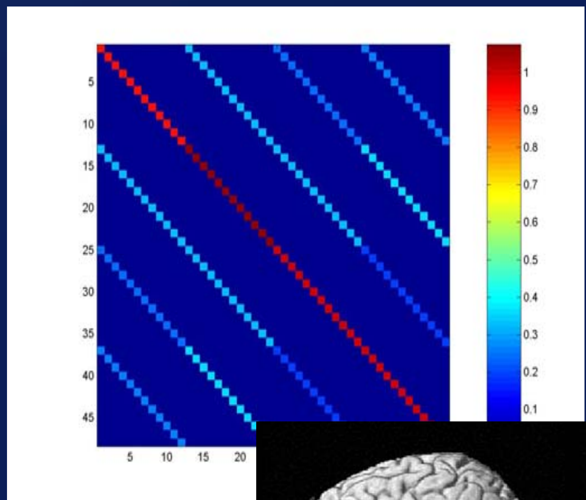




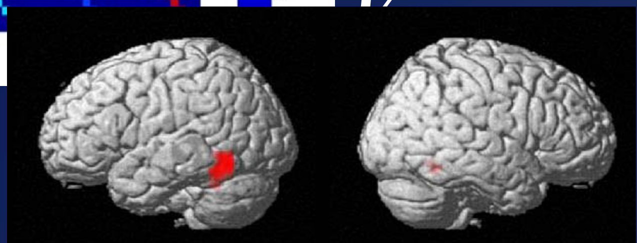


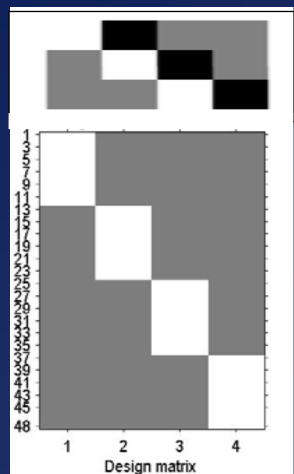


2nd level:



V





X

$$c^T = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

Summary

Linear hierarchical models are general enough for typical multi-subject imaging data (PET, fMRI, EEG/MEG).

Summary statistics are robust approximation for group analysis.

Also accomodates multiple contrasts per subject.

Thank you for your attention!

jbpoline@cea.fr