ETH zürich

University of Zurich^UZH

Translational
Neuromodeling
Unit

Classical Statistical Inference

Andreea O. Diaconescu

SPM Course 2016

With many thanks to Jakob Heinzle, Lars Kasper,
Jean-Baptiste Poline, Frederike Petzschner & Klaas E. Stephan

Overview

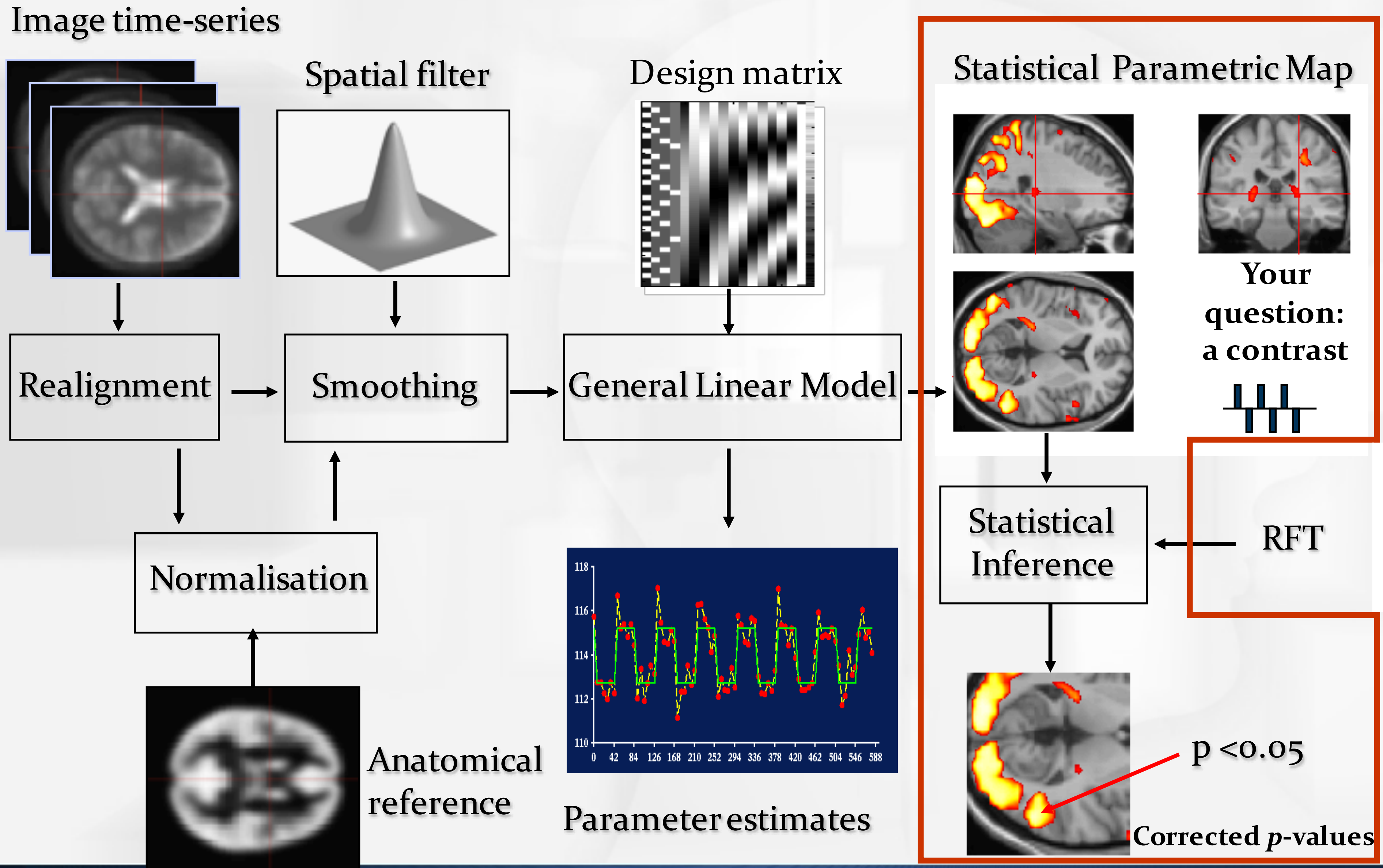


Image time-series

Spatial filter

Design matrix

Statistical Parametric Map

Realignment

Smoothing

General Linear Model

Your question: a contrast

Normalisation

Statistical Inference

RFT

Anatomical reference

Parameter estimates

p <0.05

Corrected *p*-values

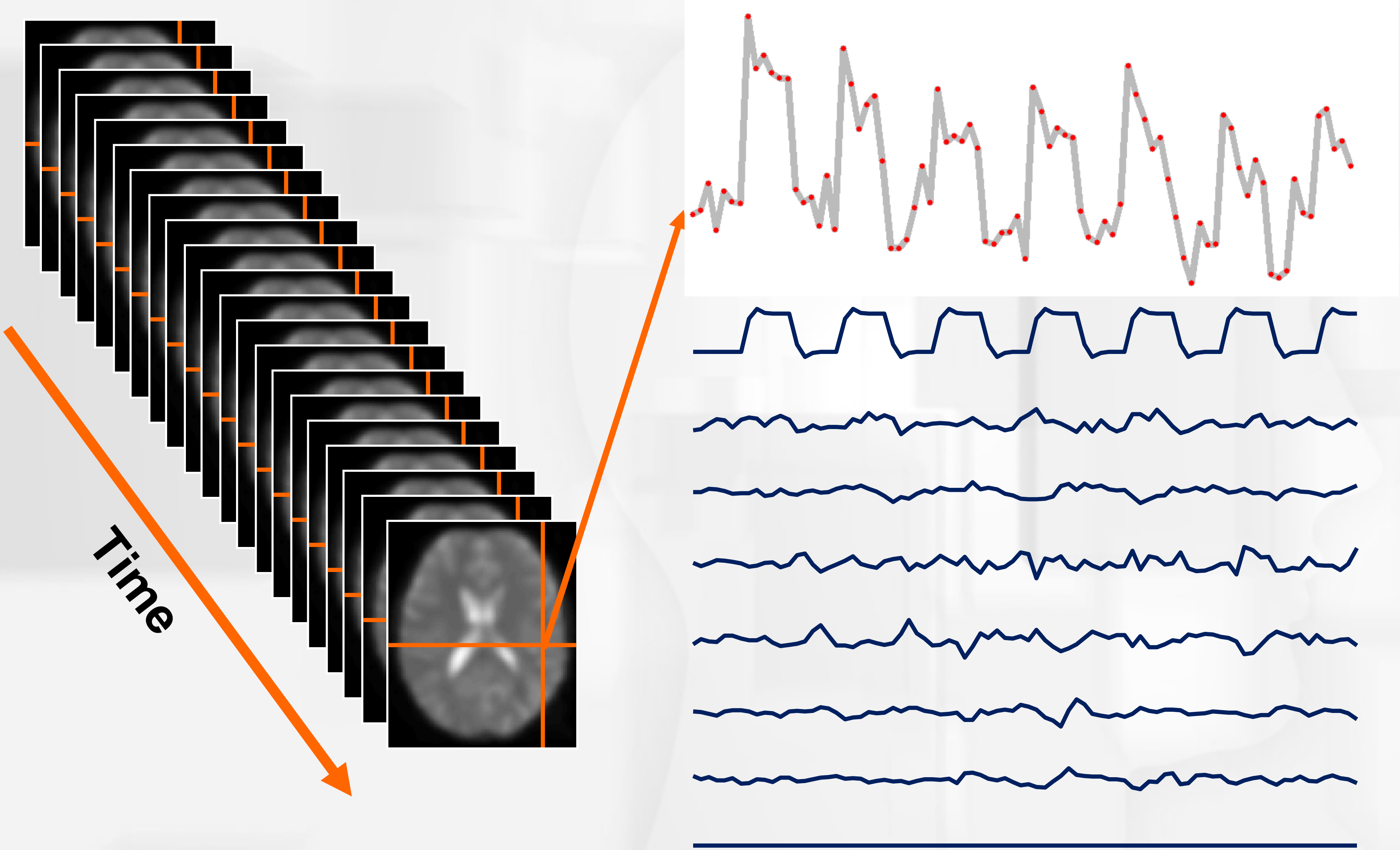
Outline

- Model and fit the data using the General Linear Model (GLM): a toy example
- T- and F-tests
  - What do they measure exactly?
- Multicollinearity

Outline

- Model and fit the data using the General Linear Model (GLM): a toy example
- T- and F-tests
  - What do they measure exactly?
- Multicollinearity

A mass-univariate approach

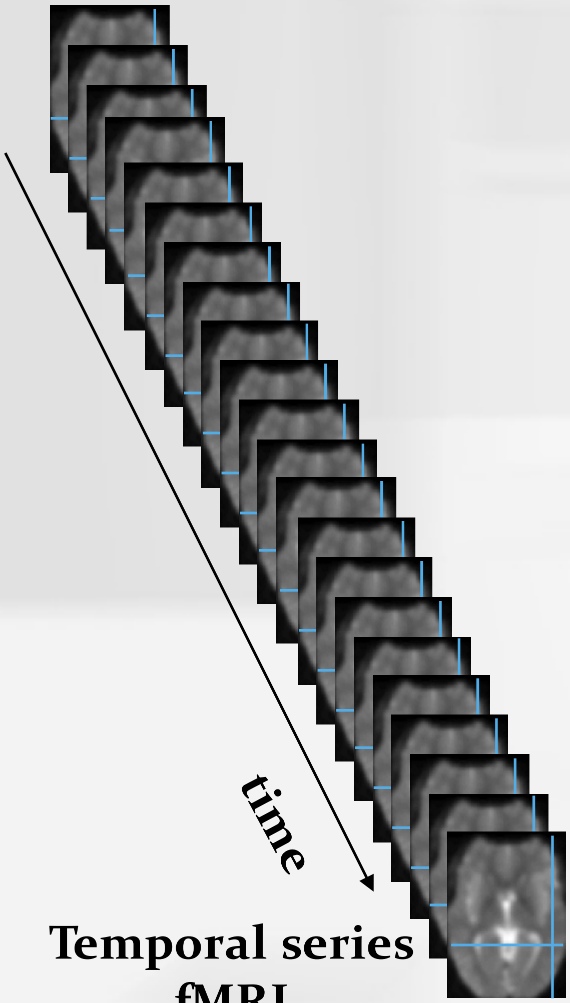


Imagine a very simple experiment…

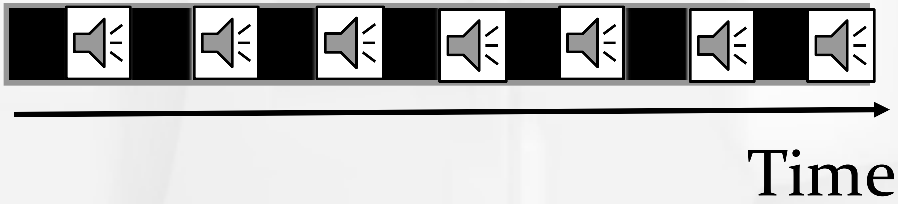
Toy Example:

Where in the brain do we represent sounds?

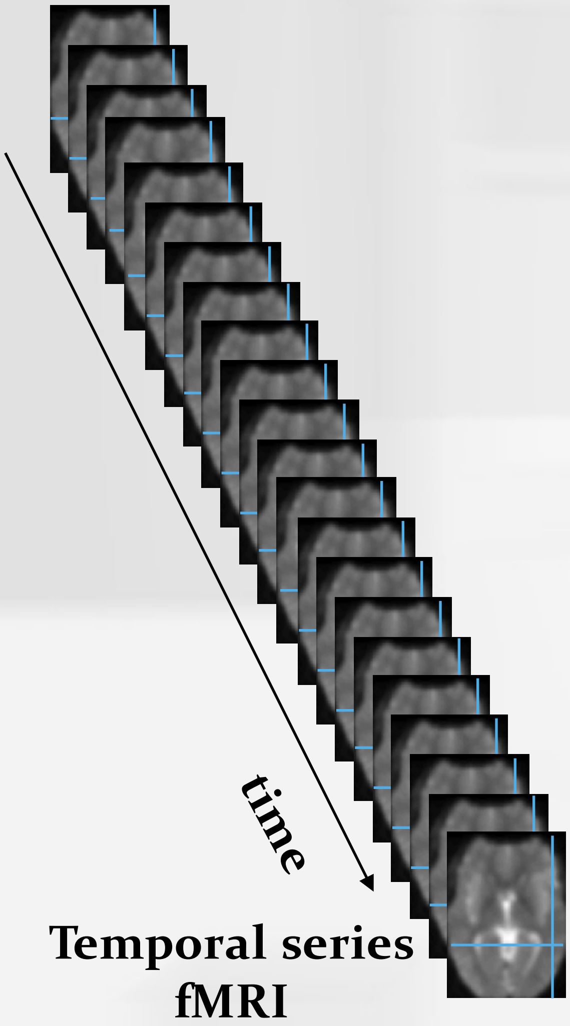
Imagine a very simple experiment…

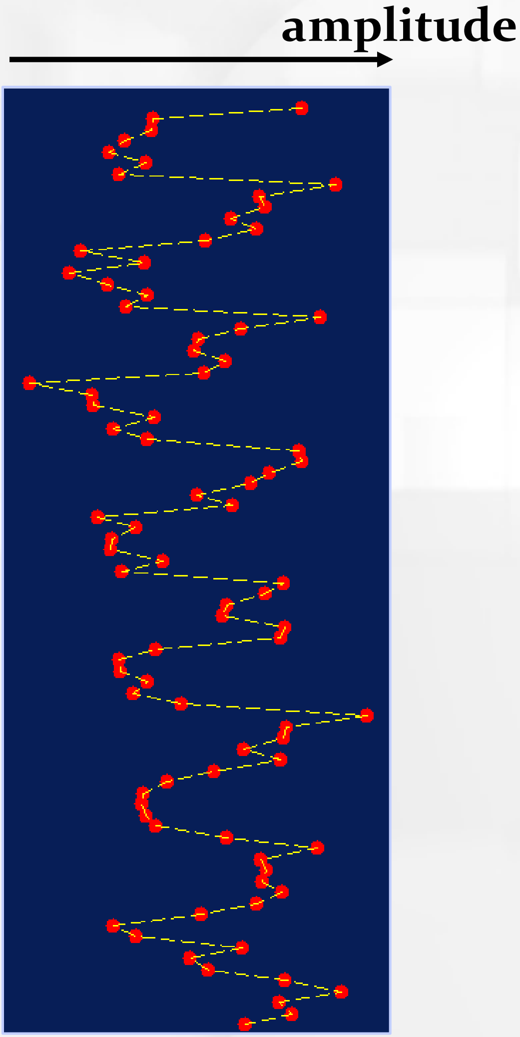


Temporal series fMRI



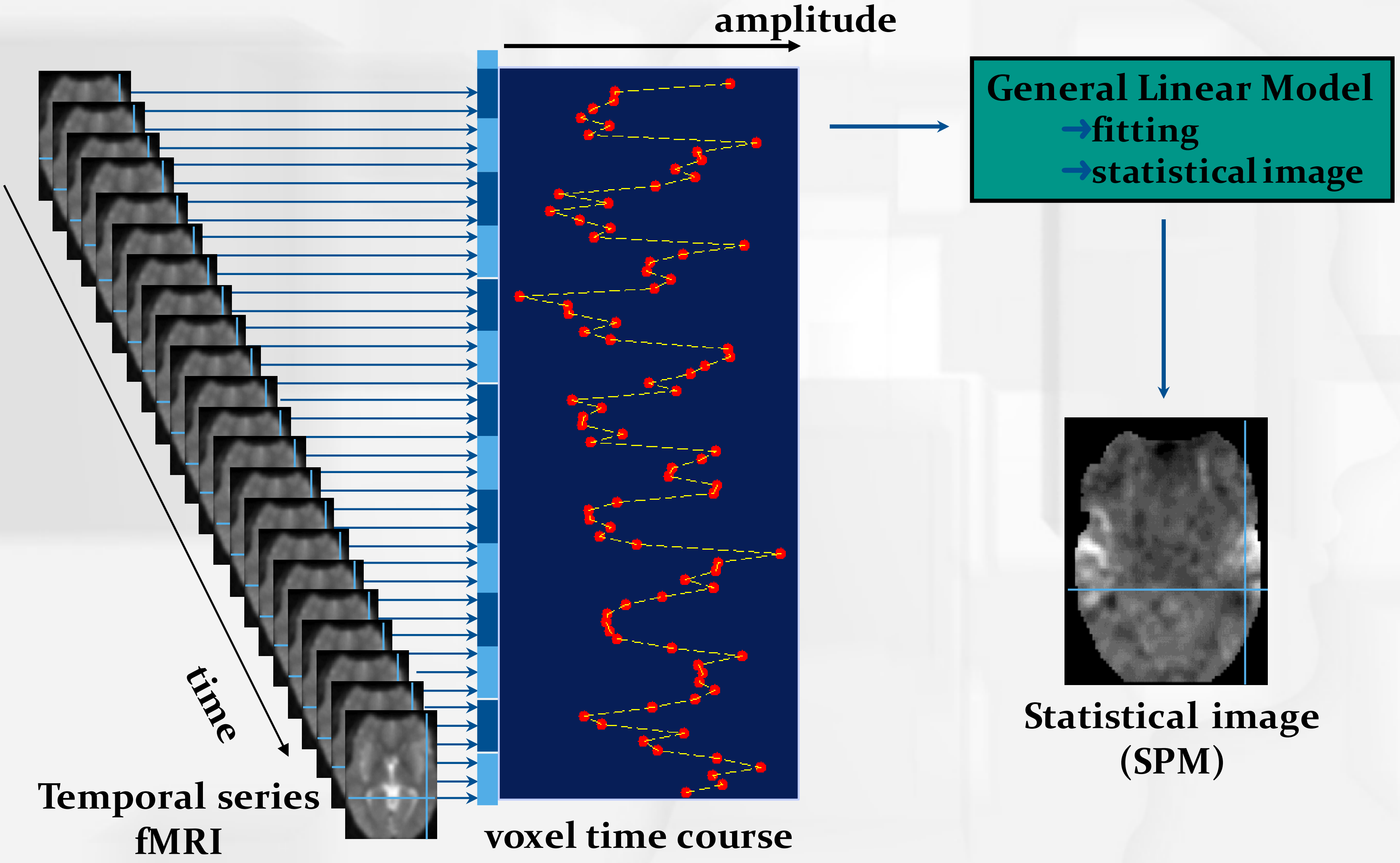
One voxel = One test (t, F, ...)





voxel time course

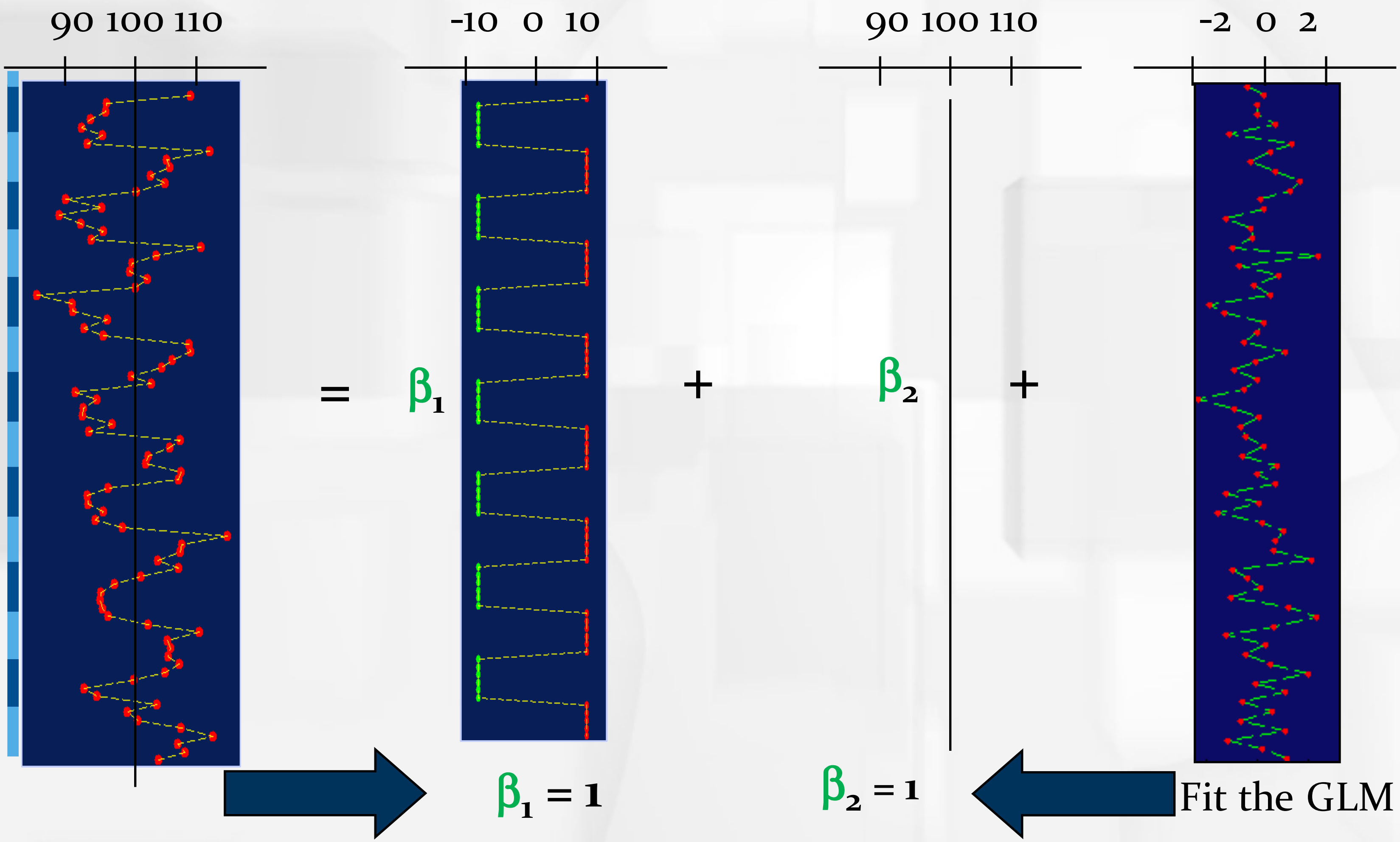
One voxel = One test (t, F, ...)



General Linear Model
→fitting
→statistical image

Statistical image
(SPM)

Regression example...



$\beta_1 = 1$

$\beta_2 = 1$

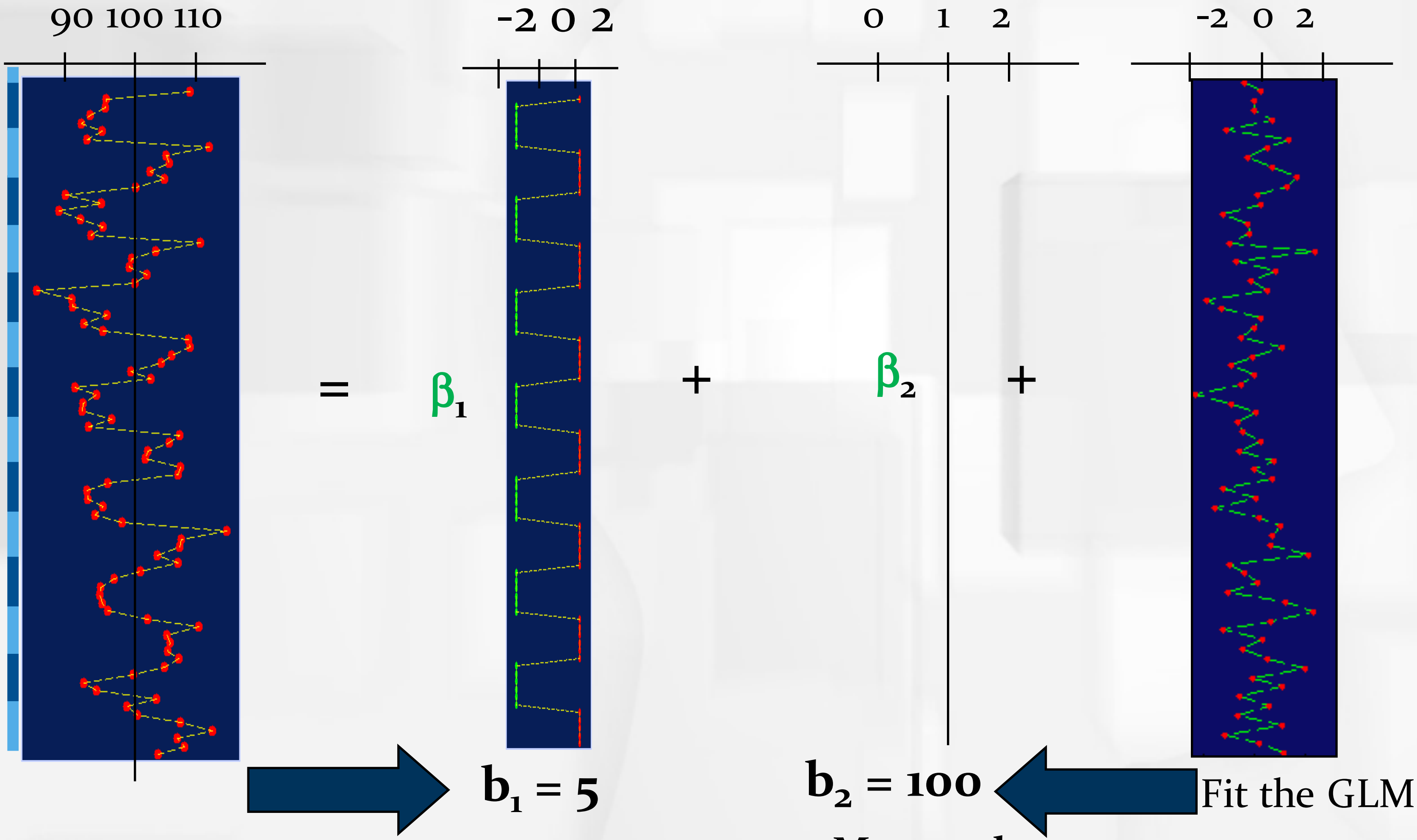
Fit the GLM

Mean value

voxel time series

box-car reference function

Regression example...



$b_1 = 5$

$b_2 = 100$

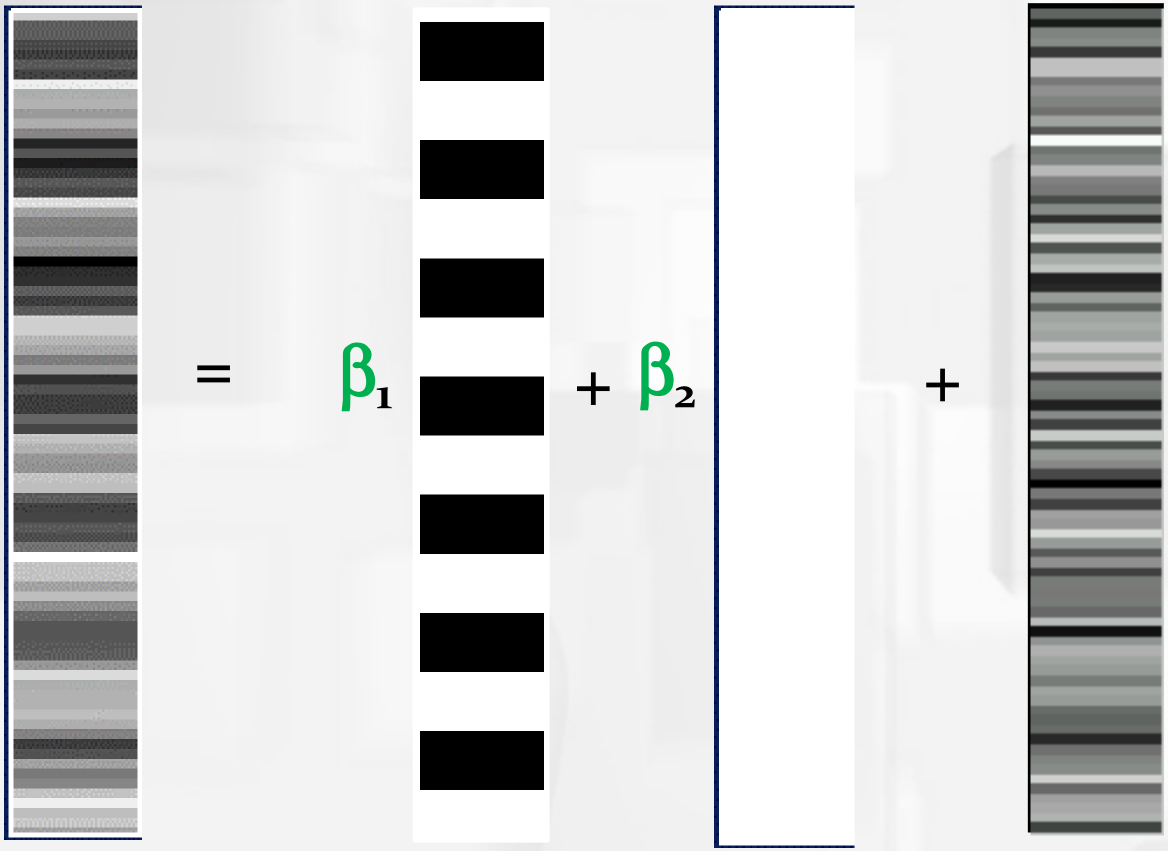
Fit the GLM

Mean value

voxel time series

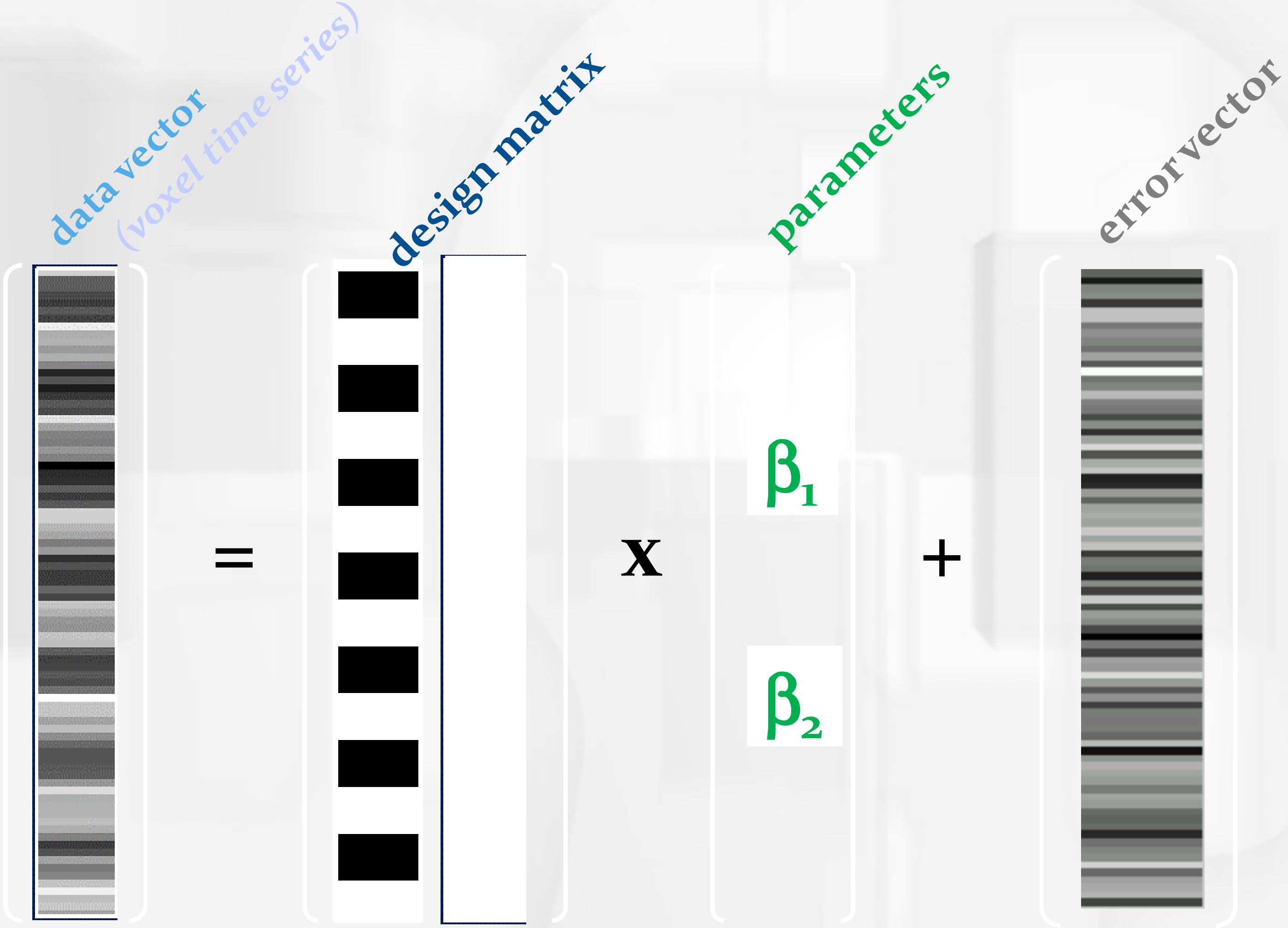
box-car reference function

...revisited : matrix form



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

Box car regression: design matrix…



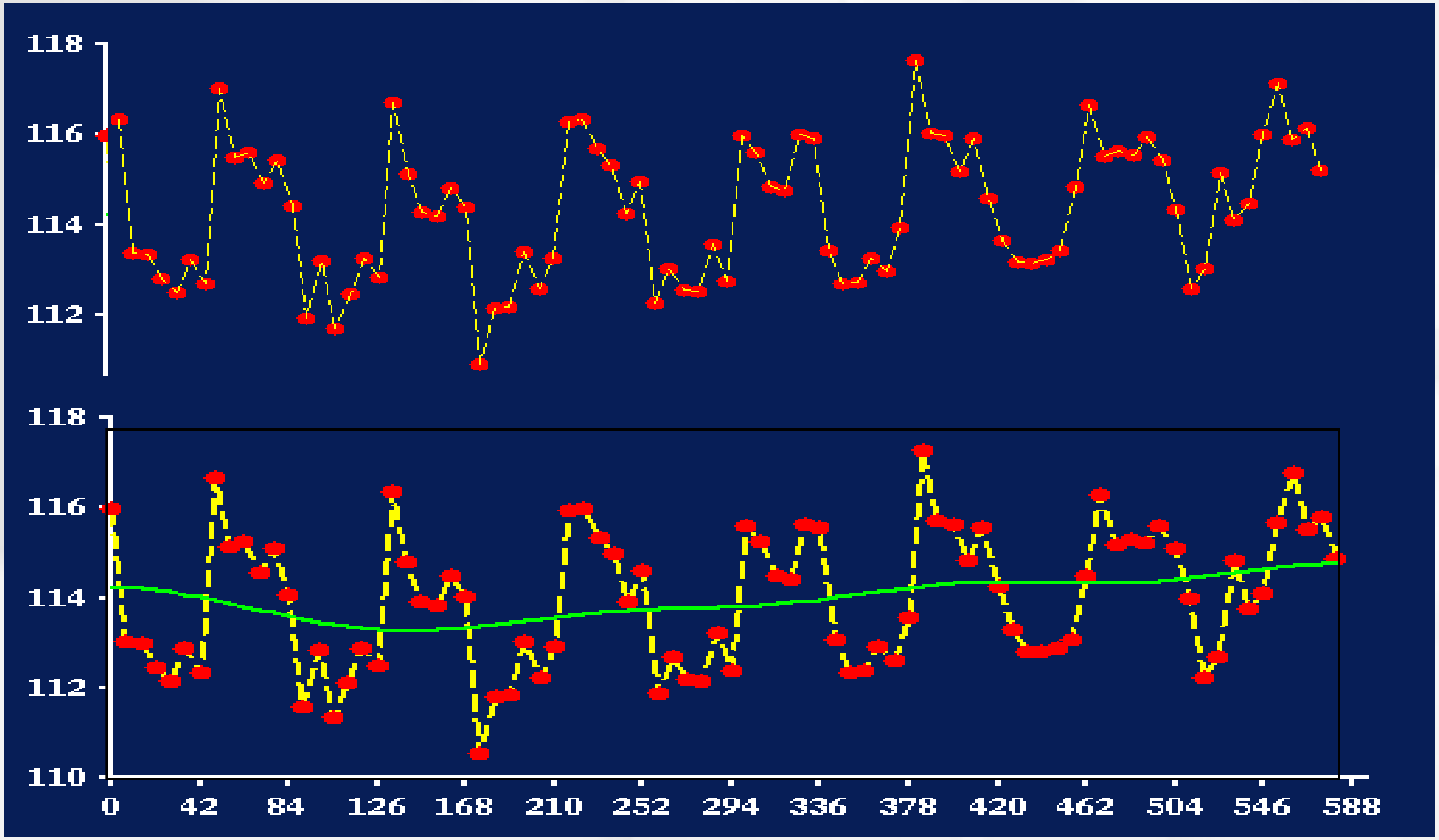
$$\underline{Y} = X \times \underline{\beta} + \underline{\varepsilon}$$

Fact: model parameters depend on regressors scaling

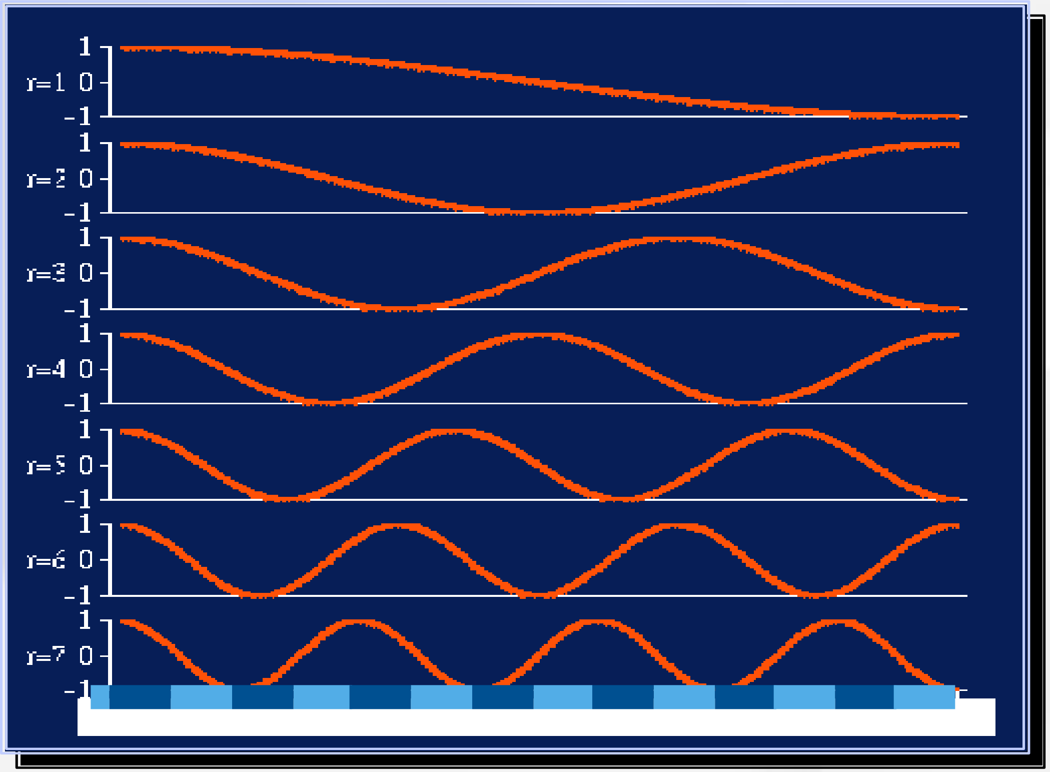
ONLY when comparing manually entered regressors (e.g., if you would like to compare two scores)

careful when comparing two (or more) manually entered regressors!

What else can we include in our design matrix?

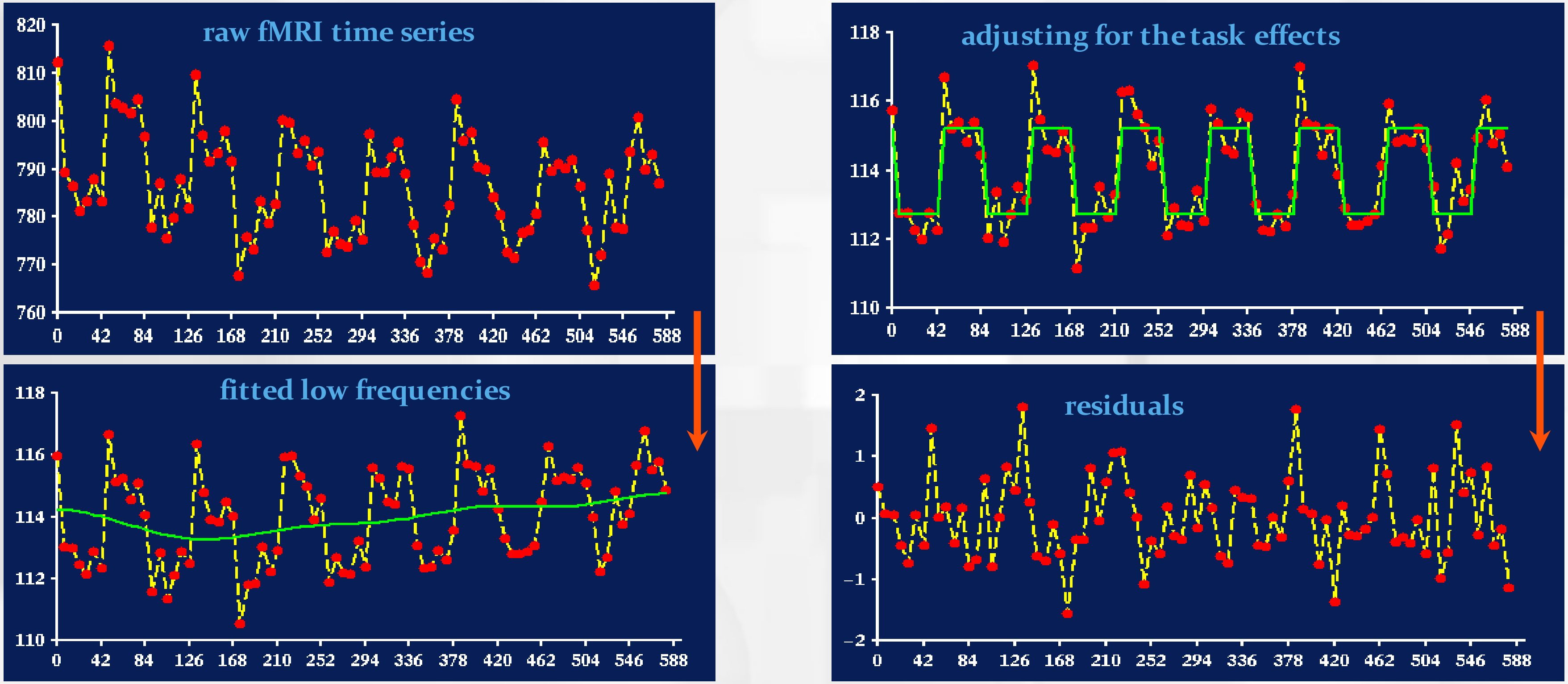


Nuisance Regressors



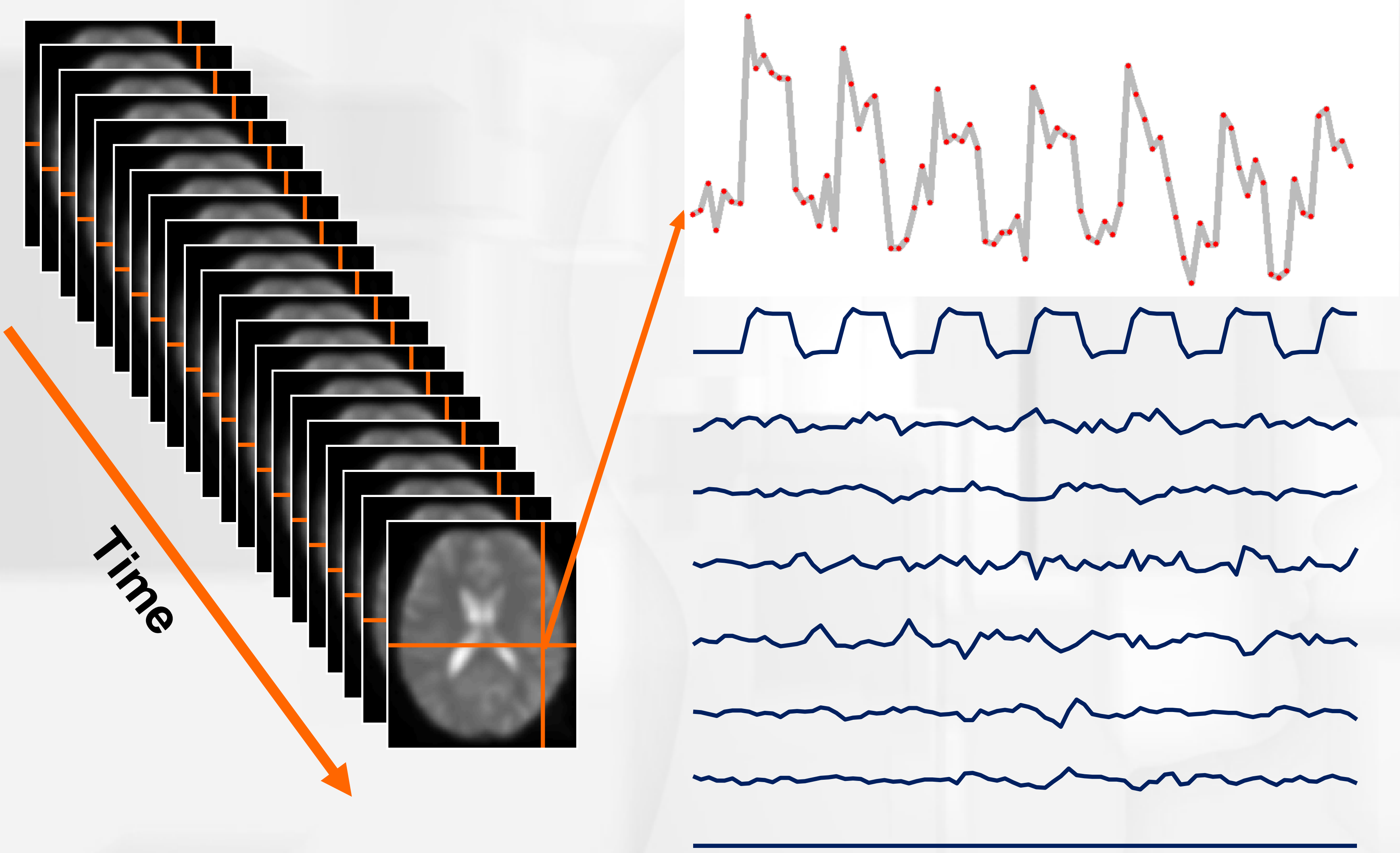
Discrete cosine transform basis functions

Fitting the model = finding some estimate of the betas

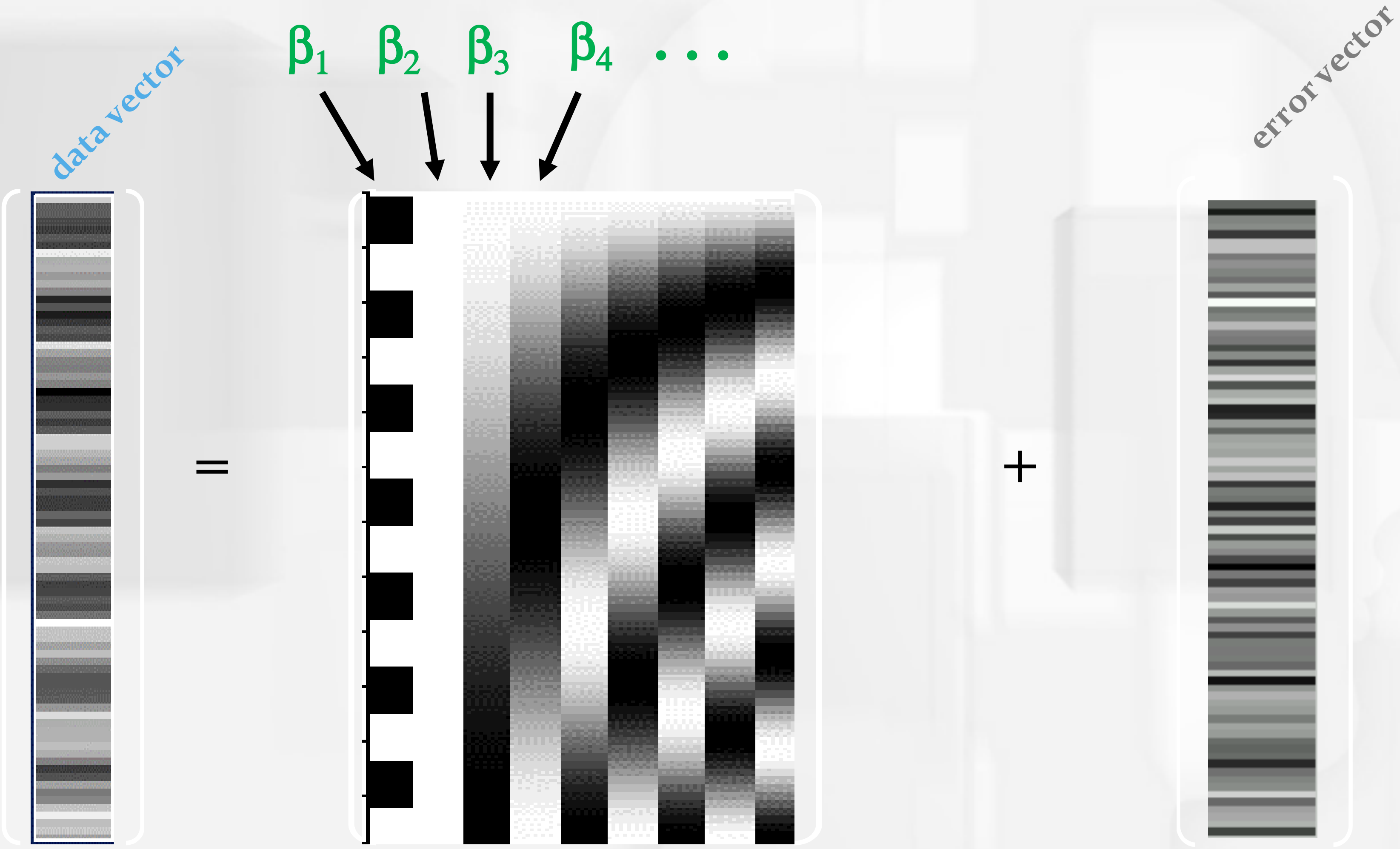


How do we find the beta estimates? By minimizing the residual variance

A mass-univariate approach

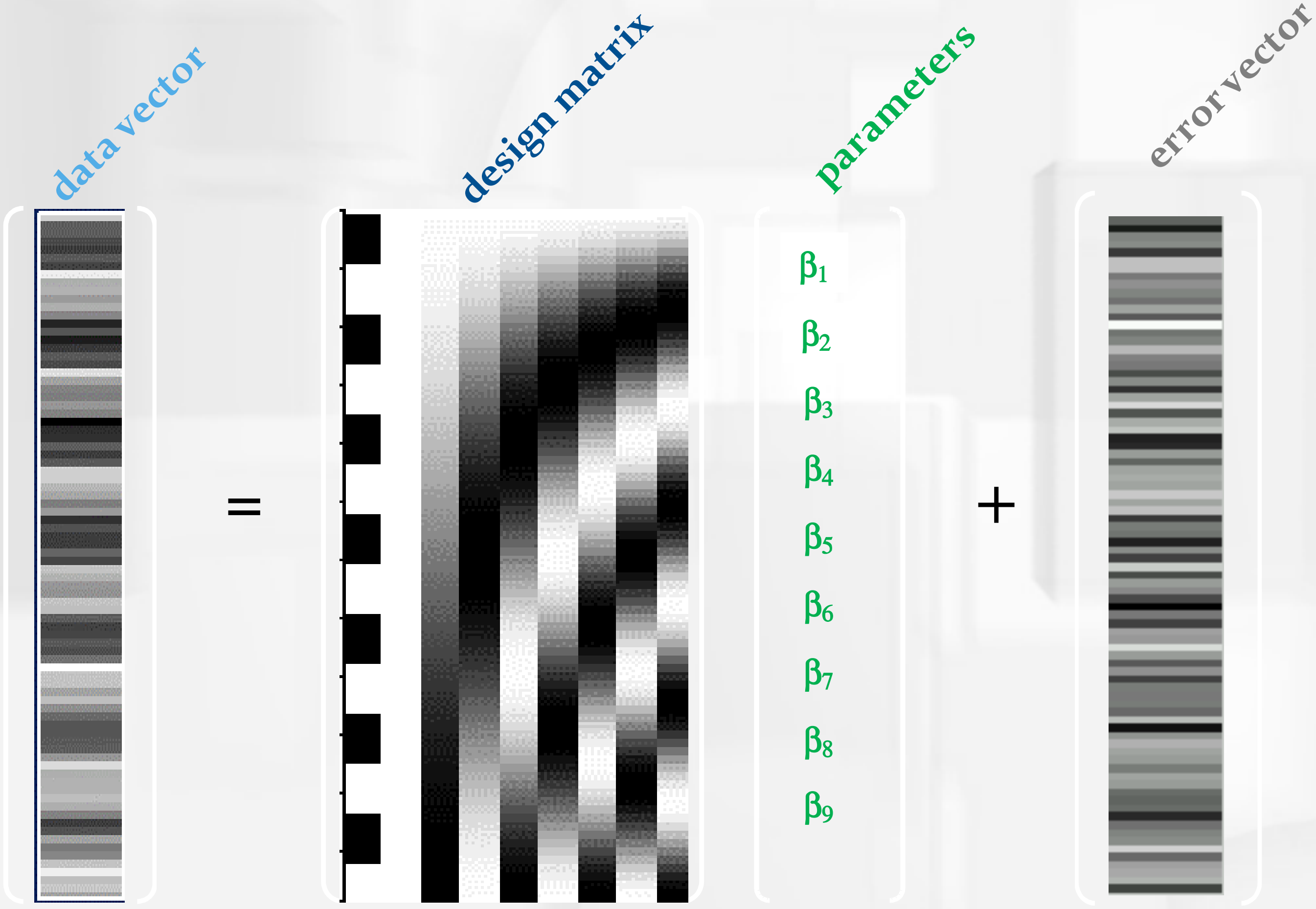


...design matrix



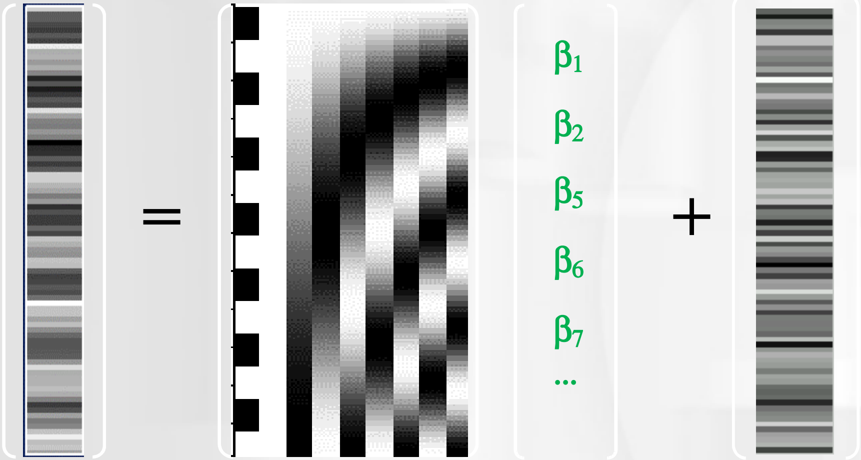
$$Y = X \times \beta + \varepsilon$$

...design matrix



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

Fitting the model = finding an estimate of beta



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

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

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

Suppose β is a candidate value for beta:

$$\left\| y_i - x_i^T \beta \right\|^2 = \sum_i \varepsilon_i^2 = \sum_i \left[y_i - x_i^T \beta \right]^2 = (y_\mathrm{i} - x_i \beta)^T (y_\mathrm{i} - x_i \beta)$$

Fitting the model = finding an estimate of beta

$$s(\beta) = (y_{\mathrm{i}} - x_i\beta)^T(y_{\mathrm{i}} - x_i\beta)$$

Since the function $s(\beta)$ is quadratic in β, it possesses a global minimum at $\beta = \hat{\beta}$

$$\hat{\beta} = \underset{\beta}{argmin}\, s(\beta) = \frac{\partial\big(s(\beta)\big)}{\partial\beta}$$

$$\frac{\partial((y_{\mathrm{i}} - x_i\beta)^T(y_{\mathrm{i}} - x_i\beta))}{\partial\beta} = -2X^Ty + 2X^TX\beta$$

We have to find a value for $\hat{\beta}$ where this expression is zero:

$$-2X^Ty + 2X^TX\hat{\beta} = 0$$

$$\hat{\beta} = (X^TX)^{-1}X^Ty$$

To summarize

The Ordinary Least Squares (OLS) estimators are:

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

These estimators minimise $\sum_i \varepsilon_i^2$

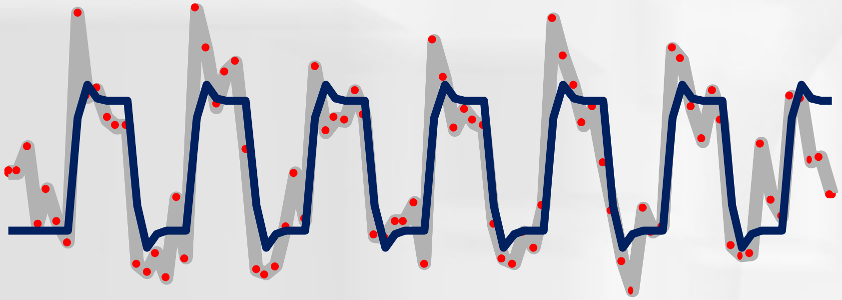
They are found solving $\dfrac{\partial(\sum_i \varepsilon_i^2)}{\partial \hat{\beta}_i}$

Under i.i.d. assumptions, the OLS estimates correspond to ML estimates:

$$\begin{tikzcd}
\varepsilon \sim N(0,\sigma^2 I) \arrow[r] \arrow[dr] & Y \sim N(X\beta, \sigma^2 I) \\
& \hat{\beta} \sim N(\beta, \boxed{\sigma^2 (X^T X)^{-1}})
\end{tikzcd}$$

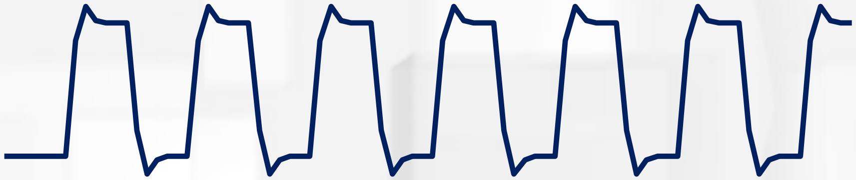
NB: precision of our estimates depends on design matrix!

To summarize

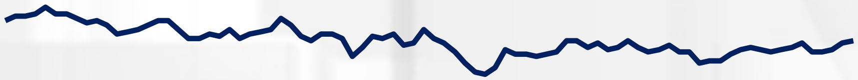


OLS estimates: $\hat{\beta} = (X^T X)^{-1} X^T y$

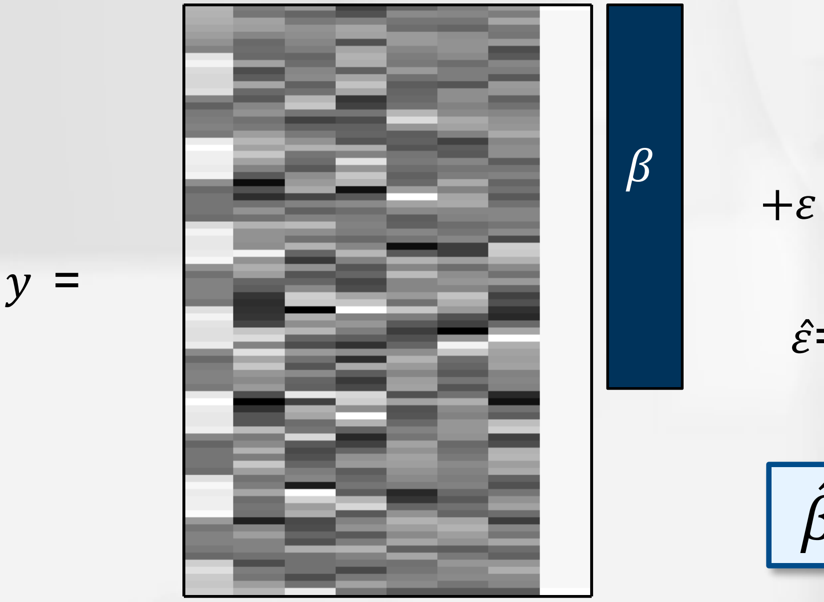
$\hat{\beta}_1 = 3.9831$



$\hat{\beta}_{2-7} = \{0.6871, 1.9598, 1.3902, 166.1007, 76.4770, -64.8189\}$



$\hat{\beta}_8 = 131.0040$



$\hat{\varepsilon} =$

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$$

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{\boxed{N-p}}$$

Take home message

- We put our model regressors (covariates) that represent how we think the signal is varying (of interest or no interest)
  - Which one to include
  - What if too many or too few?

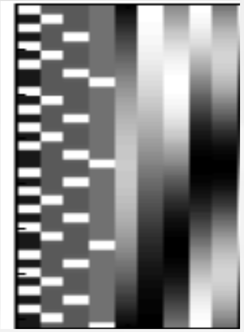
- Coefficients (or parameters) are estimated by minimizing fluctuations (variance) of the estimated noise (or residual error)

Outline

- Model and fit the data using the General Linear Model (GLM)
- T- and F-tests
  - What do they measure exactly?
- Multicollinearity

Statistical Inference: Contrasts

- We are usually not interested in the whole β vector.
- A contrast $c^T\beta$ selects a specific effect of interest:
  - ⇨ $c^T\beta$ is a linear combination of regression coefficients β



$c^T = [1\ 0\ 0\ 0\ 0\ \ldots]$

$$c^T\beta = 1\beta_1 + 0\beta_2 + 0\beta_3 + 0\beta_4 + 0\beta_5 + \ldots$$

$c^T = [0\ \text{-}1\ 1\ 0\ 0\ \ldots]$

$$c^T\beta = 0\beta_1 + \text{-}1\beta_2 + 1\beta_3 + 0\beta_4 + 0\beta_5 + \ldots$$

- Under i.i.d assumptions:

$$\boxed{c^T\hat{\beta} \sim N(c^T\beta, \sigma^2 \boxed{c^T(X^TX)^{-1}c})}$$

NB: the precision of our estimates depends on design matrix and the chosen contrast !

Statistical Inference

- ## T-test
- ## F-Test

To test a hypothesis, we construct a “test statistic”.

- **“Null hypothesis” H_0 = “there is no effect” $\Rightarrow c^T\beta = 0$**

  This is what we want to disprove.

  ⇨ The “alternative hypothesis” H_1 represents the outcome of interest.

- **The test statistic T**

  The test statistic summarises the evidence for H_0.

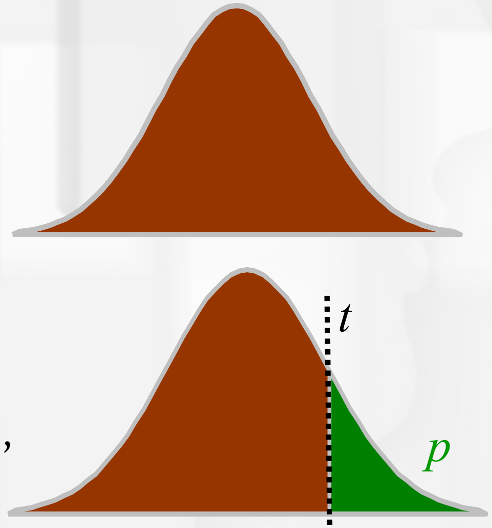
  ⇨ We need to know the distribution of T under the null hypothesis.

- **Observation of test statistic t, a realisation of T**

  A p-value summarises evidence against H_0.

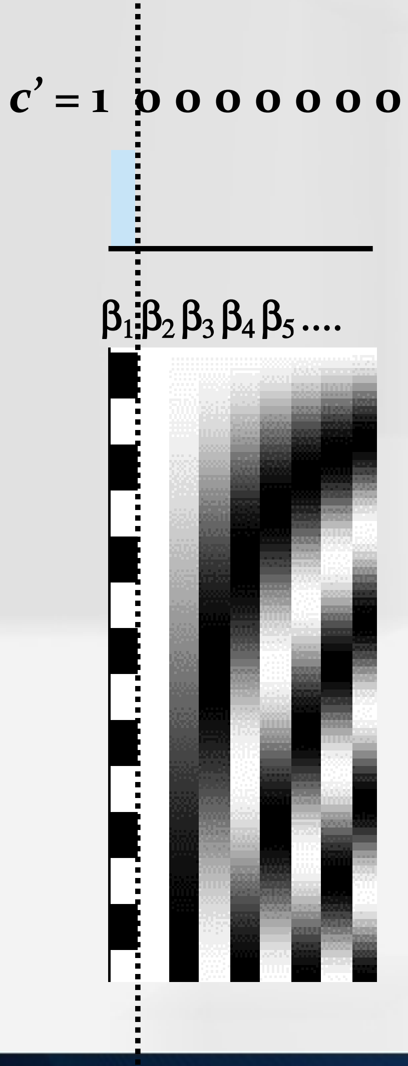
  This is the probability of observing t, or a more extreme value, under the null hypothesis:

$$\boxed{p(T \geq t \mid H_0)}$$



Null Distribution of T

T-test: one dimensional contrast SPM {t}



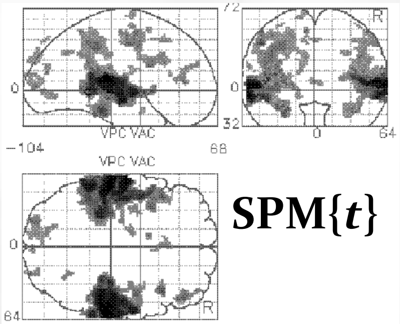
A contrast = a weighted sum of parameters: $c^T\beta$

$\beta_1 > 0$?

Compute $1x\beta_1 + 0x\beta_2 + 0x\beta_3 + 0x\beta_4 + 0x\beta_5 + \ldots = c^T\beta$
$c^T = [1\ 0\ 0\ 0\ 0\ \ldots.]$

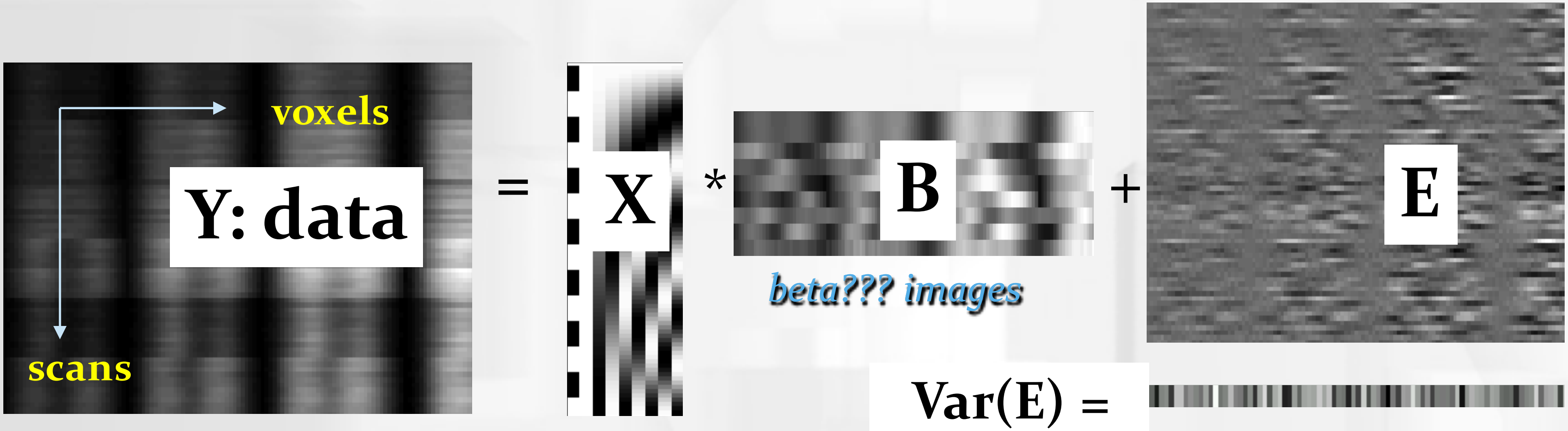
divide by estimated standard deviation of β_1

$$T = \frac{\textit{contrast of estimated parameters}}{\sqrt{\text{variance estimate}}}; \quad T = \frac{c^T\hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T \boxed{(X^TX)}^{-1} c}}$$



SPM{t}

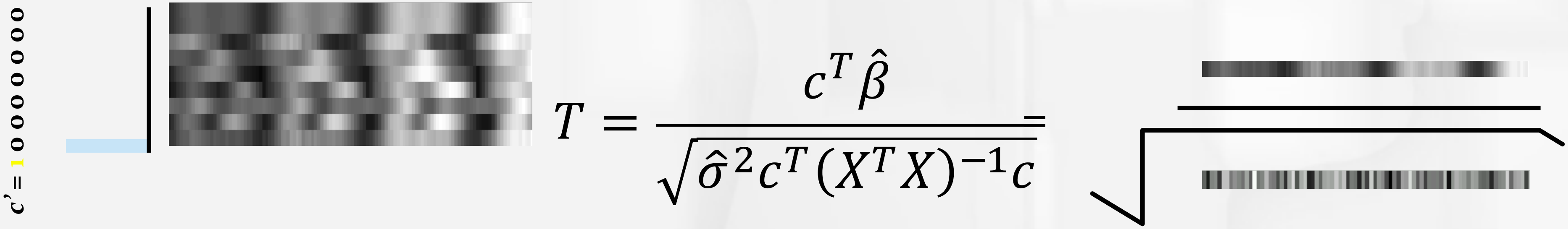
From one time series to an image



beta??? images

Var(E) = s²

spm_ResMS



$$T = \frac{c^T \hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}} =$$

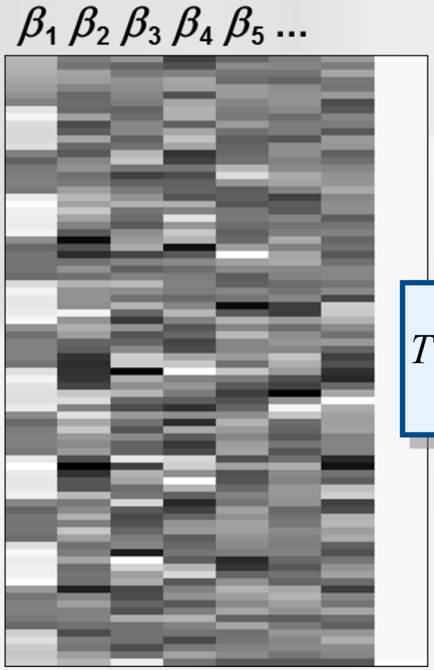
spm_con??? images

spm_t??? images

T-test: a simple example

Passive word listening versus rest

$c^T = 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0$



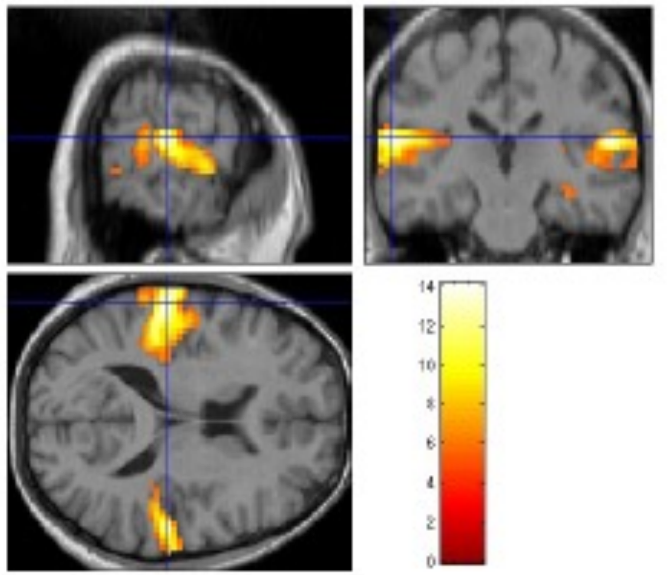
Q: activation during listening ?

Null hypothesis:

$$H_0 \colon c^T \beta = 0$$

Test statistic:

$$T = \frac{c^T \hat{\beta}}{\sqrt{\mathrm{var}(c^T \hat{\beta})}} = \frac{c^T \hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T \left(X^T X\right)^{-1} c}} \sim t_{N-p}$$

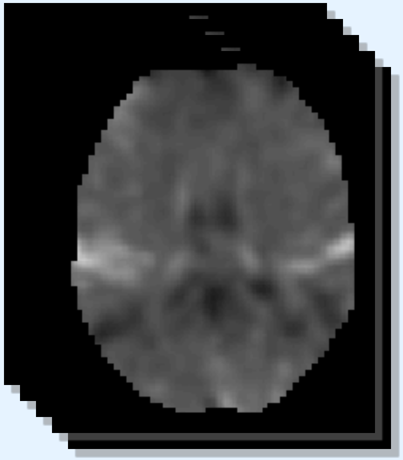


SPMresults: Threshold T = 3.2057 {p<0.001}
voxel-level

	$(Z_{\underline{o}})$	$p_{\text{uncorrected}}$	Mm	mm	mm
13.94	**Inf**	**0.000**	**-63**	**-27**	**15**
12.04	Inf	0.000	-48	-33	12
11.82	Inf	0.000	-66	-21	6
13.72	**Inf**	**0.000**	**57**	**-21**	**12**
12.29	Inf	0.000	63	-12	-3
9.89	7.83	0.000	57	-39	6
7.39	**6.36**	**0.000**	**36**	**-30**	**-15**
6.84	**5.99**	**0.000**	**51**	**0**	**48**
6.36	**5.65**	**0.000**	**-63**	**-54**	**-3**

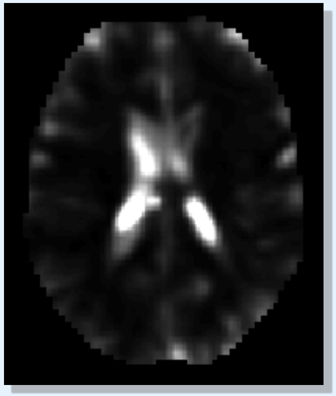
T-contrast in SPM

- For a given contrast *c*:



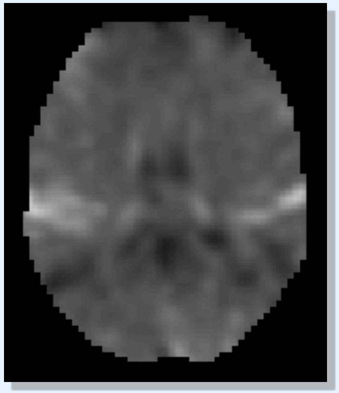
beta_???? images

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



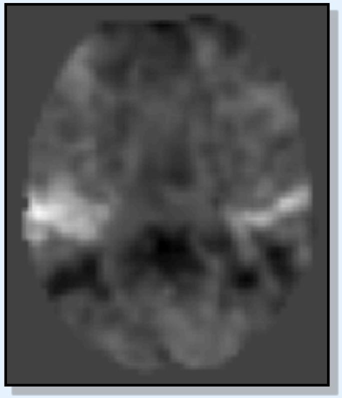
ResMS image

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{N - p}$$



con_???? image

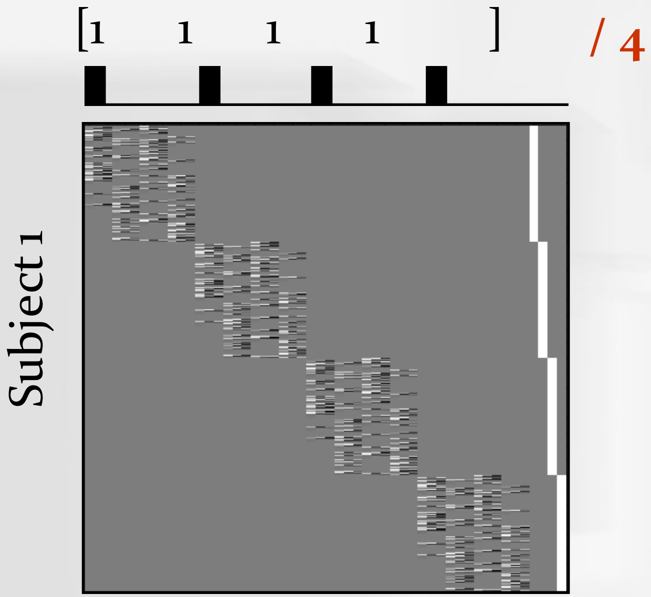
$$c^T \hat{\beta}$$

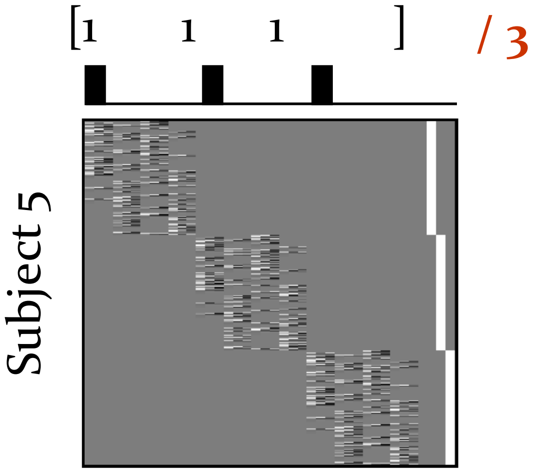


spmT_???? image

SPM{*t*}

Scaling issue





$$T = \frac{c^T \hat{\beta}}{\sqrt{\operatorname{var}(c^T \hat{\beta})}} = \frac{\boxed{c^T \hat{\beta}}}{\sqrt{\hat{\sigma}^2 \boxed{c}^T \left(X^T X\right)^{-1} \boxed{c}}}$$

☐ The *T*-statistic does not depend on the scaling of the contrast.

☐ Contrast $c^T \hat{\beta}$ depends on scaling.

➢ Be careful of the interpretation of the contrasts $c^T \hat{\beta}$ themselves (e.g., for a second level analysis):

sum ≠ average

T-test: summary

T-test is a *signal-to-noise* measure (ratio of estimate to standard deviation of estimate).

- ☐ Alternative hypothesis:

$$\mathrm{H_o}: \quad c^T \beta = 0 \qquad \text{vs} \qquad \mathrm{H_A}: \quad c^T \beta > 0$$

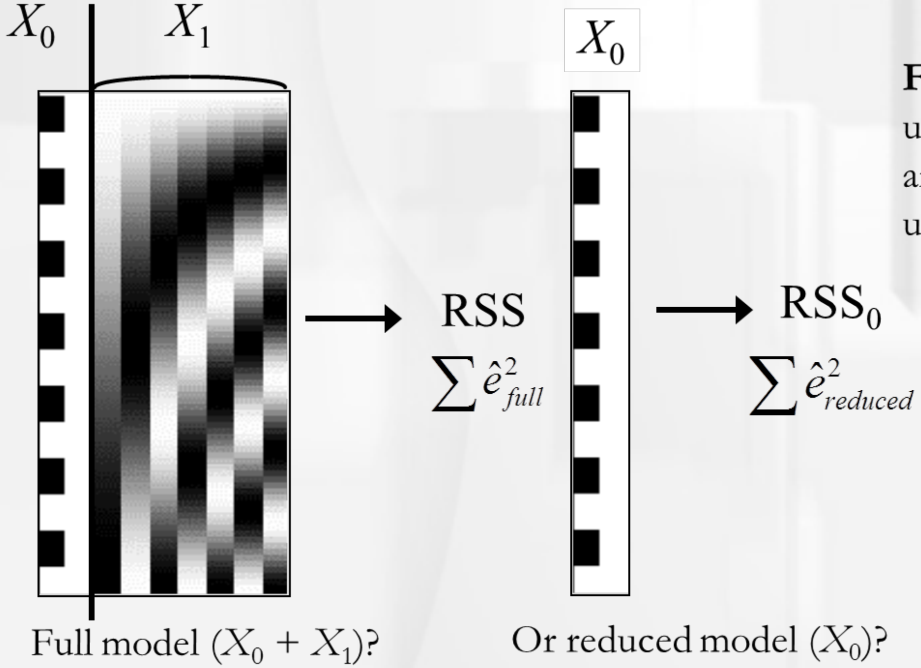
- ☐ *T*-contrasts are simple combinations of the betas; the T-statistic does not depend on the scaling of the regressors or the scaling of the contrast.

Statistical Inference

- T-test
- F-Test

Model comparison: Full vs. reduced model

Null Hypothesis H_0: True model is X_0 (reduced model)



Full model ($X_0 + X_1$)?

Or reduced model (X_0)?

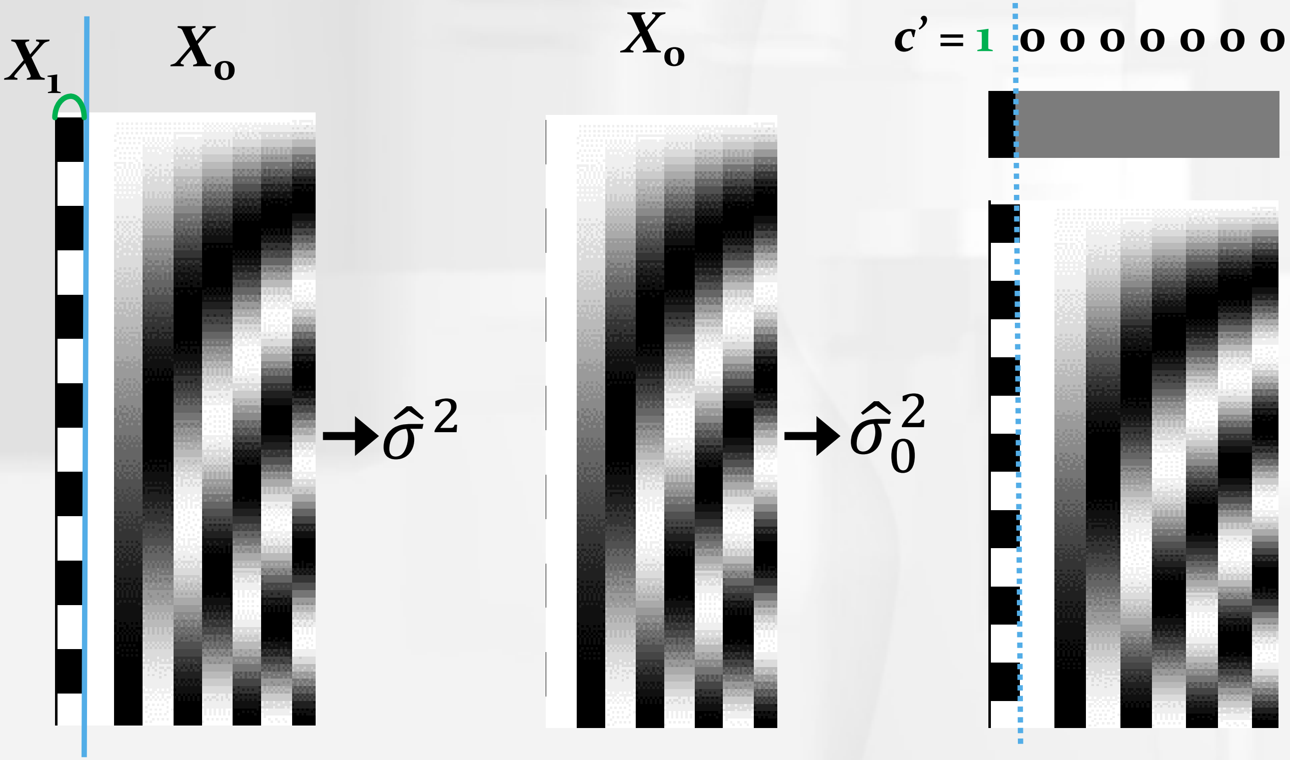
F-statistic: ratio of unexplained variance under X_0 and total unexplained variance under the full model

$$\boxed{F \propto \frac{RSS_0 - RSS}{RSS}}$$

F-test: model comparison

Does the block structure model anything ?

Null Hypothesis H_0: True model is X_0 (reduced model)



Full model ($X_0 + X_1$)? Or reduced model (X_0)?

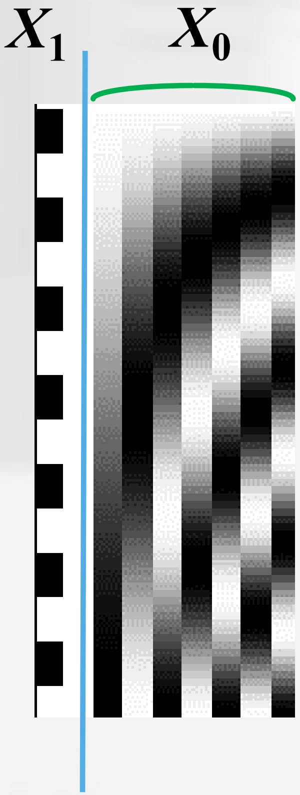
$$F = \frac{\text{unexplained variance under } X_0}{\text{total unexplained variance under the full model}}$$

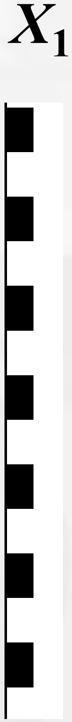
$$F \sim \left(\hat{\sigma}_0^2 - \hat{\sigma}^2\right) / \hat{\sigma}^2$$

F-test: model comparison

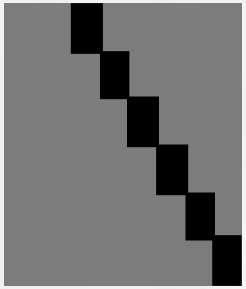
Does the nuisance regressor model anything ?

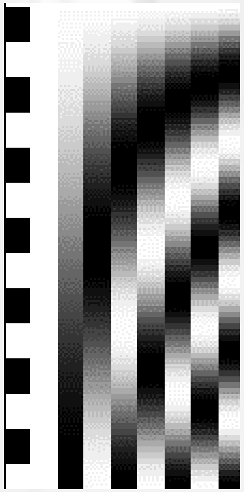
Null Hypothesis H$_0$: True model is X_0 (reduced model)





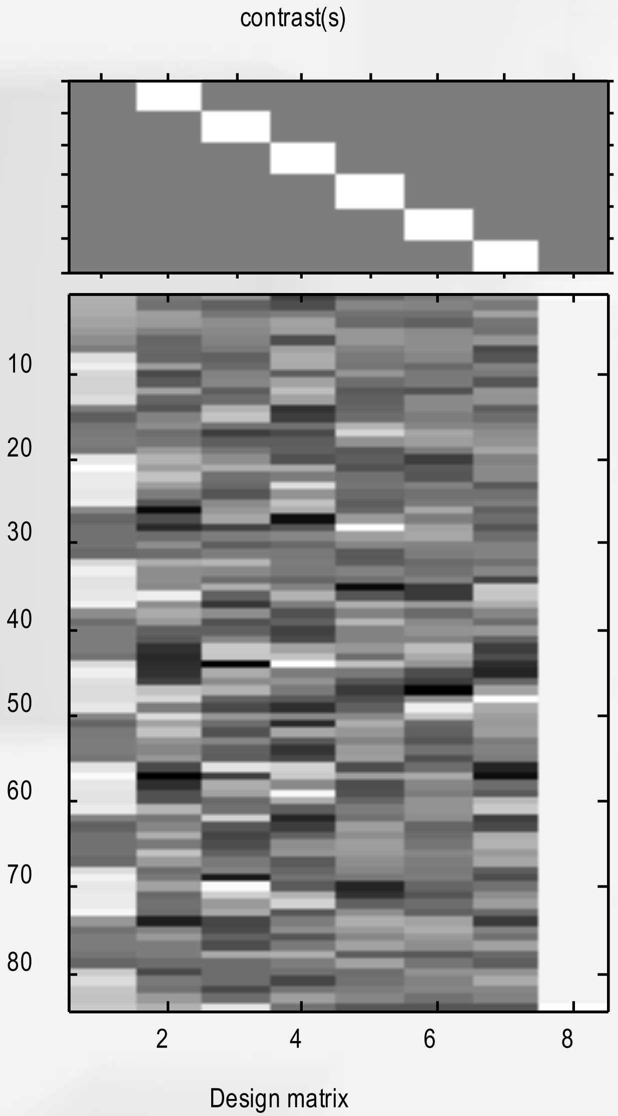
$$c' = \begin{matrix} 0&0&1&0&0&0&0&0 \\ 0&0&0&1&0&0&0&0 \\ 0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&0&1 \end{matrix}$$

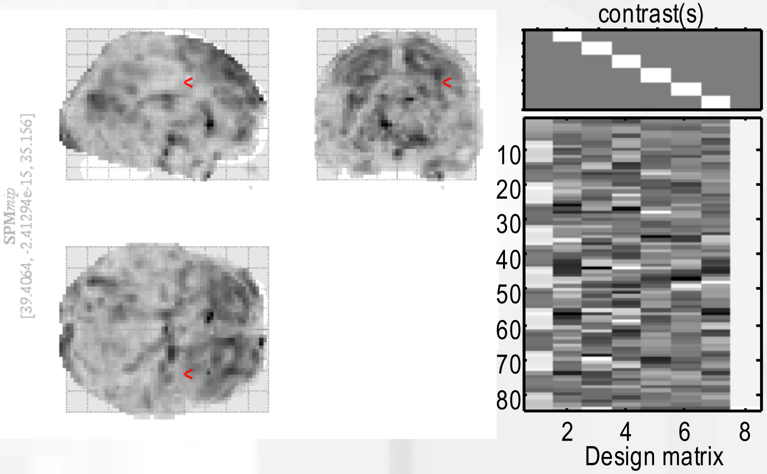


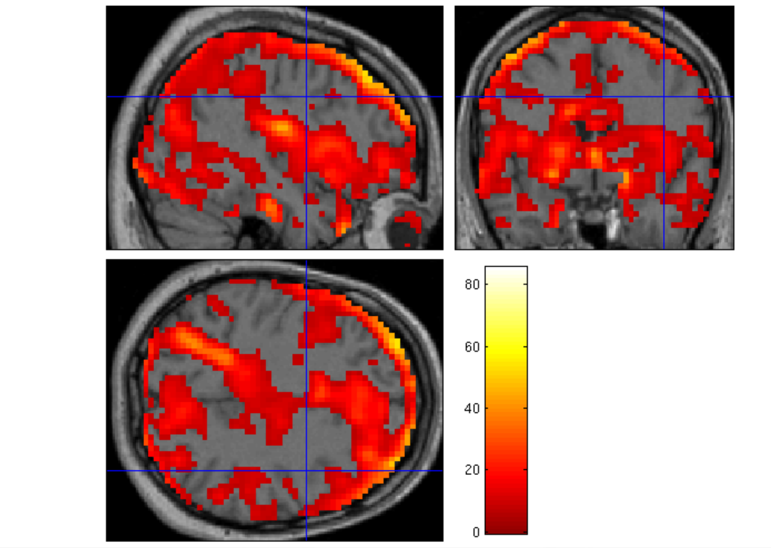


Full model (X_0+ X_1)? Or reduced model (X_0)?

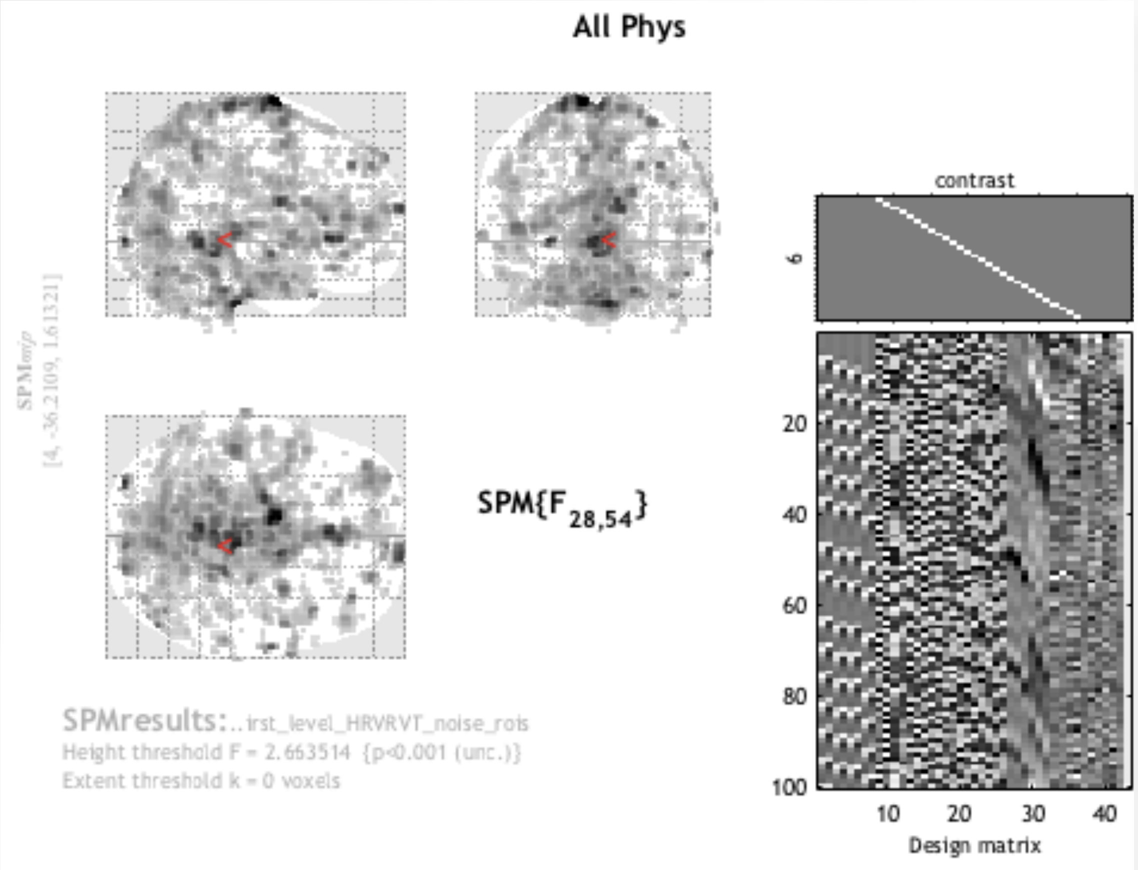
F-test example: movement related effects







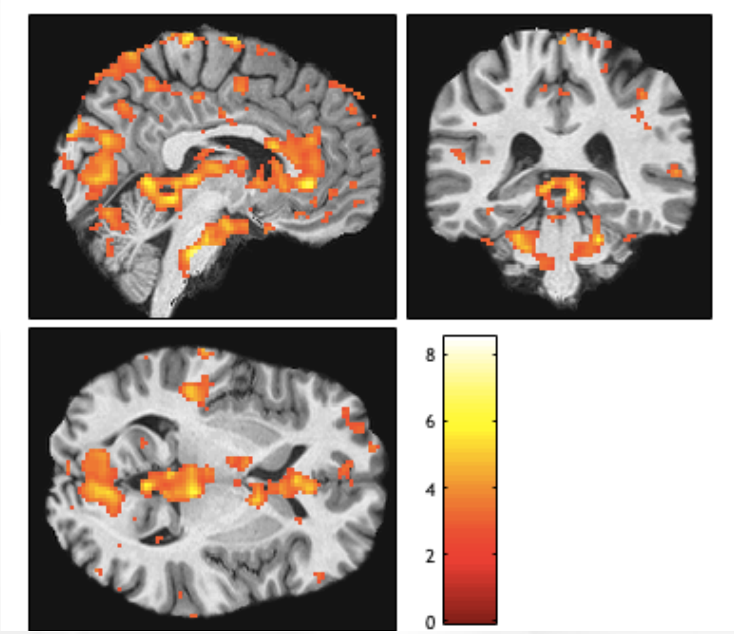
F-test example: physiological-noise related effects

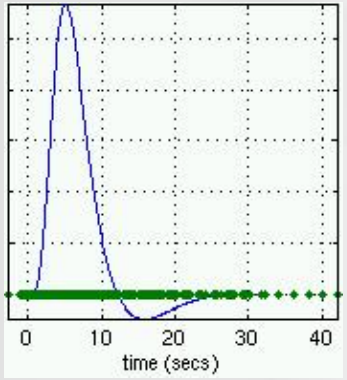
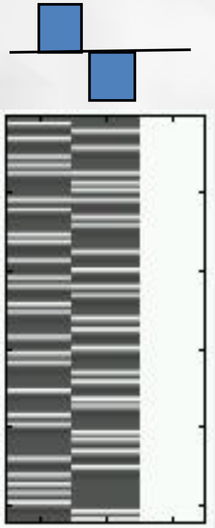
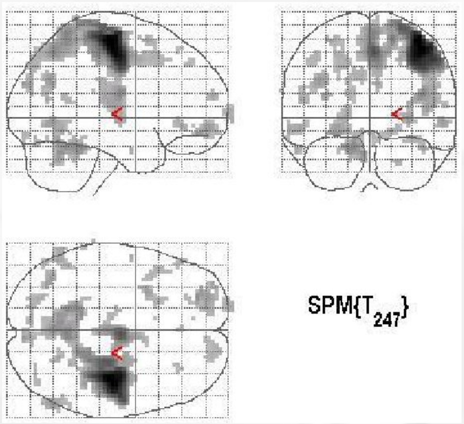
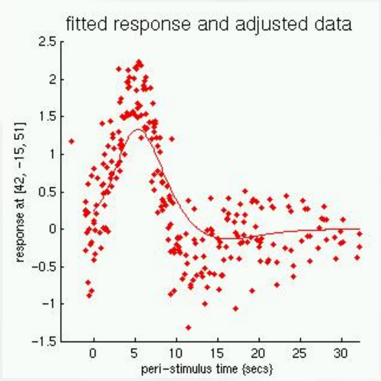
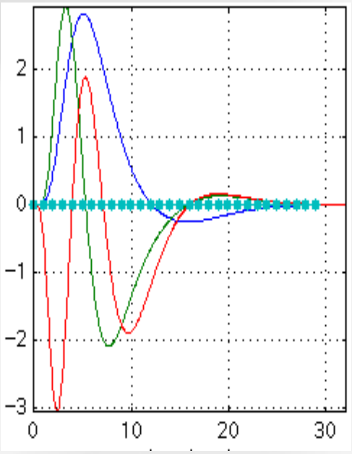
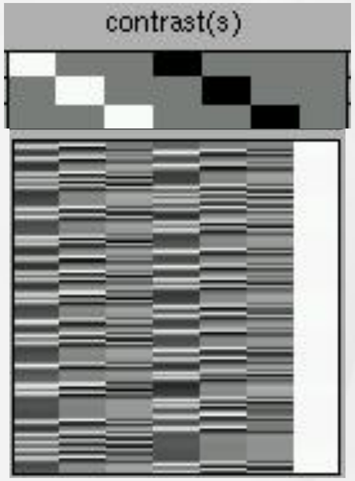
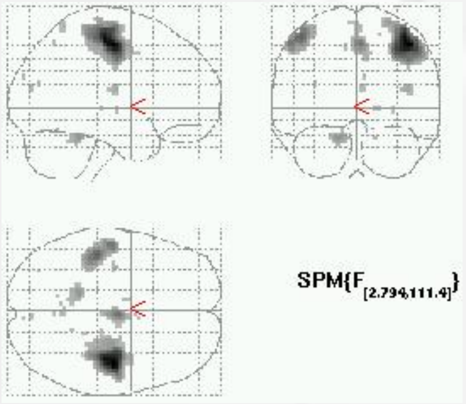
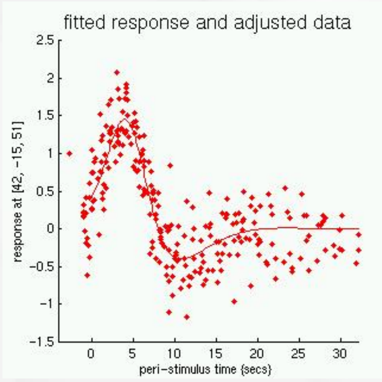


All Phys

SPM{F$_{28,54}$}

SPMresults:..irst_level_HRVRVT_noise_rois
Height threshold F = 2.663514 {p<0.001 (unc.)}
Extent threshold k = 0 voxels



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

T- and F-tests: 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
  - F tests are the square of one or several combinations of the betas

- When testing “simple 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.

Outline

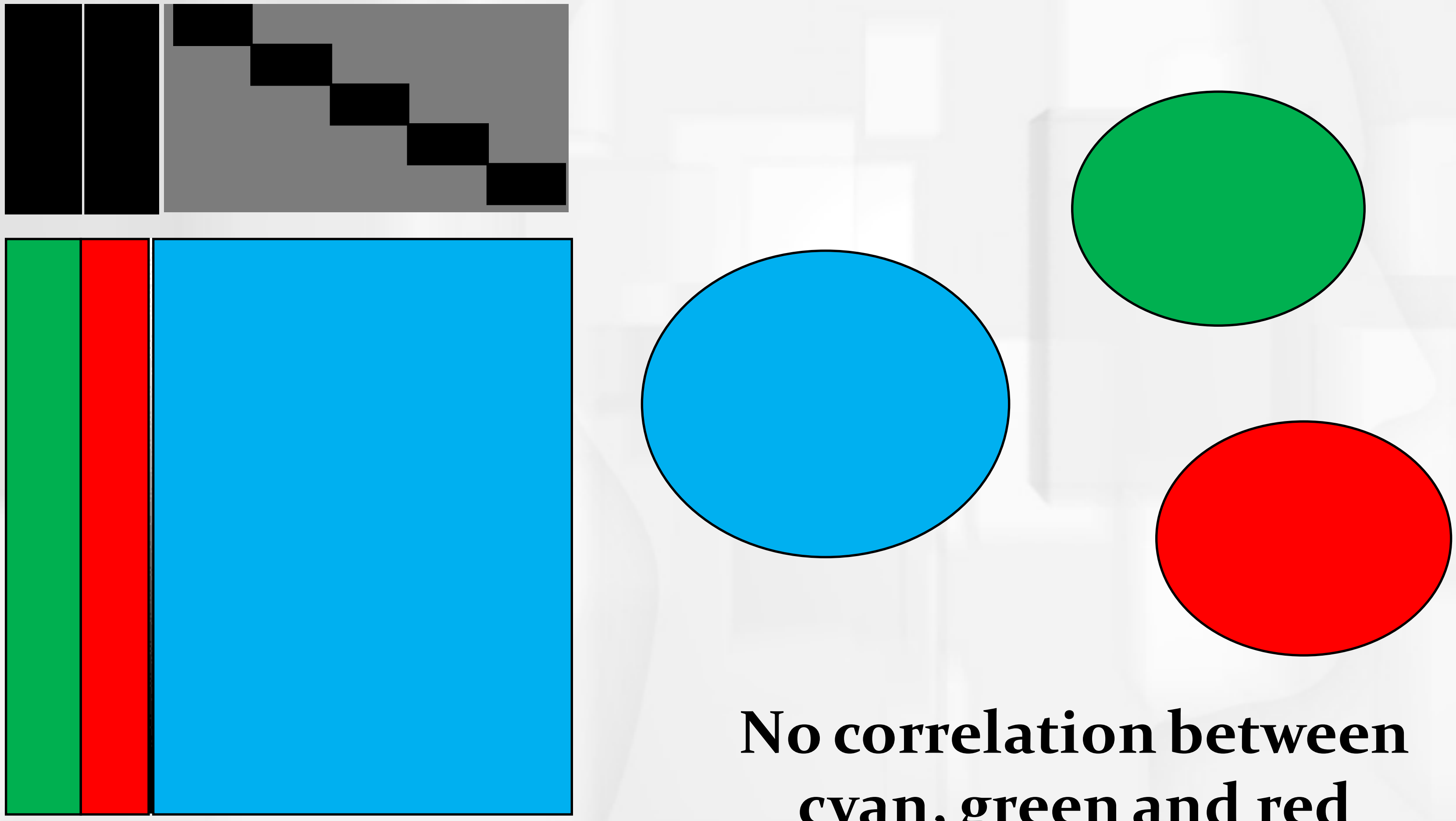
- Model and fit the data using the General Linear Model (GLM)
- T- and F-tests
    - What do they measure exactly?
- Multicollinearity

What is multicollinearity?

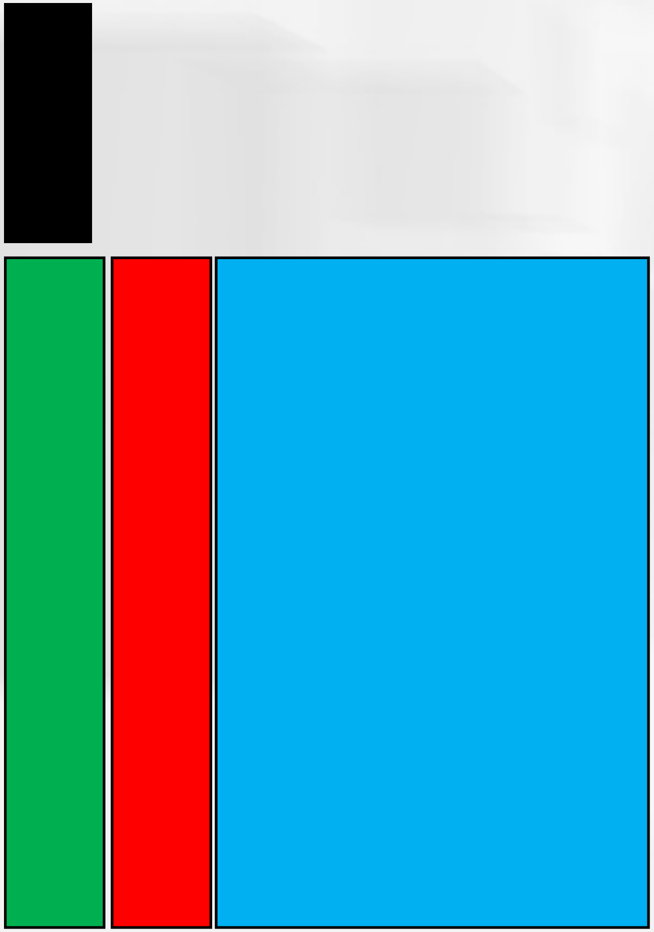
- Multicollinearity is a problem of fitting linear (regression) models when two or more predictor variables are **highly correlated**: one can be linearly predicted from the others with a substantial degree of accuracy.

- How do we counteract it?

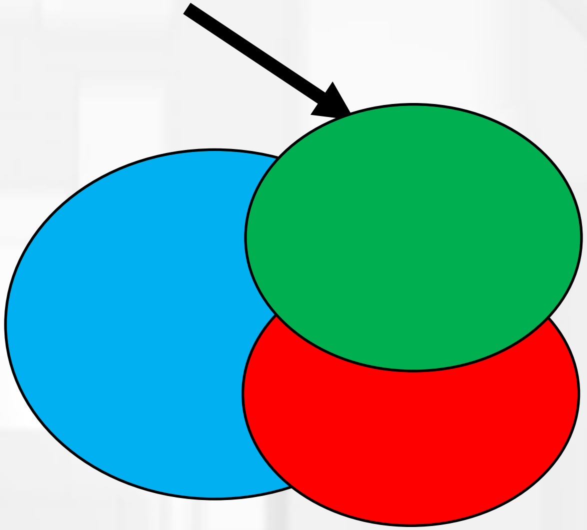
« Additional variance » : Again



No correlation between cyan, green and red



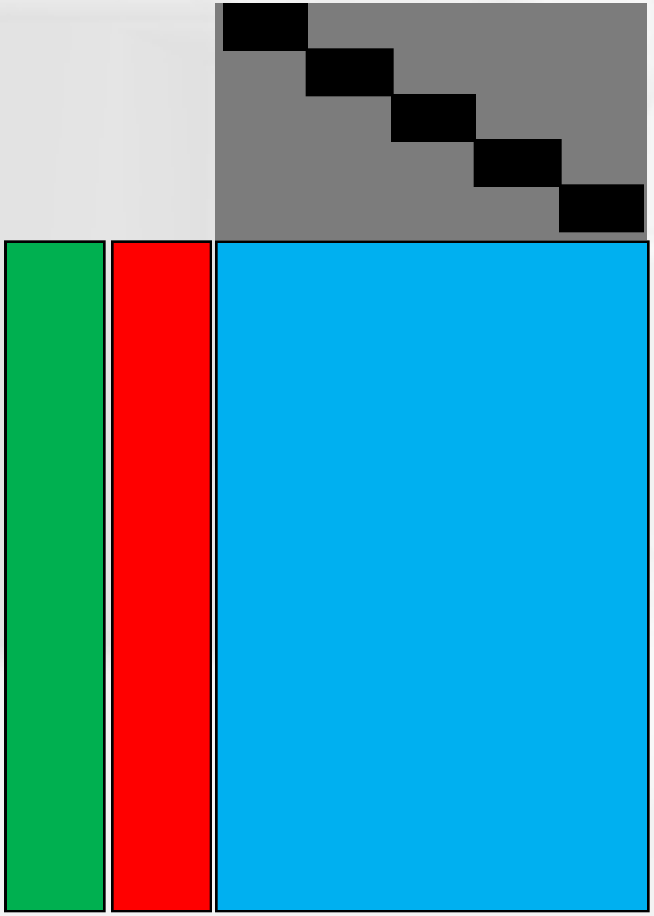
Testing for the green



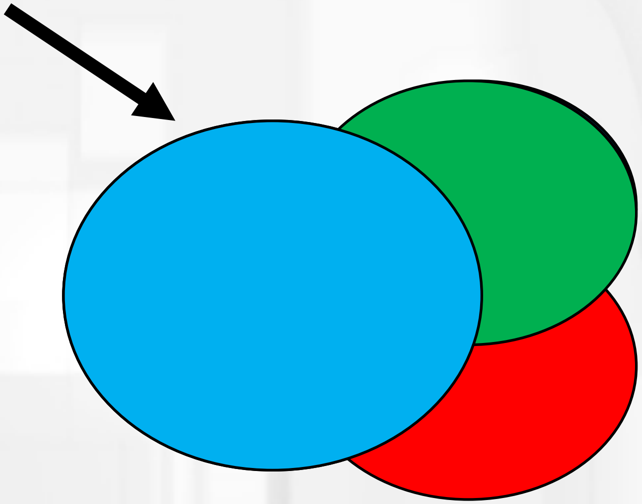
correlated regressors, for example

green: response to outcome

red: reward PE

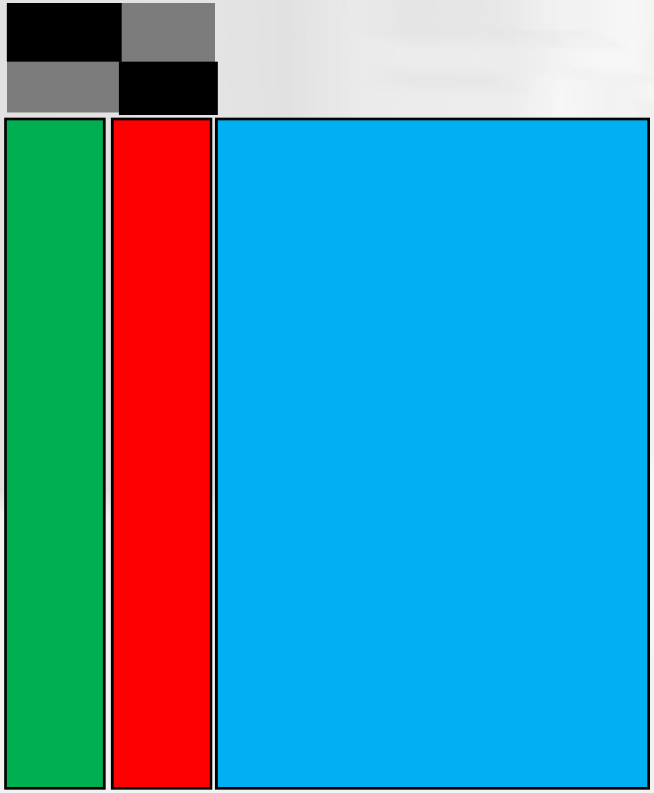


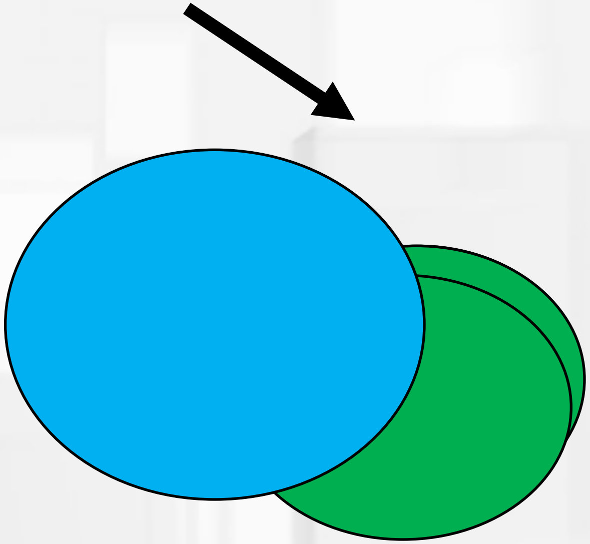
Testing for the cyan



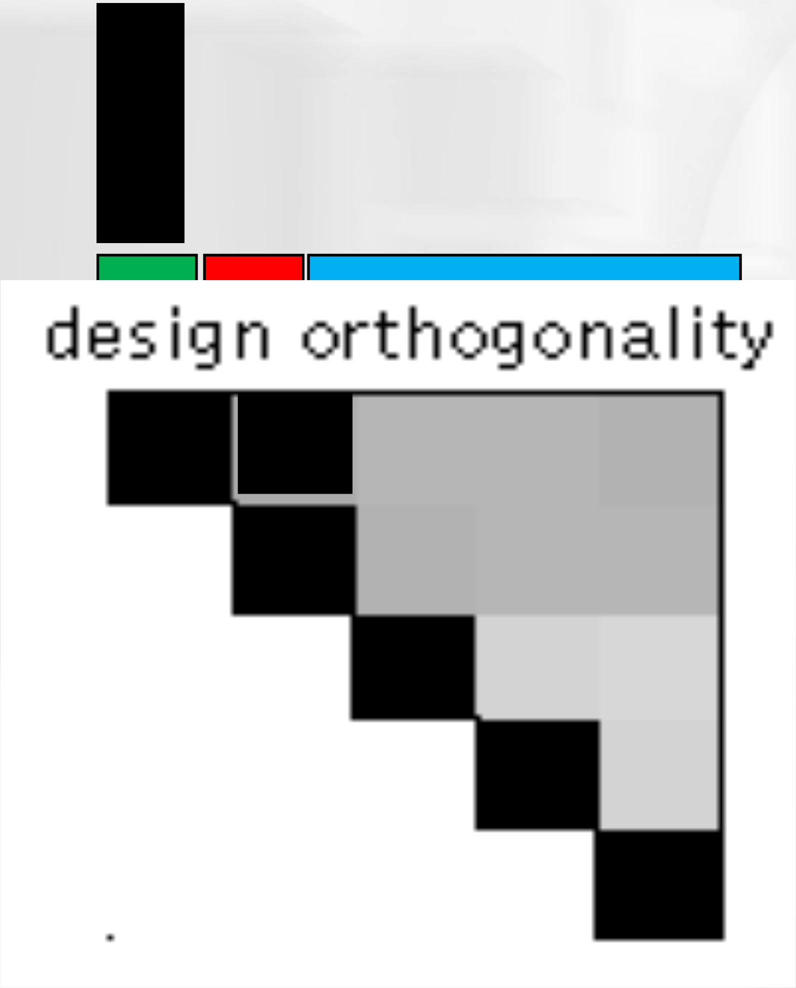
correlated regressors

Testing for the outcome and PE

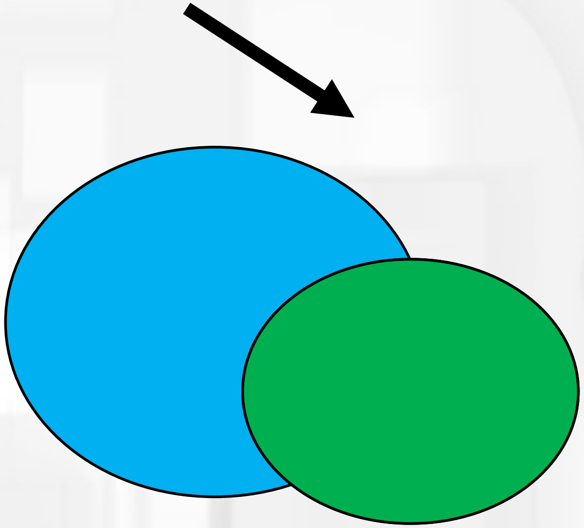




If significant ? Could be both!

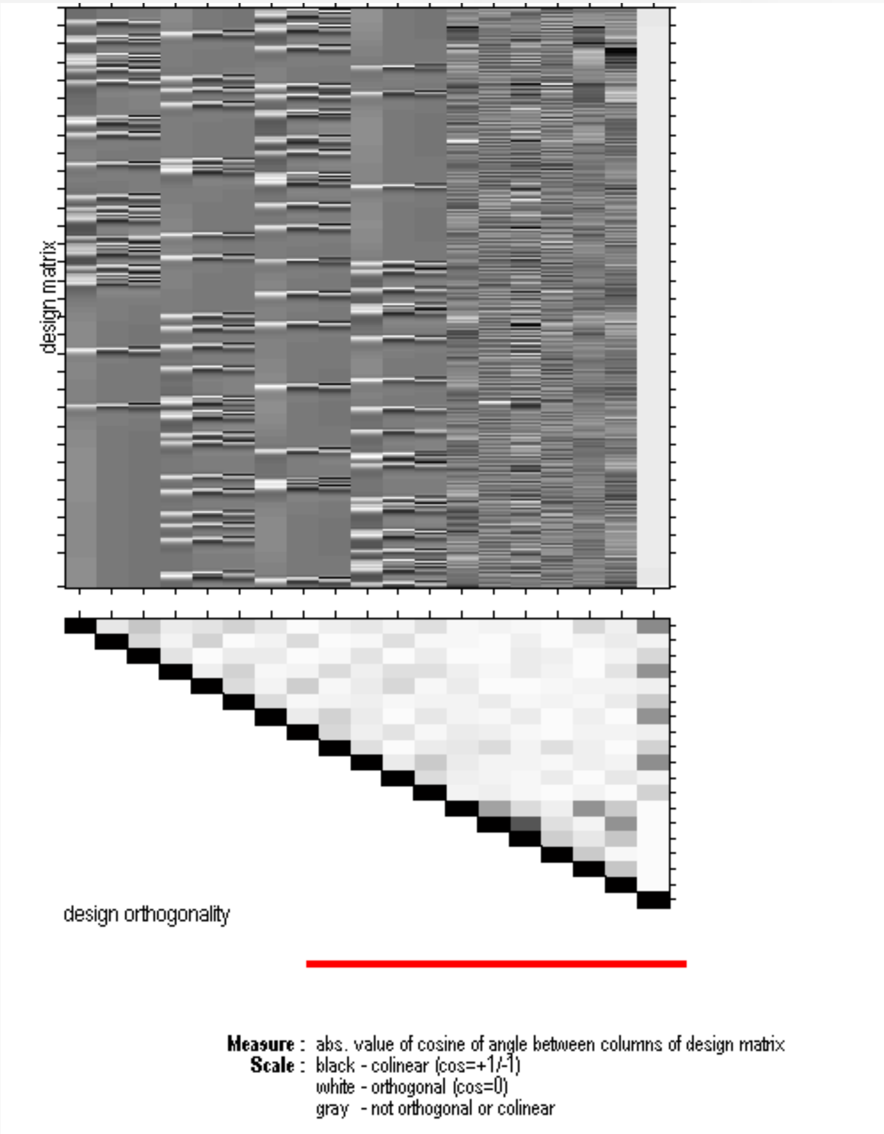


Testing for the outcome



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

Design orthogonality



Measure : abs. value of cosine of angle between columns of design matrix
Scale : black - colinear (cos=+1/-1)
white - orthogonal (cos=0)
gray - not orthogonal or colinear

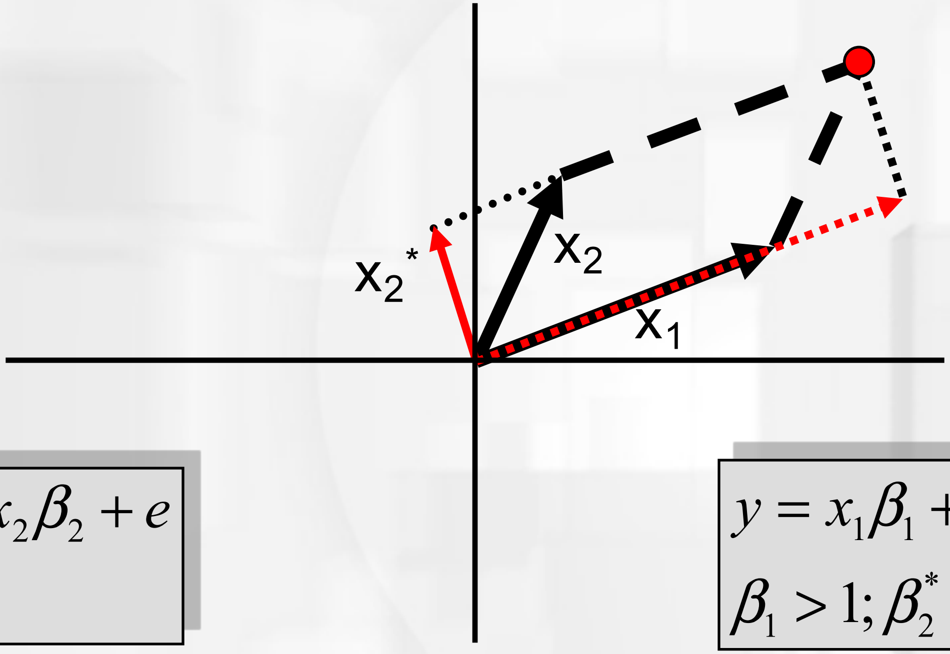
For each pair of columns of the design matrix, the orthogonality matrix depicts the magnitude of the **cosine of the angle** between them, with the range 0 to 1 mapped from white to black.

$$\cos\alpha = \frac{ab}{|a||b|}$$

If both vectors have **zero mean** then the cosine of the angle between the vectors is equivalent to the **correlation** between the two variates.

$$\cos\alpha = corr_{a,b}$$

Orthogonalization



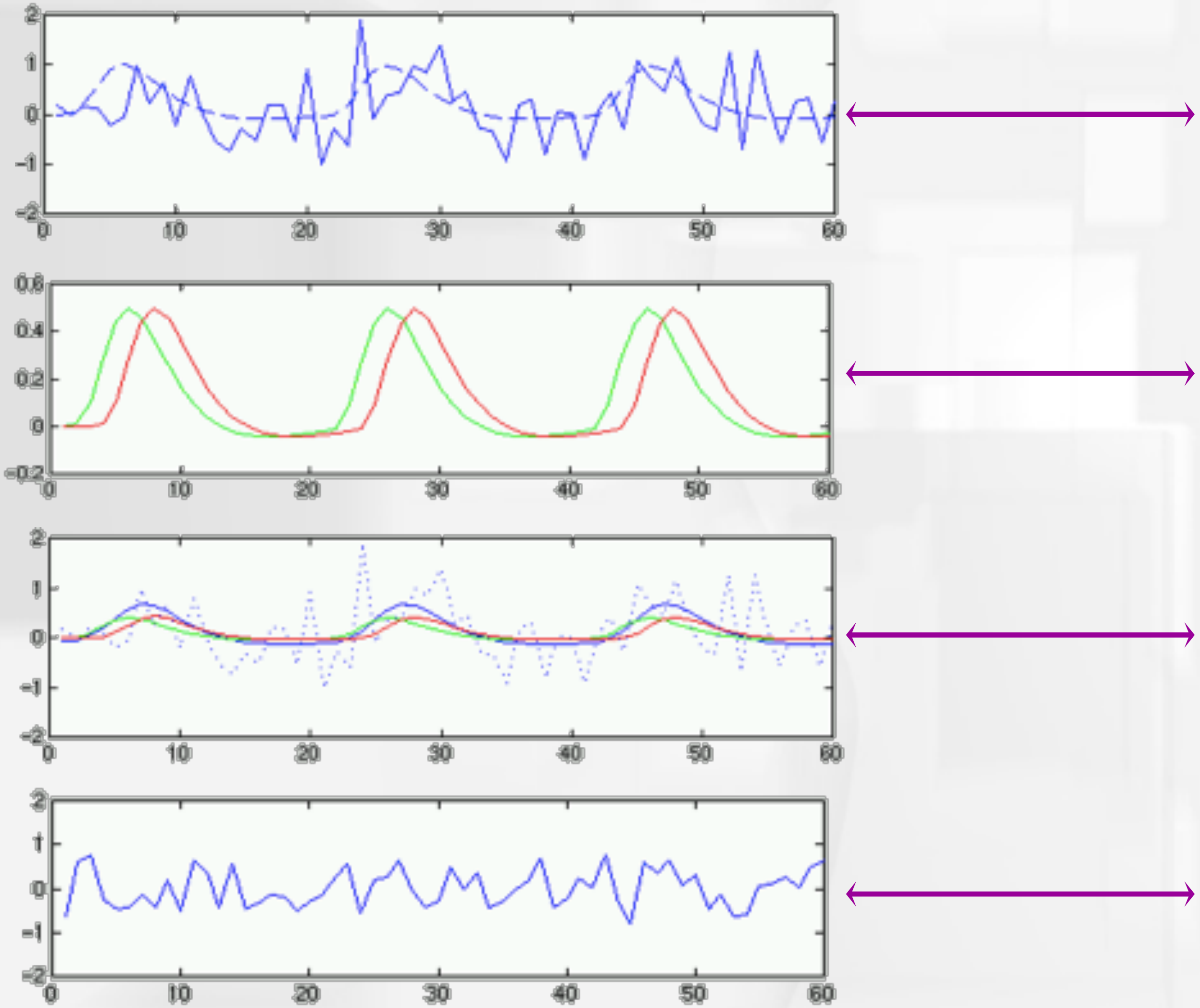
$$y = x_1\beta_1 + x_2\beta_2 + e$$
$$\beta_1 = \beta_2 = 1$$

$$y = x_1\beta_1 + x_2^*\beta_2^* + e$$
$$\beta_1 > 1; \beta_2^* = 1$$

Correlated regressors = explained variance is shared between regressors

When x_2 is orthogonalized with regard to x_1, only the parameter estimate for x_1 changes, not that for x_2!

Before Orthogonalization



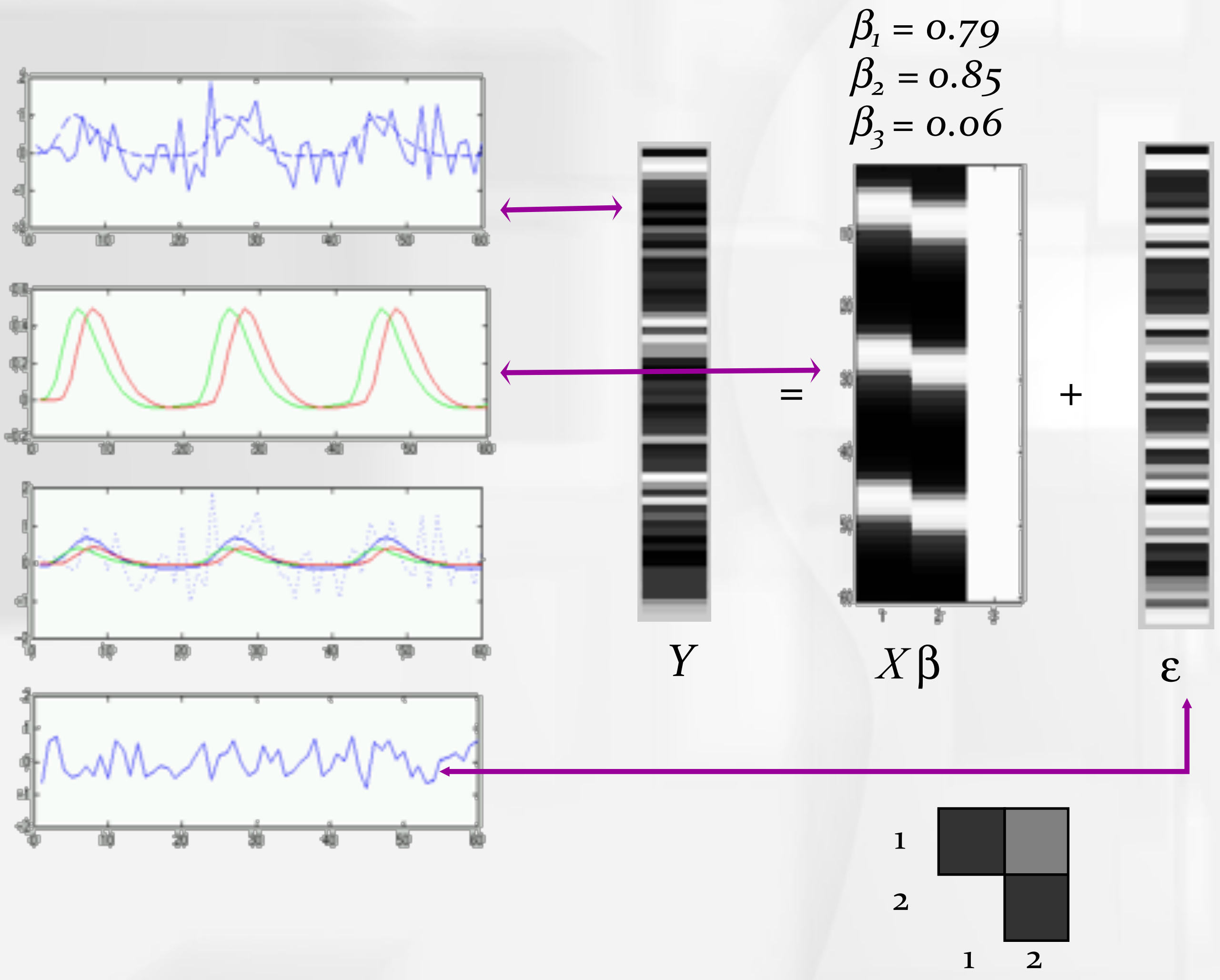
True signal

Model (green and red)

Fit (blue: total fit)

Residual

Before Orthogonalization



$\beta_1 = 0.79$

$\beta_2 = 0.85$

$\beta_3 = 0.06$

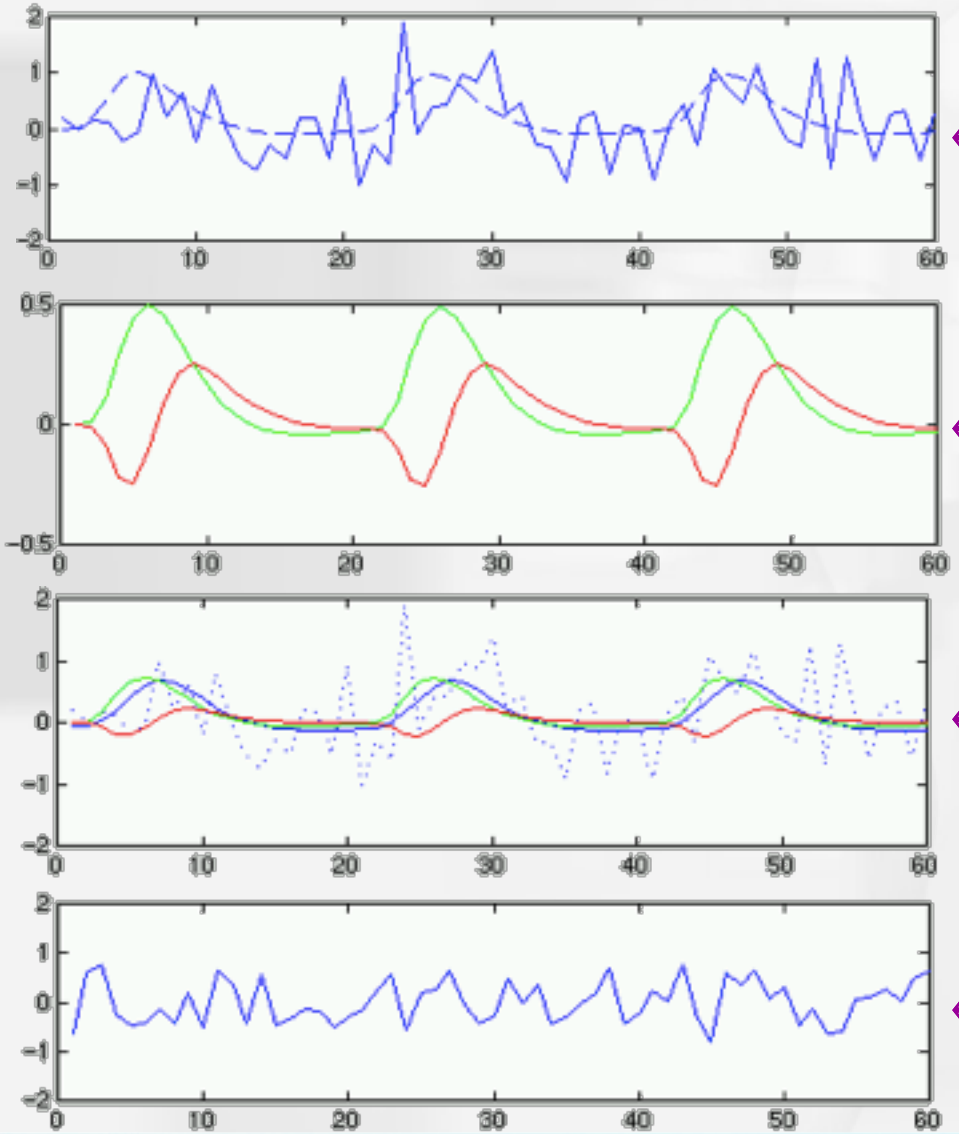
Residual var. = 0.3

$p(Y|\ \beta_1 = 0)$ ⇨
p-*value* = 0.08
(t-test)

$P(Y|\ \beta_2 = 0)$ ⇨
p-*value* = 0.07
(t-test)

$p(Y|\ \beta_1 = 0, \beta_2 = 0)$ ⇨
p-*value* = 0.002
(F-test)

After Orthogonalization



True signal

Model (green and red)
red regressor has been orthogonalised with respect to the green one
⇔ remove everything that correlates with the green regressor

Fit (does not change)

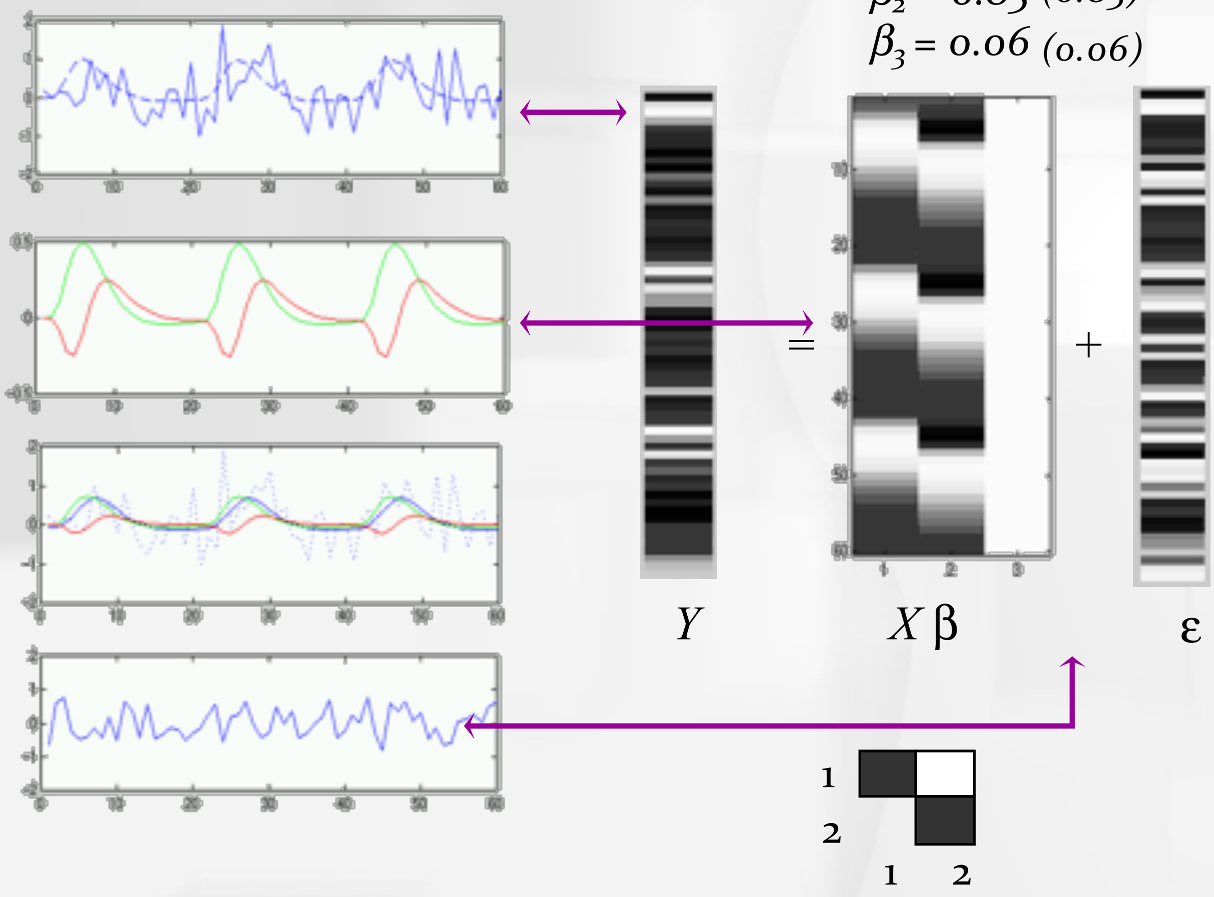
Residuals (do not change)

After Orthogonalization

$\beta_1 = 1.47 \quad (0.79)$

$\beta_2 = 0.85 \quad (0.85)$

$\beta_3 = 0.06 \quad (0.06)$



$Residual\ var. = 0.3$

$p(Y|\ \beta_1 = 0)$

$p\text{-}value = 0.0003$

(t-test)

does change

$p(Y|\ \beta_2 = 0)$

$p\text{-}value = 0.07$

(t-test)

does not change

$p(Y|\ b_1 = 0,\ b_2 = 0)$

$p\text{-}value = 0.002$

(F-test)

does not change

Multicollinearity: Take Home

- Orthogonalisation = decorrelation.
- *Parameters and test on the non-modified regressor change.*
- Rarely solves the problem as it requires assumptions about which regressor to uniquely attribute the common variance.
  ⇨ use F-tests to assess overall significance.
- Original regressors may not matter: it is the contrast you are testing which should be as decorrelated as possible from the rest of the design matrix – e.g. factorial designs are optimal
- For model-based fMRI analyses, this could be an issue (Discussion: Friday)
- Solution: design efficiency

Design efficiency

$$T = \frac{c^T\hat{\beta}}{\sqrt{\mathrm{var}(c^T\hat{\beta})}}$$

- The aim is to minimize the standard error of a *t*-contrast (i.e. the denominator of a t-statistic).

$$\mathrm{var}(c^T\hat{\beta}) = \hat{\sigma}^2 c^T (X^T X)^{-1} c$$

- This is equivalent to maximizing the efficiency *e*:

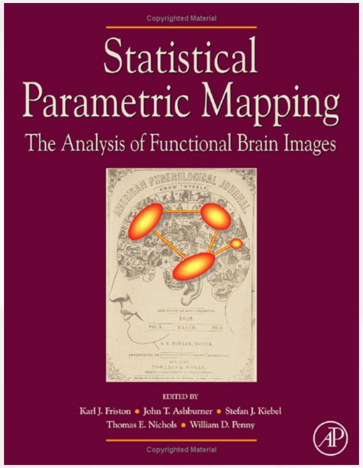
$$e(\hat{\sigma}^2, c, X) = (\underbrace{\boxed{\hat{\sigma}^2}}_{\text{Noise variance}} c^T \underbrace{\boxed{(X^T X)}}_{\text{Design variance}}^{-1} c)^{-1}$$

- If we assume that the noise variance is independent of the specific design:

$$e(c, X) = (c^T (X^T X)^{-1} c)^{-1}$$

- This is a relative measure: all we can really say is that one design is more efficient than another (for a given contrast).

Bibliography:



Copyrighted Material

Statistical Parametric Mapping

The Analysis of Functional Brain Images

EDITED BY

Karl J. Friston • John T. Ashburner • Stefan J. Kiebel

Thomas E. Nichols • William D. Penny

Copyrighted Material

Statistical Parametric Mapping: The Analysis of Functional Brain Images. Elsevier, 2007.

- *Plane Answers to Complex Questions: The Theory of Linear Models*. R. Christensen, Springer, 1996.
- *Statistical parametric maps in functional imaging: a general linear approach*. K.J. Friston et al, Human Brain Mapping, 1995.
- *Ambiguous results in functional neuroimaging data analysis due to covariate correlation*. A. Andrade et al., NeuroImage, 1999.
- *Estimating efficiency a priori: a comparison of blocked and randomized designs*. A. Mechelli et al., NeuroImage, 2003.