Translational Neuromodeling Unit

Classical Statistical Inference

Andreea O. Diaconescu

October 9th, 2015

Translational Neuromodeling Unit
Institute for Biomedical Engineering
University of Zurich and ETH Zurich

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

Overview

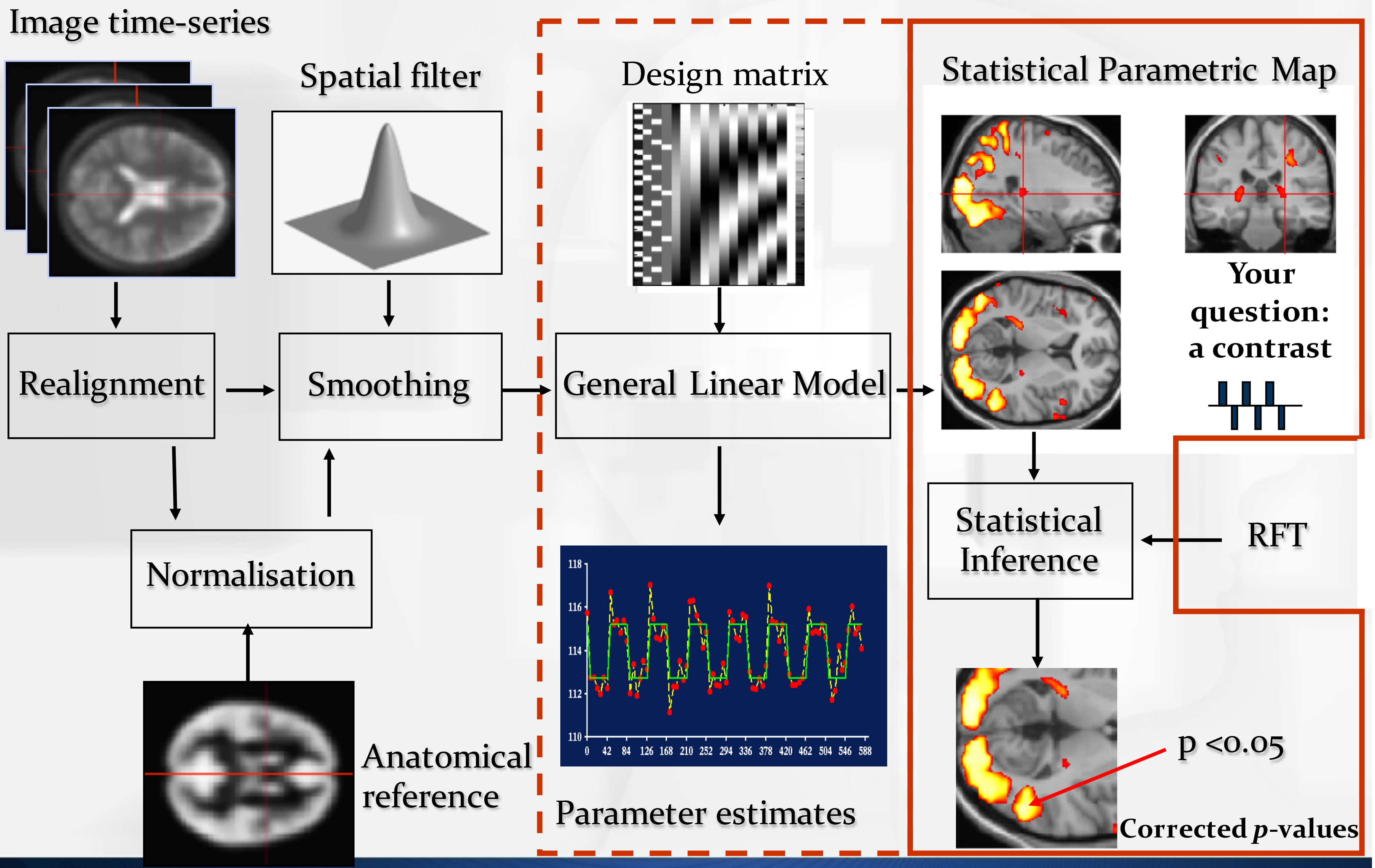


Image time-series
Spatial filter
Design matrix
Statistical Parametric Map
Your question: a contrast
Realignment
Smoothing
General Linear Model
Statistical Inference
RFT
Normalisation
Anatomical reference
Parameter estimates
p <0.05
Corrected p-values

Outline

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

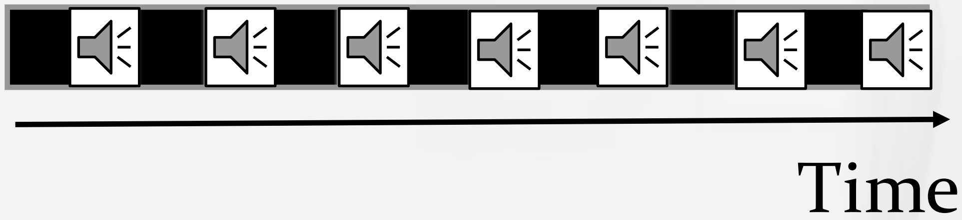
Outline

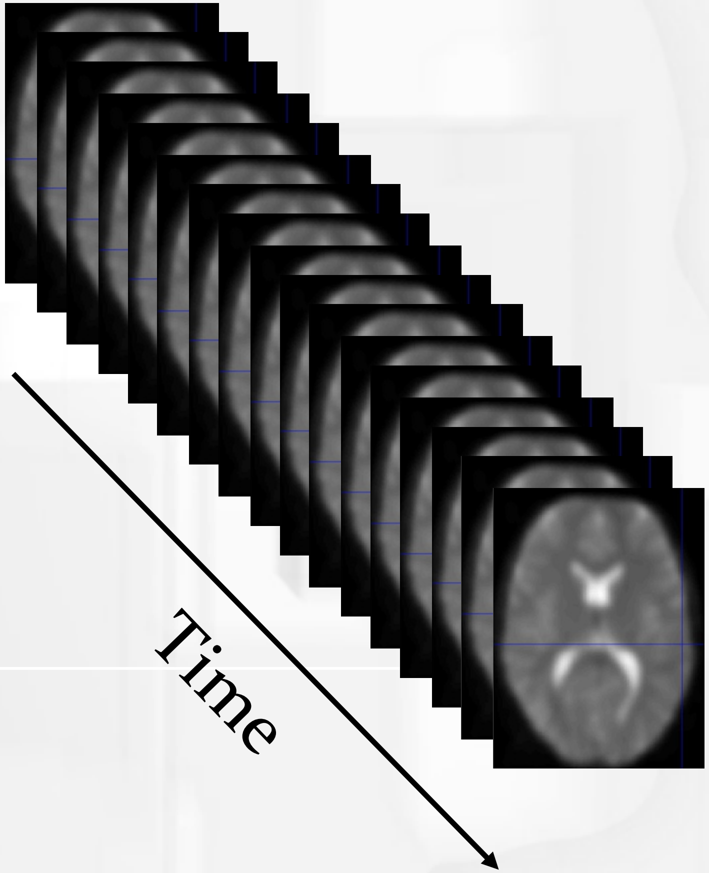
- Model and fit the data using the General Linear Model (GLM)
- T- and F-tests
  - What do they measure exactly?
- Multicollinearity

Research Question:

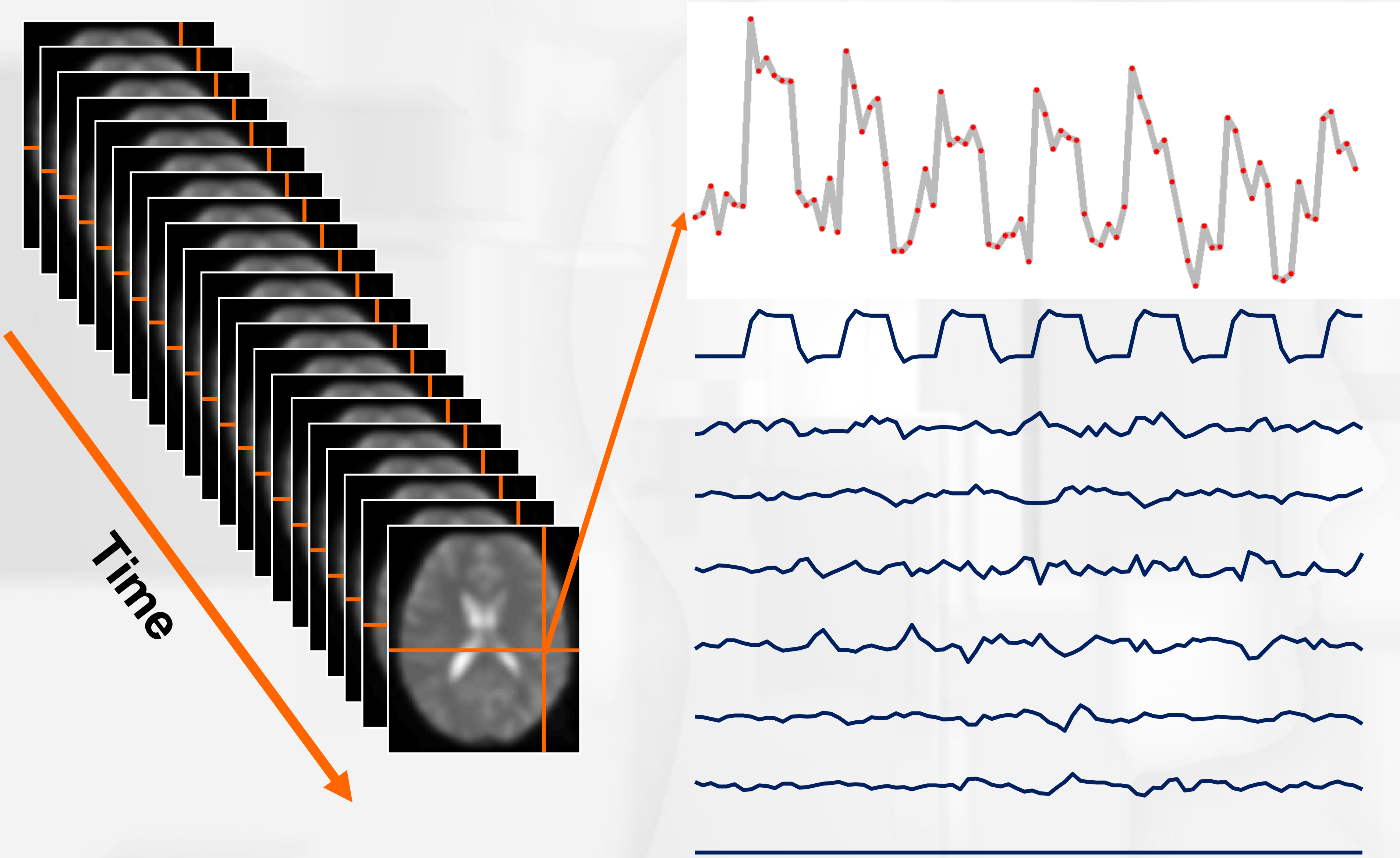
Where in the brain do we represent listening to sounds?

Imagine a very simple experiment…

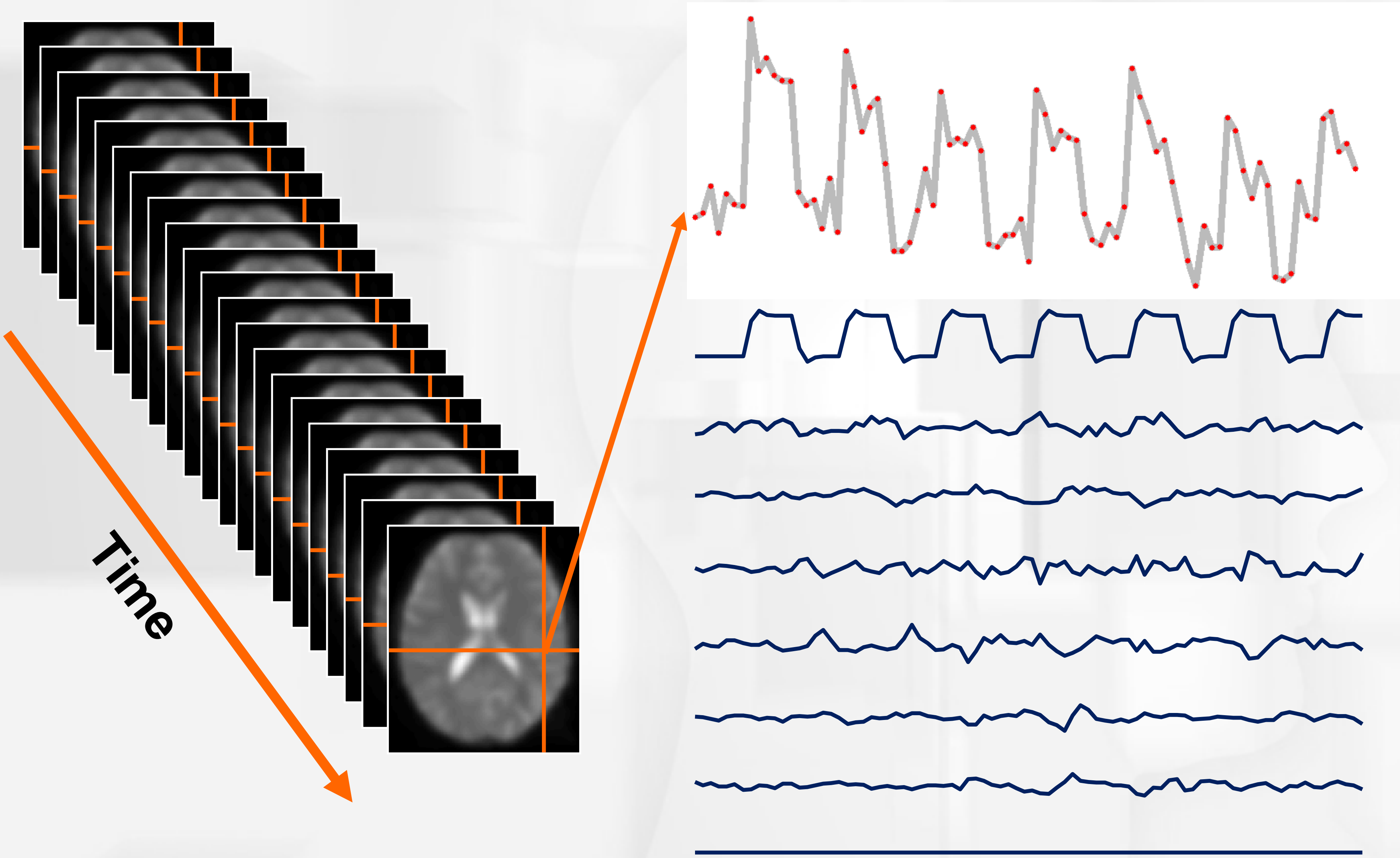




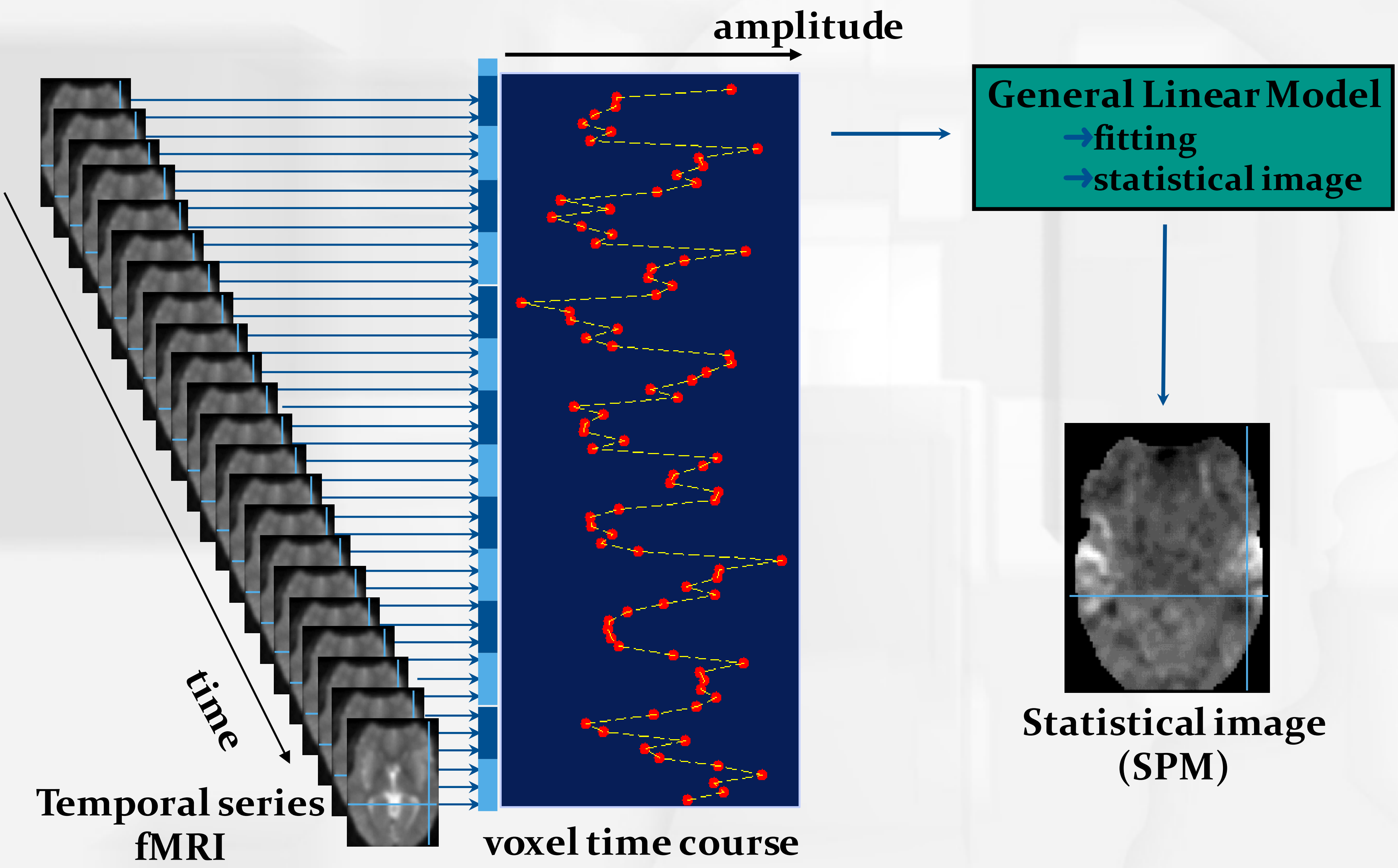
A mass-univariate approach



A mass-univariate approach

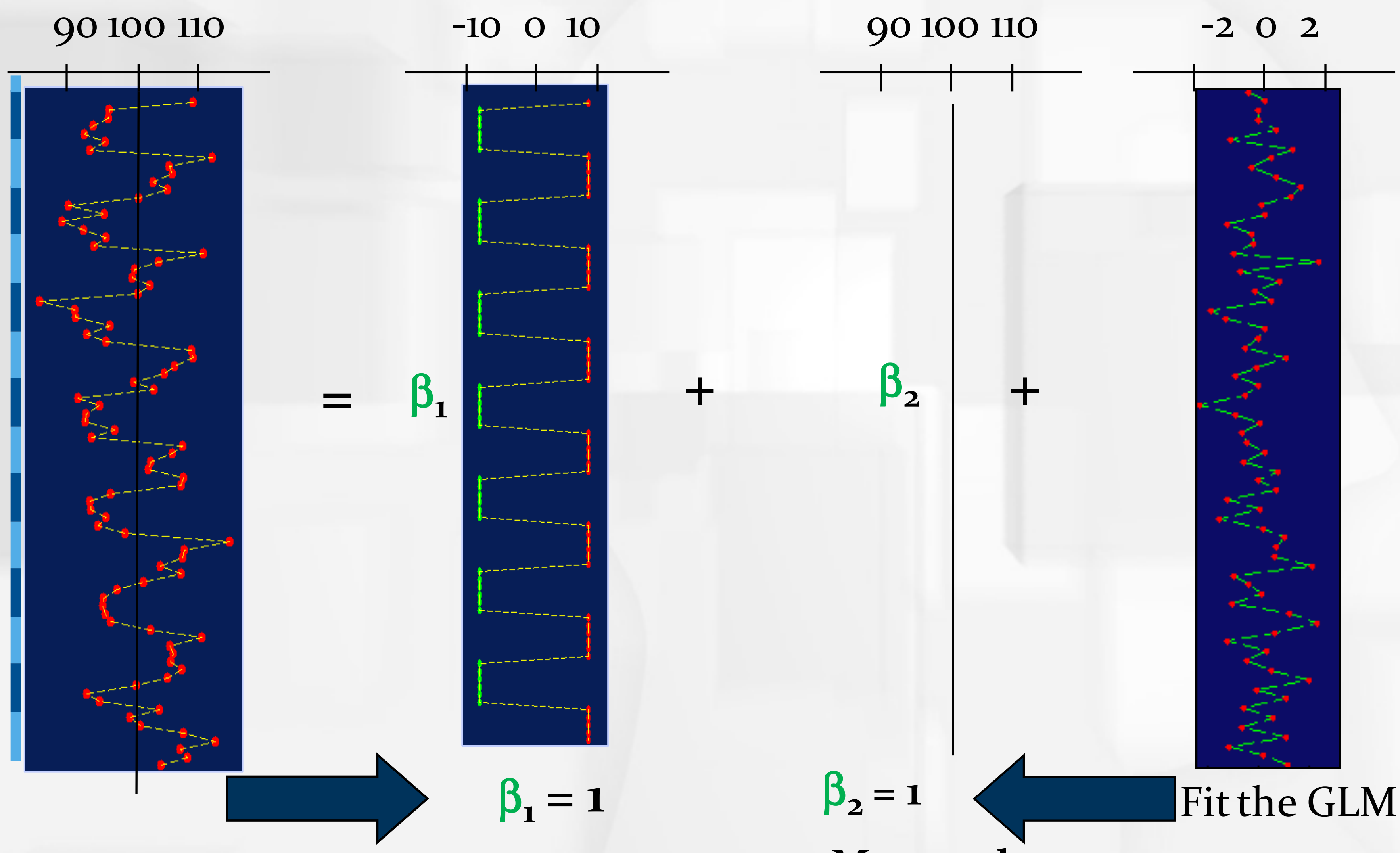


One voxel = One test (t, F, ...)



Regression example...

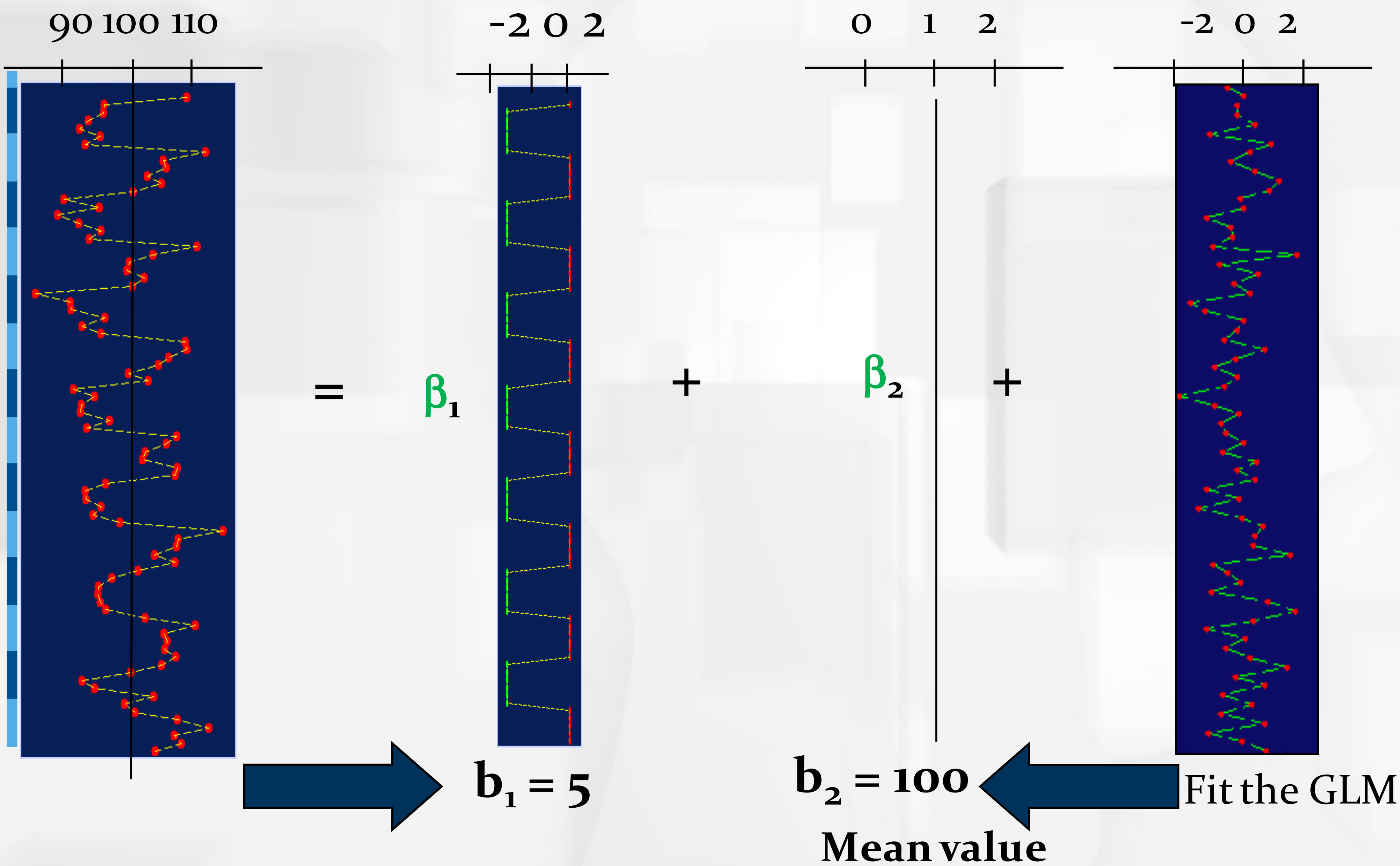
GLM | T-Test | F-Test | Multicollinearity



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

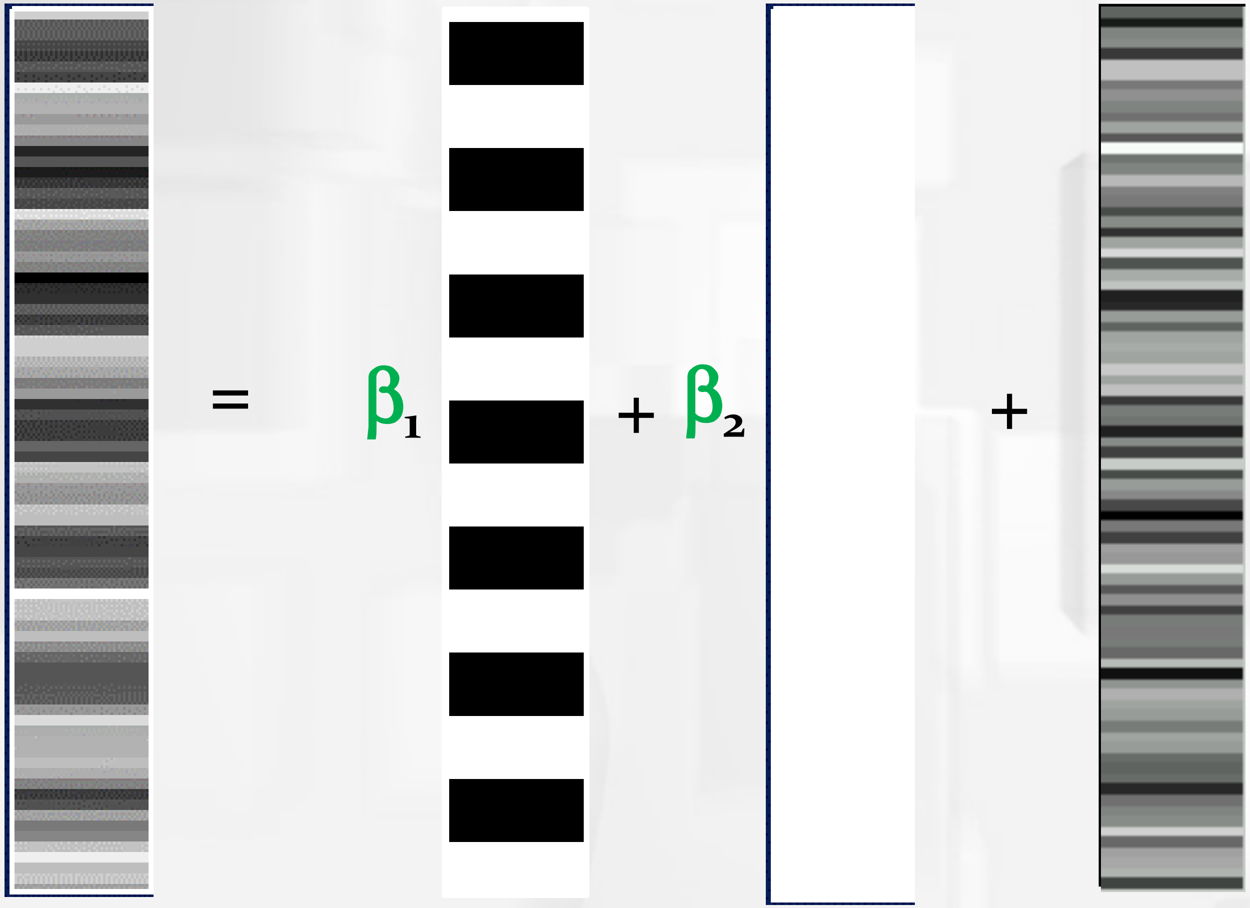
Regression example…

GLM | T-Test | F-Test | Multicollinearity



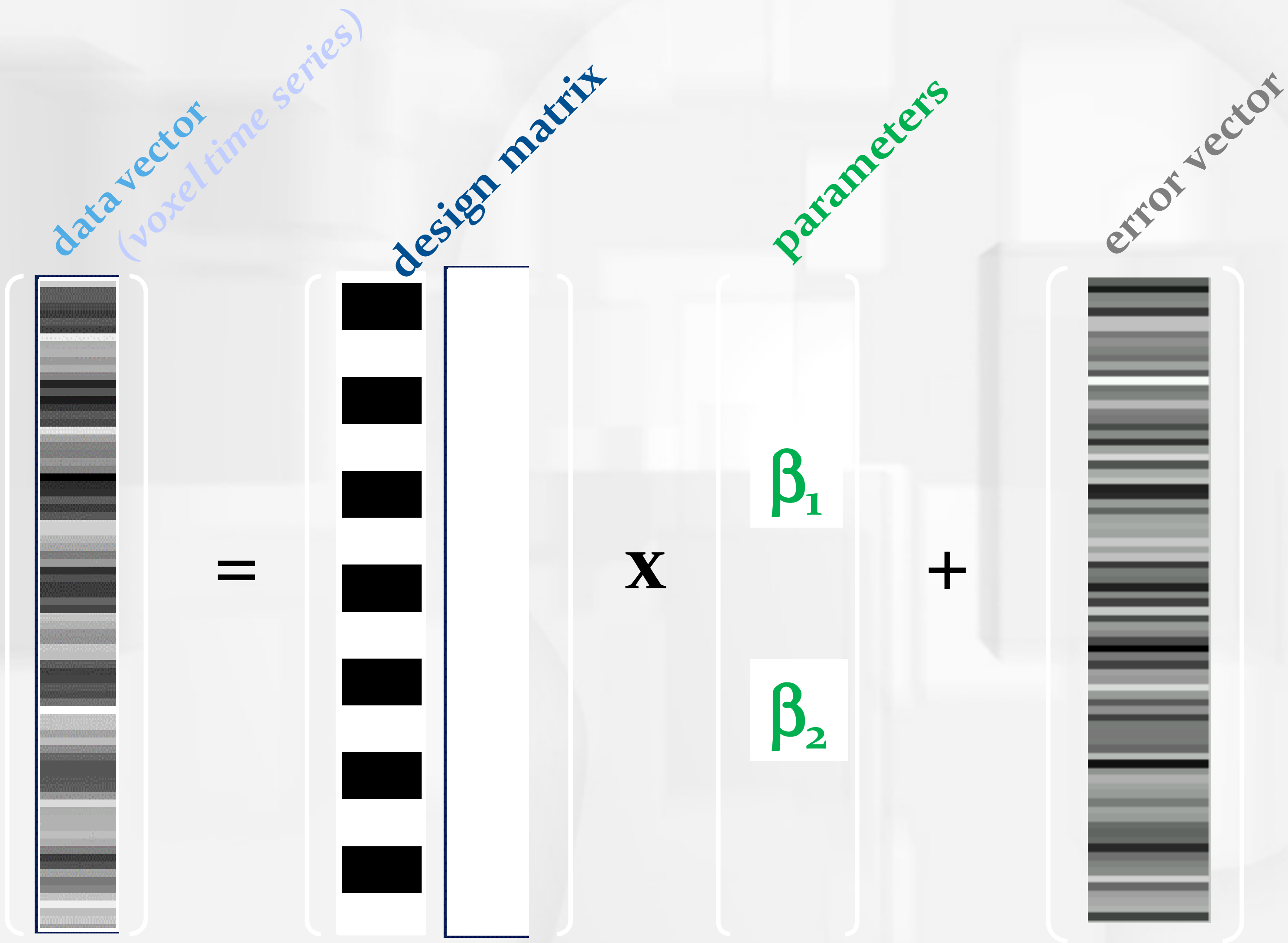
voxel time series **box-car reference function**

...revisited : matrix form



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

Box car regression: design matrix…



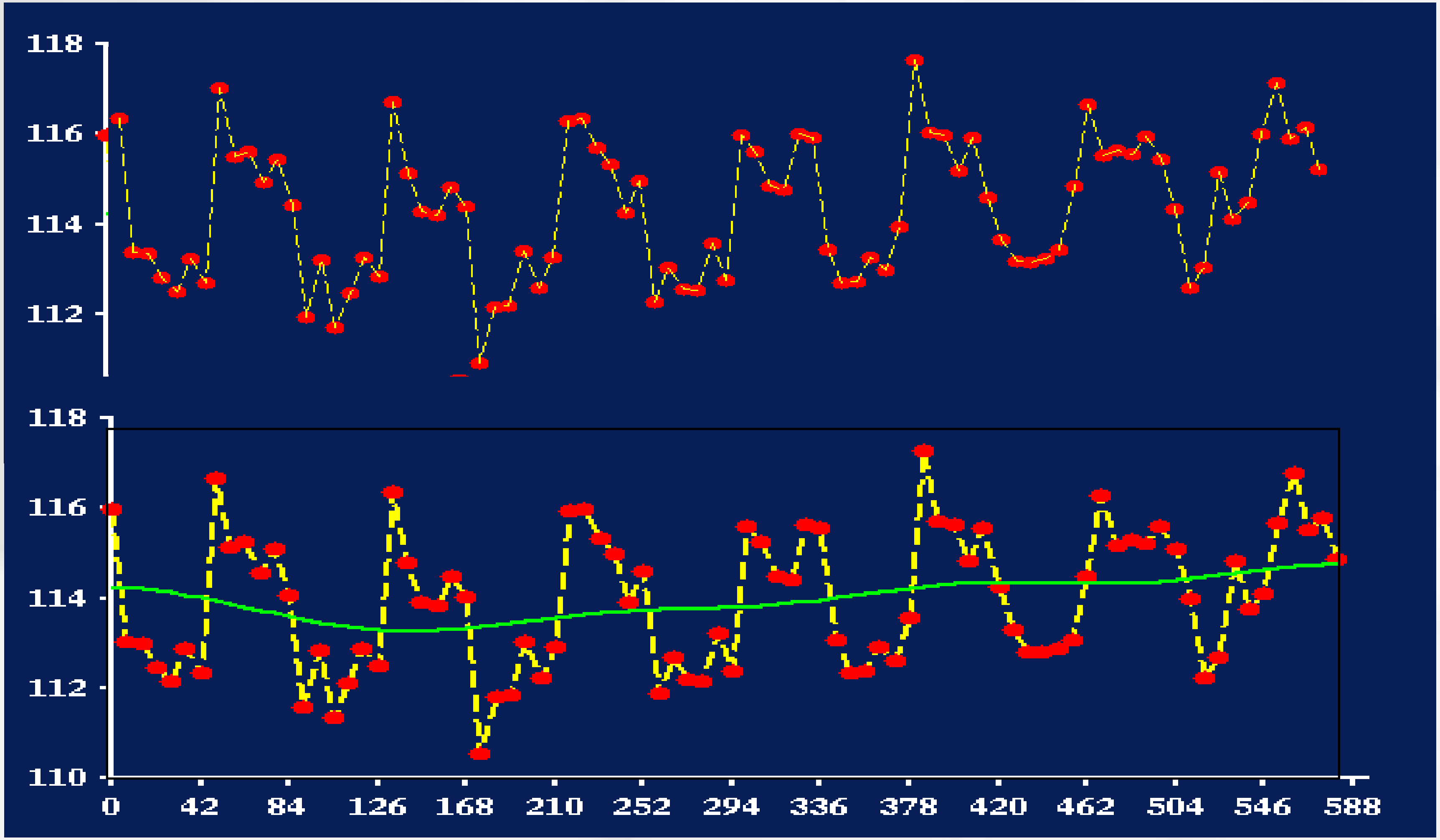
$$\underline{Y} = X \times \underline{\beta} + \underline{\varepsilon}$$

Fact: model parameters depend on regressors scaling

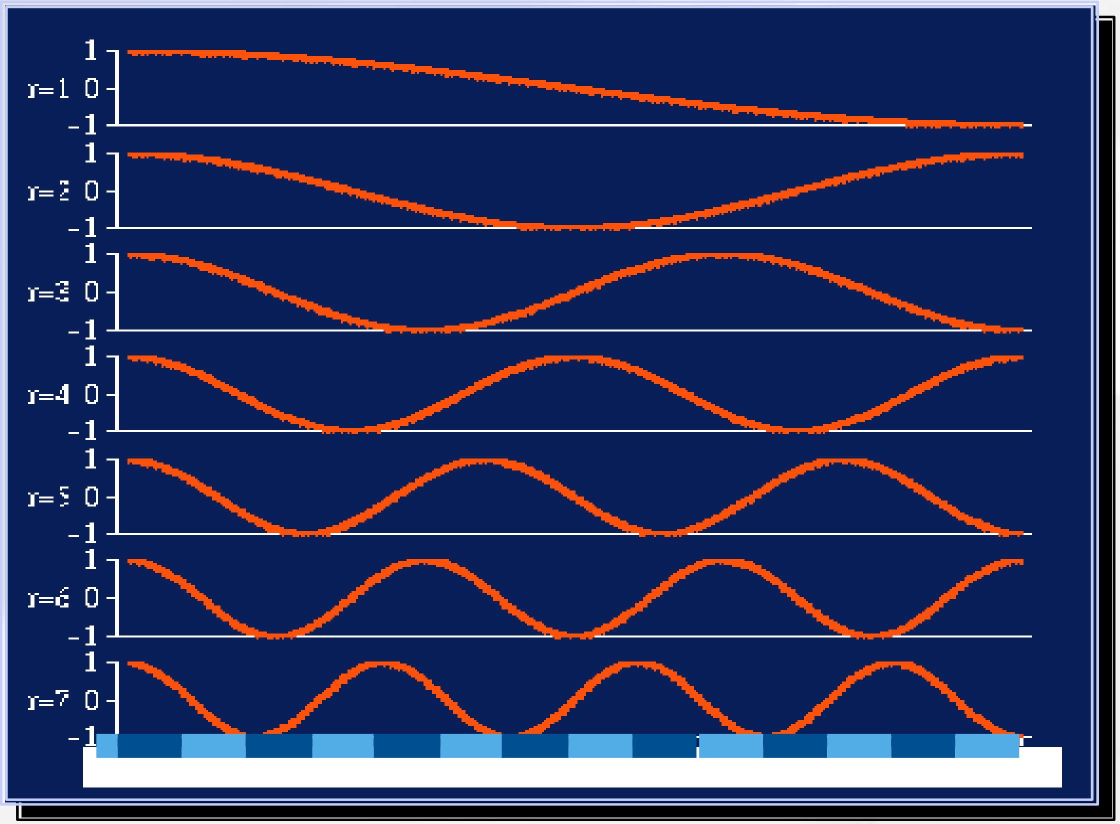
ONLY when comparing manually entered regressors (e.g., if you would like to compare two scores)

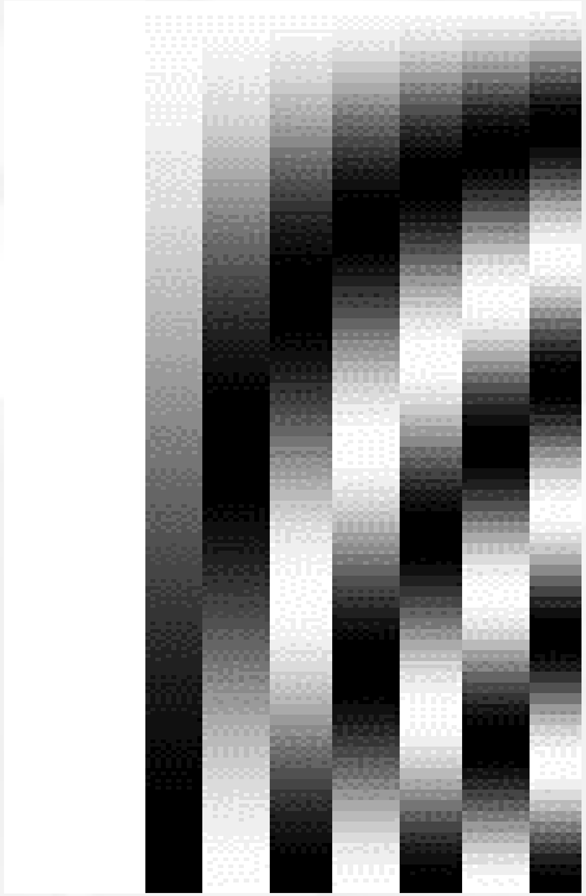
careful when comparing two (or more) manually entered regressors!

What if we believe that there are drifts?



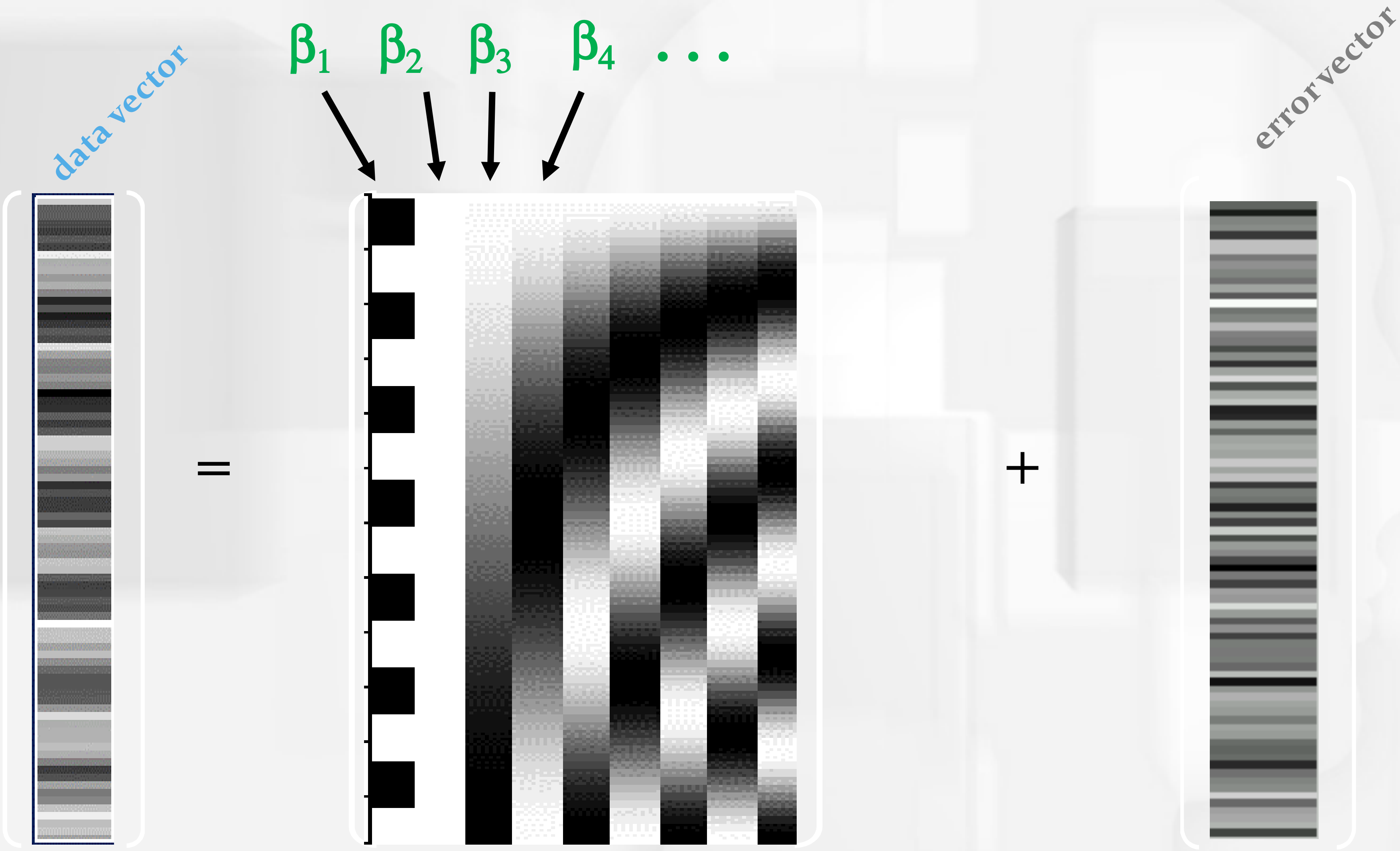
Add more reference functions / covariates ...





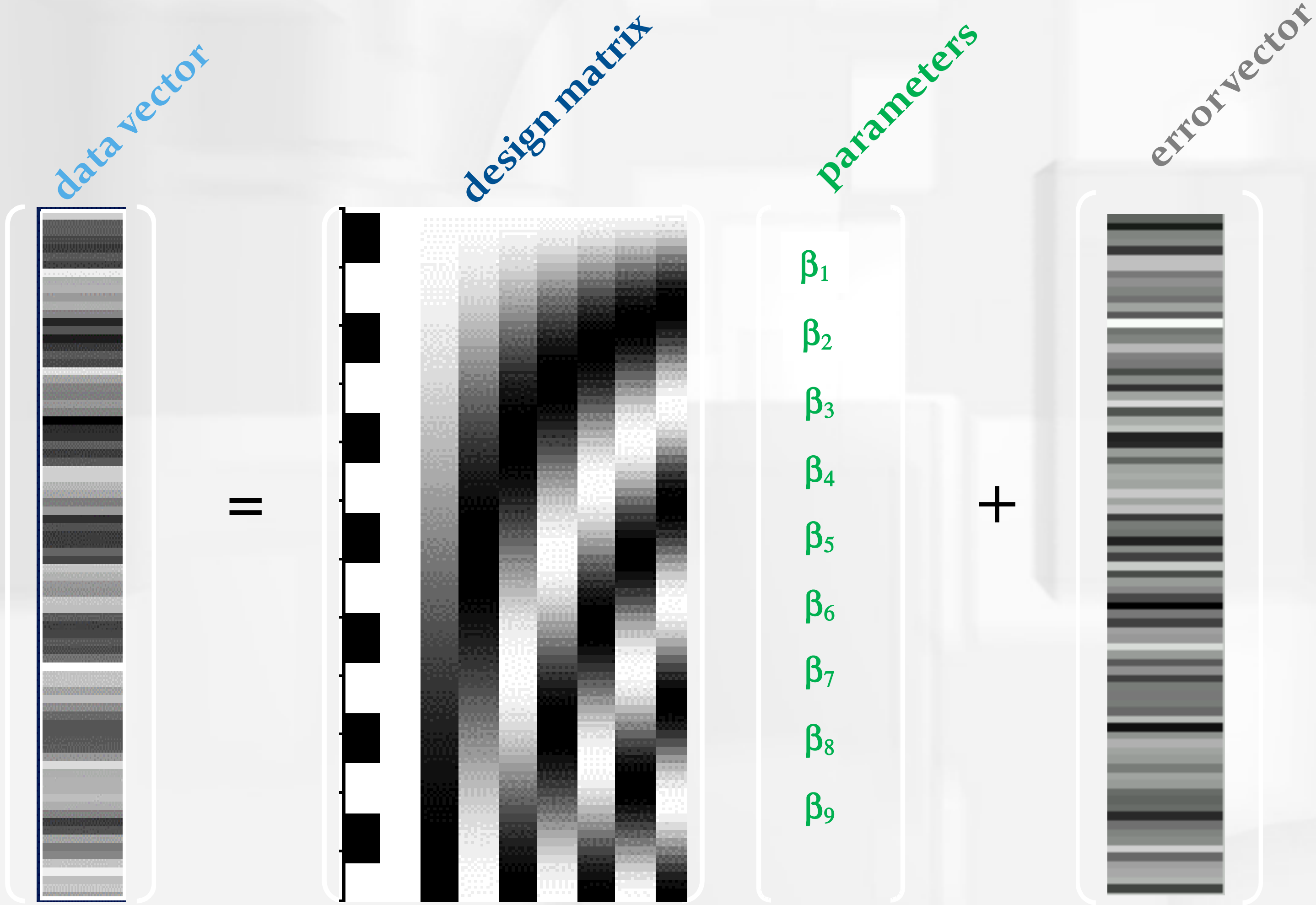
Discrete cosine transform basis functions

...design matrix



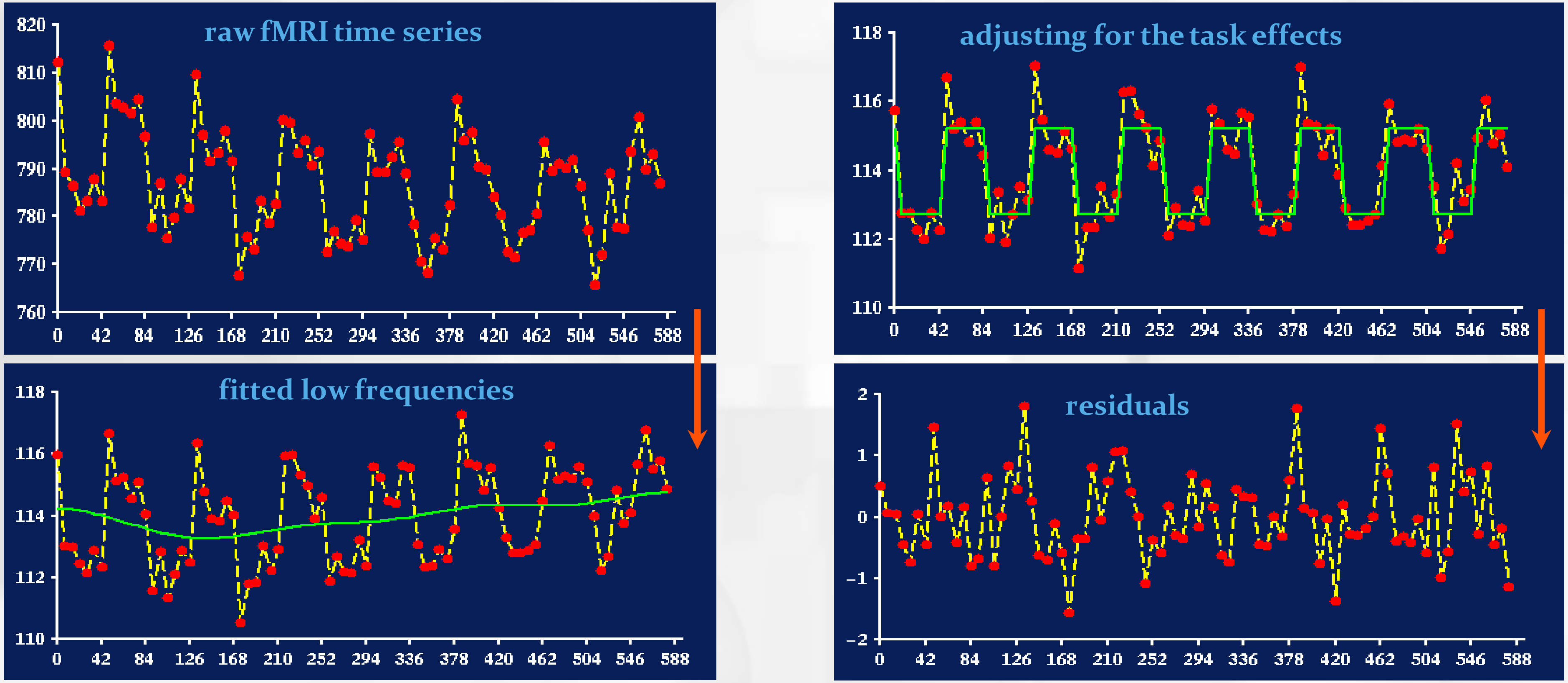
$$Y = X \times \beta + \varepsilon$$

...design matrix



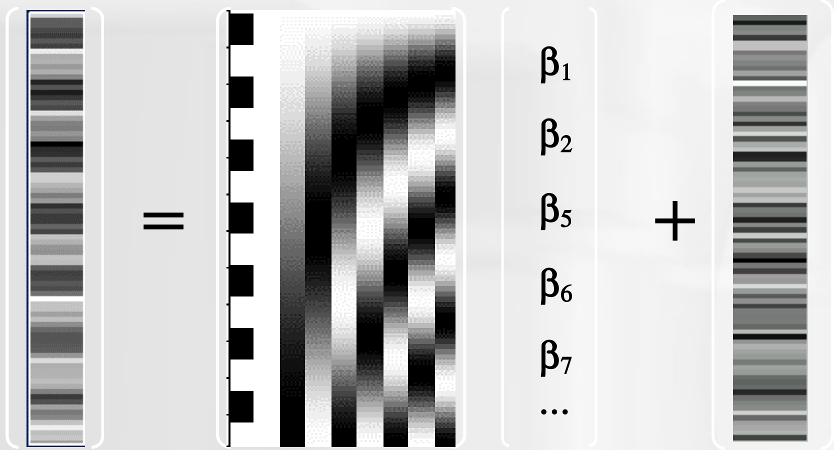
$$Y = X \times \beta + \varepsilon$$

Fitting the model = finding some estimate of the betas



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

Fitting the model = finding some estimate of the betas



Y = **X** x β + ε

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

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

$$\left\| y_i - x_i^T \beta \right\|^2 = \sum_i \left[y_i - x_i^T \beta \right]^2 = (y - X\beta)^T (y - x\beta)$$

when β are estimated: $\hat{\beta}$

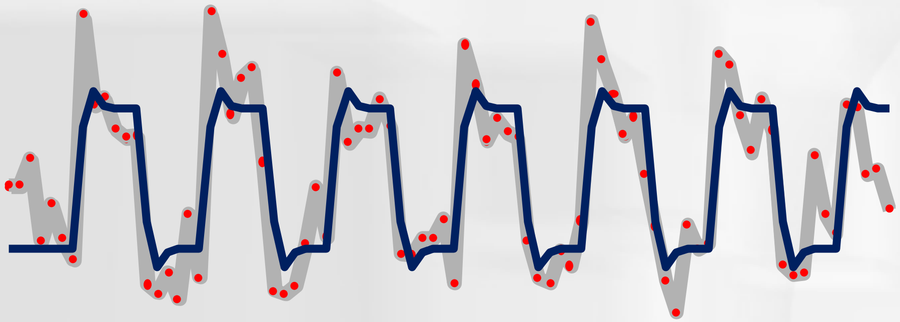
when ε is estimated: $\hat{\epsilon}$

estimated SD of ε: $\hat{\sigma}$

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

To summarize

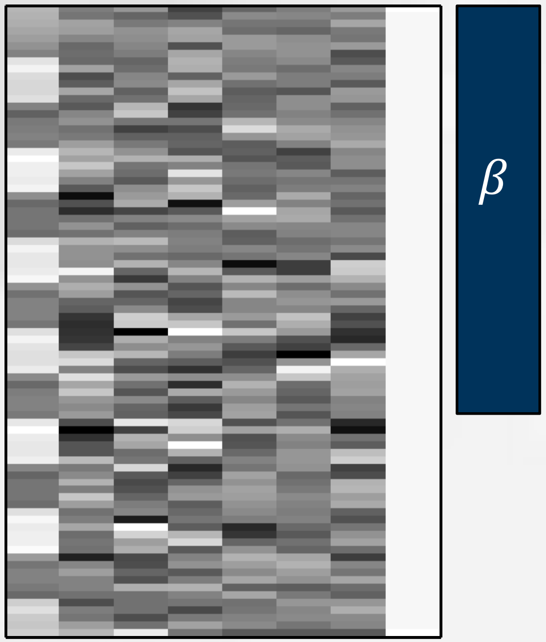
GLM | T-Test | F-Test | Multicollinearity



i.i.d. assumptions: $\varepsilon \sim N(0, \sigma^2 I)$

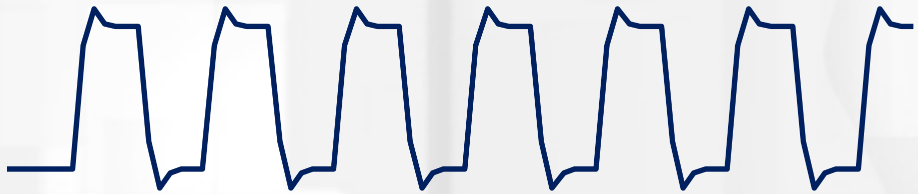
OLS estimates: $\hat{\beta} = (X^T X)^{-1} X^T y$

$y =$

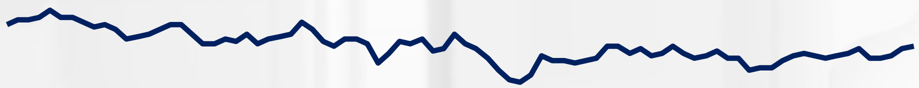


$+\varepsilon$

$\hat{\beta}_1 = 3.9831$



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



$\hat{\beta}_8 = 131.0040$

$\hat{\varepsilon} =$

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

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

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

- T-test
- F-Test

To test an hypothesis, we construct a "test statistic".

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

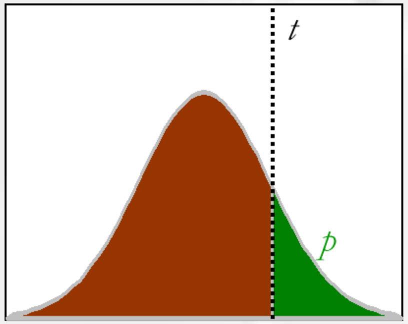
  This is what we want to disprove.

  ⇨ The "alternative hypothesis" H_1 represents the outcome of interest.

- A *p*-value summarises evidence against H_0.
  This is the probability of observing t, or a more extreme value, under the null hypothesis:

$$p(T \geq t \mid H_0)$$

- The conclusion about the hypothesis:
  We reject H_0 in favour of H_1 if $t > u_a$



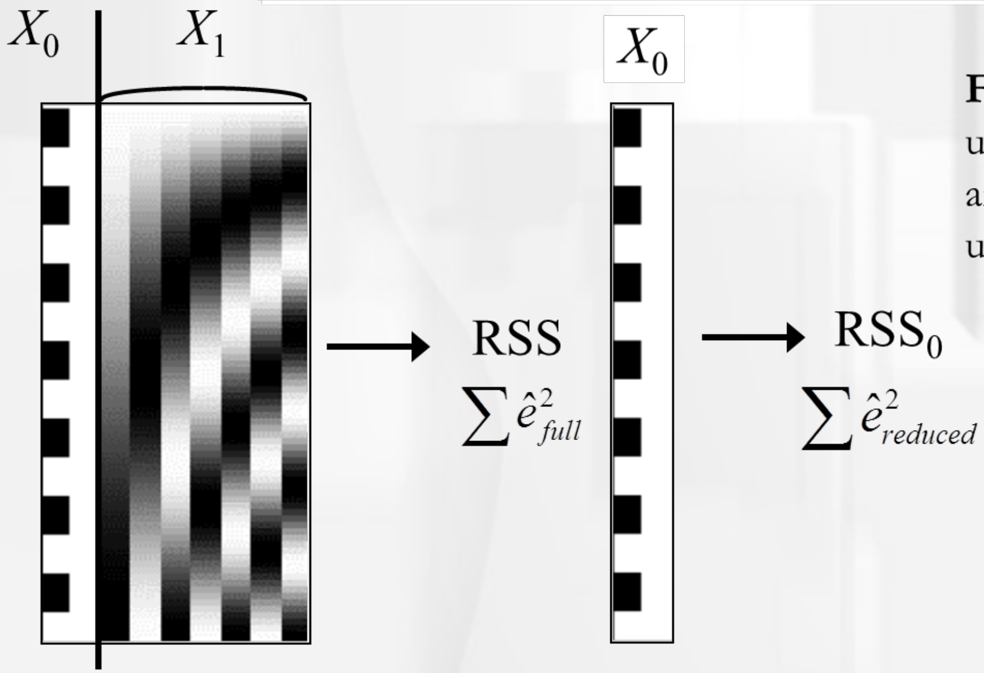
Null Distribution of T

Statistical Inference

- T-test
- F-Test

Model comparison: Full vs. reduced model

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



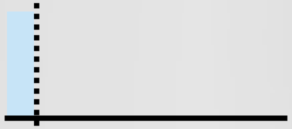
Full model ($X_0 + X_1$)? Or reduced model (X_0)?

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

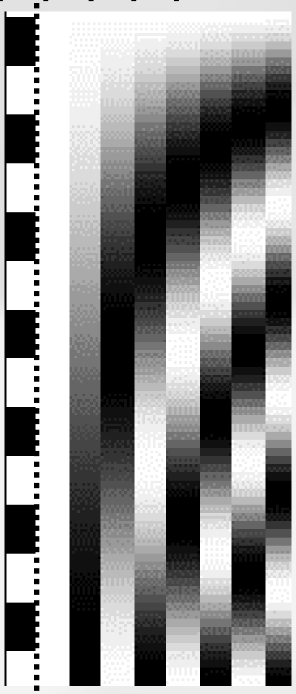
$$F \propto \frac{RSS_0 - RSS}{RSS}$$

T-test: one dimensional contrast SPM {t}

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



$\beta_1\ \beta_2\ \beta_3\ \beta_4\ \beta_5\ ....$



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

$\beta_1 > 0$?

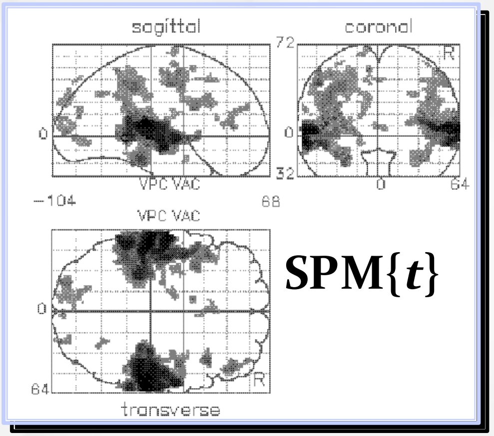
Compute $1x\beta_1 + 0x\beta_2 + 0x\beta_3 + 0x\beta_4 + 0x\beta_5 + \ldots = c^T\beta$

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

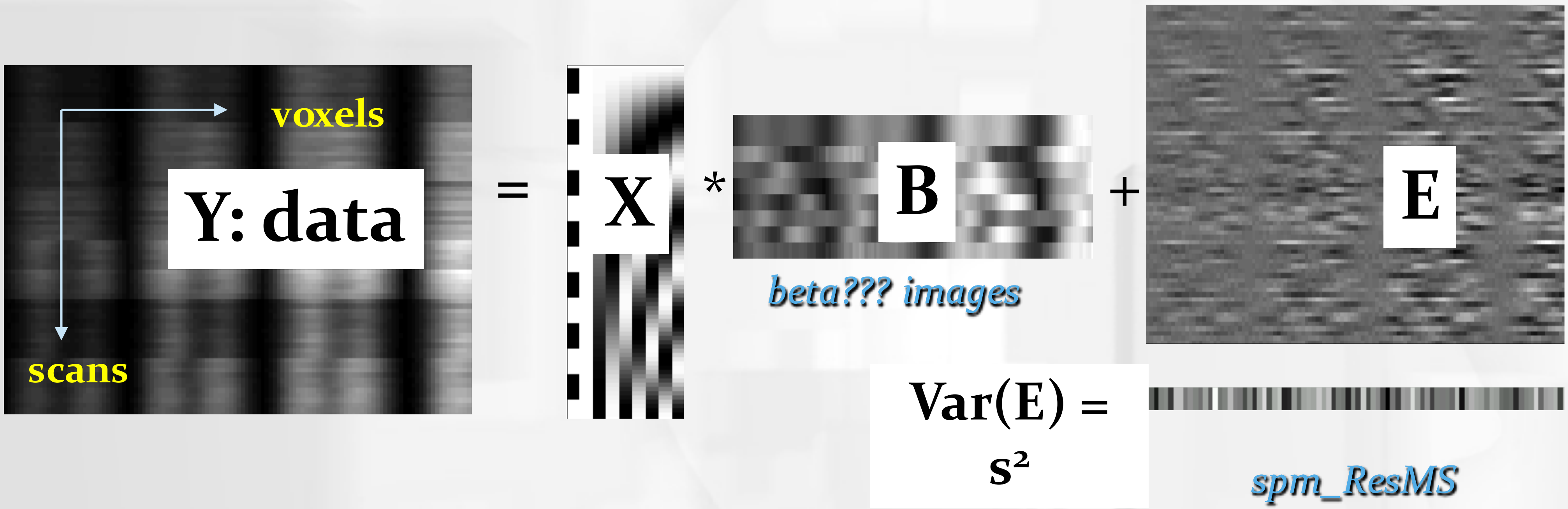
divide by estimated standard deviation of β_1

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

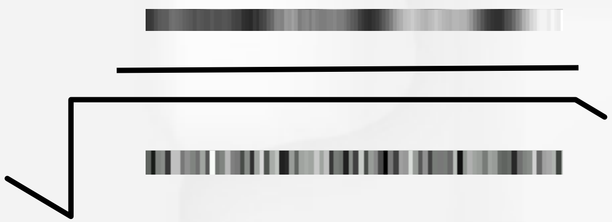
$$T = \frac{c^T\hat{\beta}}{\sqrt{\hat{\sigma}c^T(X^TX)^{-1}c}}$$



From one time series to an image



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

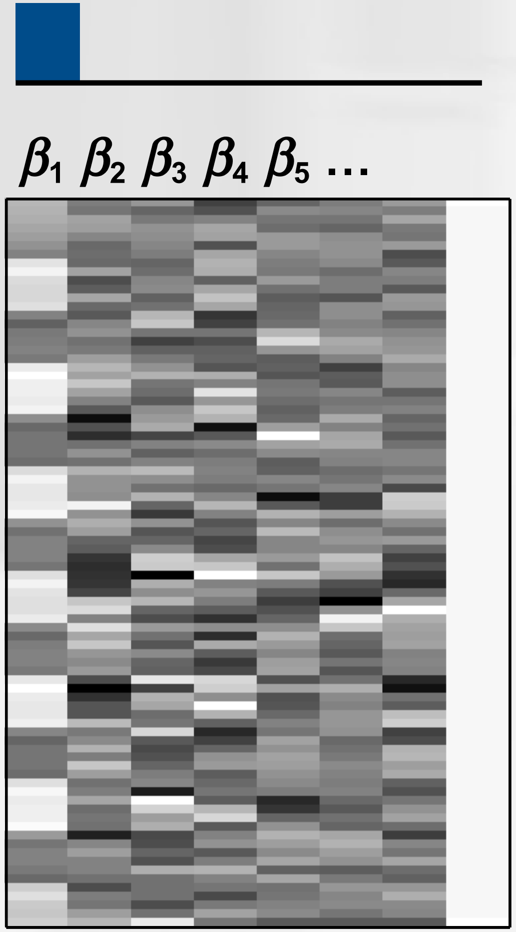


spm_con??? images

spm_t??? images

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

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



Question:

effect of interest > 0 ?

=

amplitude > 0 ?

=

$\beta_1 = c^T\beta > 0$?

Null hypothesis:

$H_0: c^T\beta = 0$

Test statistic:

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

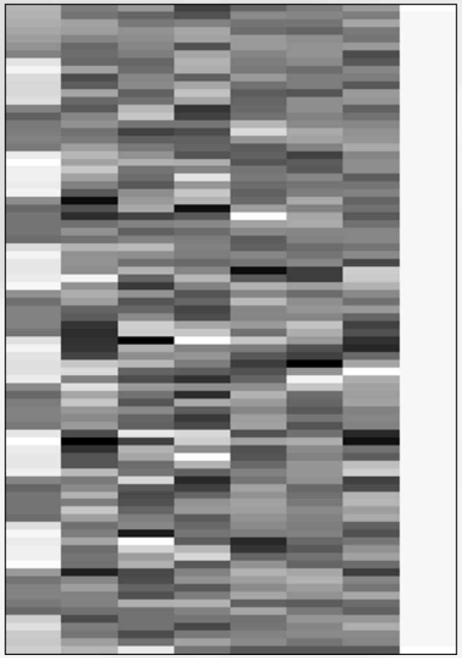
$$T = \frac{c^T\hat{\beta}}{\sqrt{\mathrm{var}(c^T\hat{\beta})}} = \frac{c^T\hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}} \sim t_{N-p}$$

T-test: a simple example

- Passive word listening versus rest

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

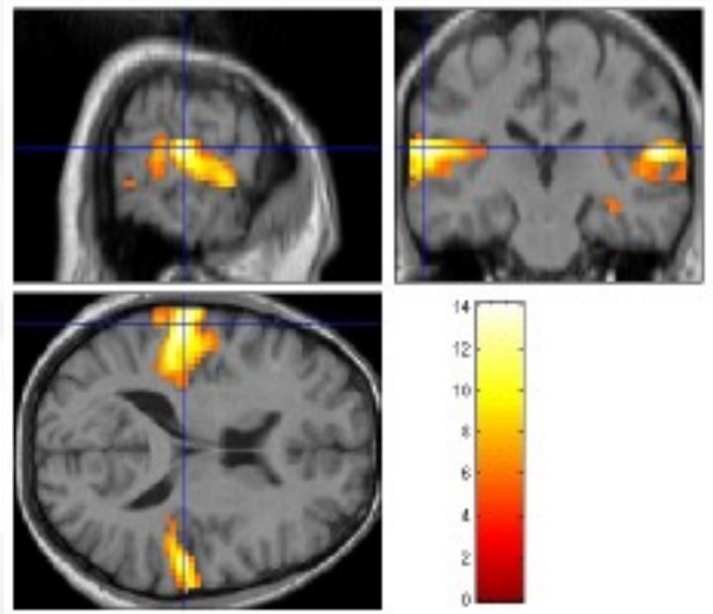
$\beta_1\ \beta_2\ \beta_3\ \beta_4\ \beta_5\ \ldots$



Q: activation during listening ?

Null hypothesis: $b_1 = 0$

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

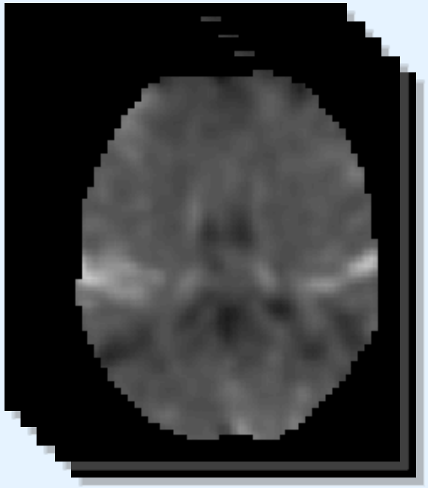


SPMresults: Threshold T = 3.2057 {p<0.001}
voxel-level

	$(Z_{\equiv})$	$p_{\text{uncorrected}}$	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**

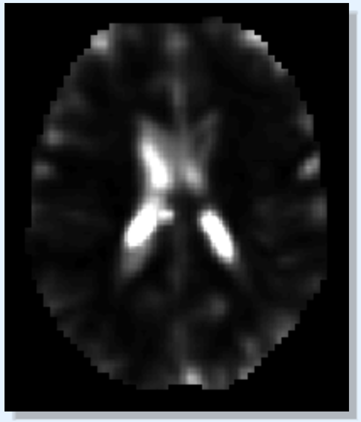
T-contrast in SPM

- For a given contrast c:



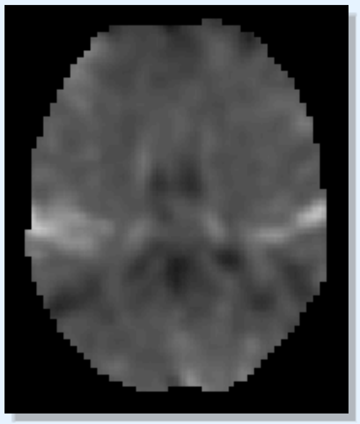
beta_???? images

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



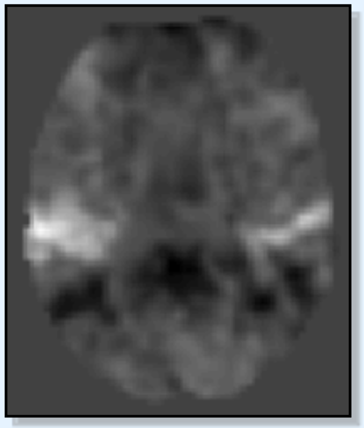
ResMS image

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



con_???? image

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



spmT_???? image

SPM{t}

T-test: summary

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

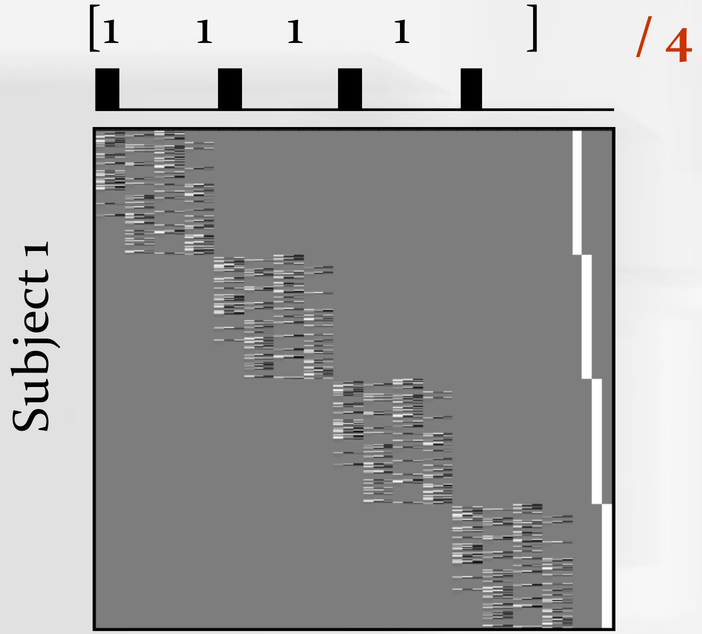
- Alternative hypothesis:

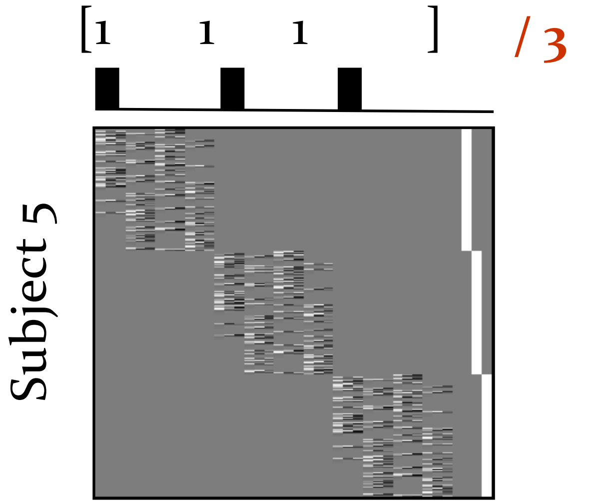
$$\text{H}_0\text{:} \quad c^T\beta = 0 \qquad \text{vs} \qquad \text{H}_A\text{:} \quad c^T\beta > 0$$

- *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.

Scaling issue

GLM | T-Test | F-Test | Multicollinearity





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

The T-statistic does not depend on the scaling of the regressors.

- ❑ 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 (eg, for a second level analysis):

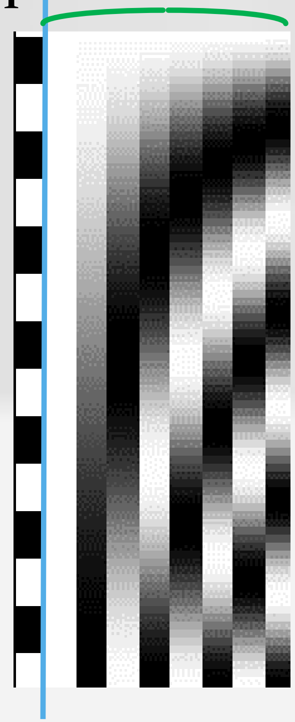
sum ≠ average

F-test: model comparison

H_0: True model is X_0

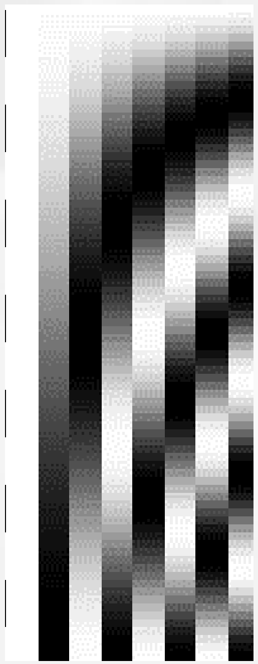
H_0: $\beta_1 = 0$

X_1 X_0



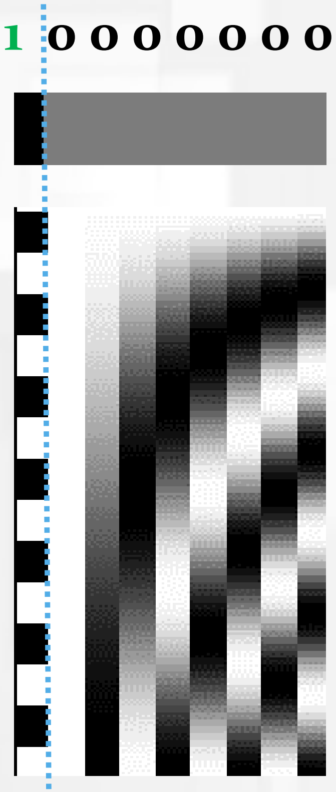
$\rightarrow \hat{\sigma}^2$

X_0



$\rightarrow \hat{\sigma}_0^2$

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



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

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

Both "activation" and "deactivations" are tested.

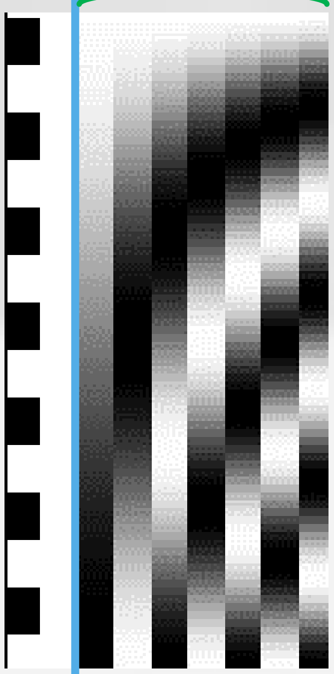
This (full) model ? Or this one?

F-test: model comparison

Tests multiple linear hypotheses : Does X1 model anything ?

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

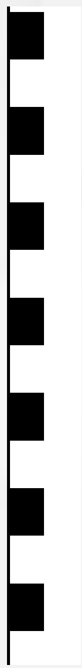
X_0 X_1



$\longrightarrow \hat{\sigma}^2$

This (full) model ?

X_0



$\longrightarrow \hat{\sigma}_0^2$

Or this one?

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

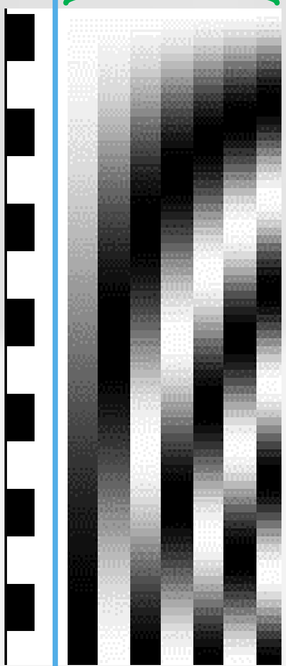
$F \sim (\hat{\sigma}_0^2 - \hat{\sigma}^2) \,/\, \hat{\sigma}^2$

F-test: model comparison

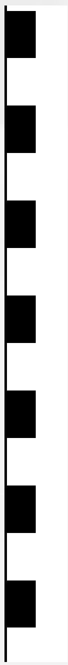
Tests multiple linear hypotheses. Example : do drift functions model anything?

H_0: True model is X_0

X_0 X_1



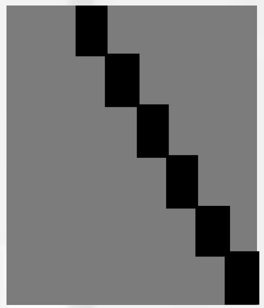
X_0

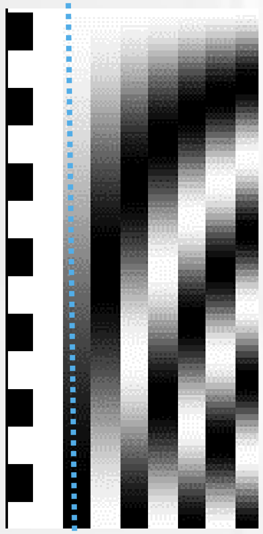


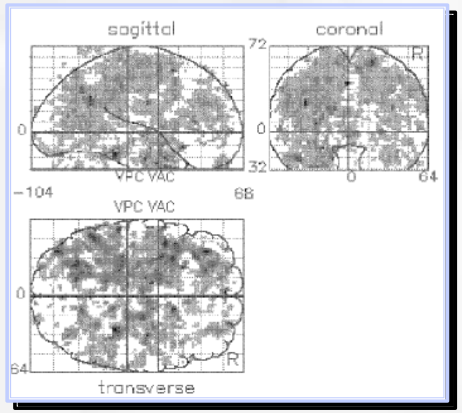
This (full) model ? Or this one?

H_0: $\beta_{3\text{-}9} = (0\ 0\ 0\ 0\ \ldots)$

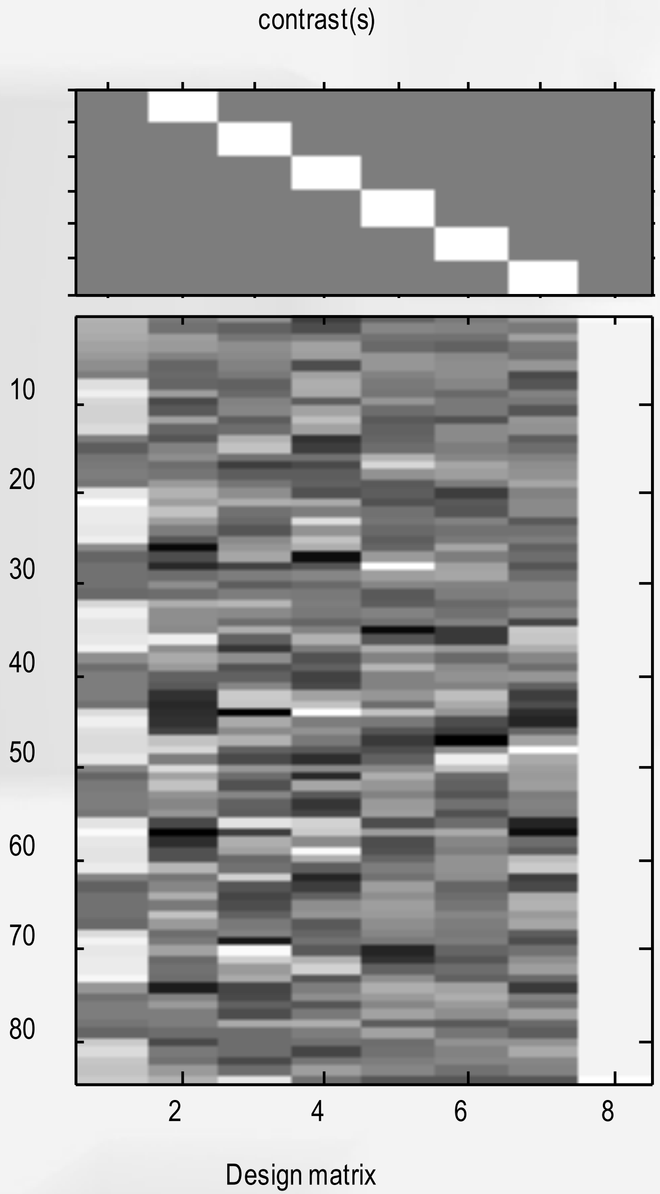
$$c' = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

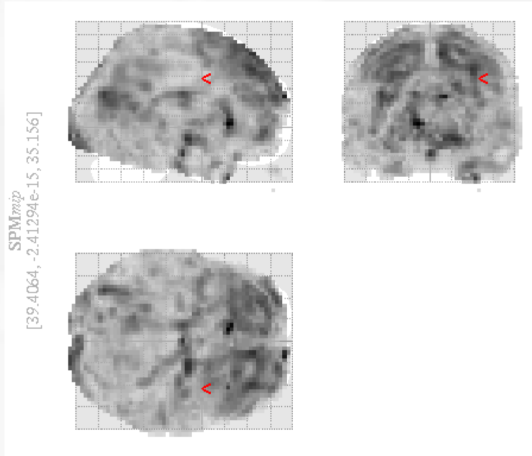


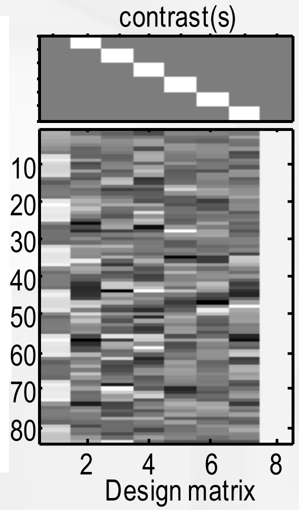


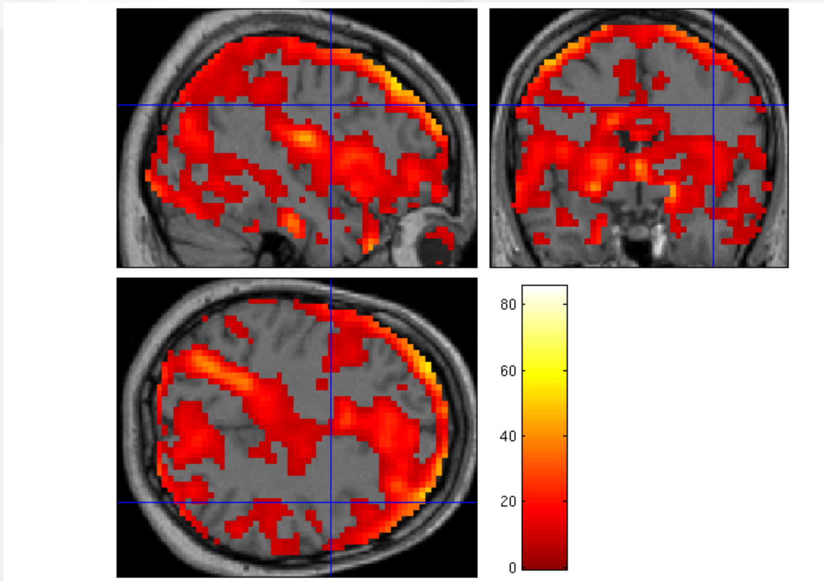


F-test example: movement related effects

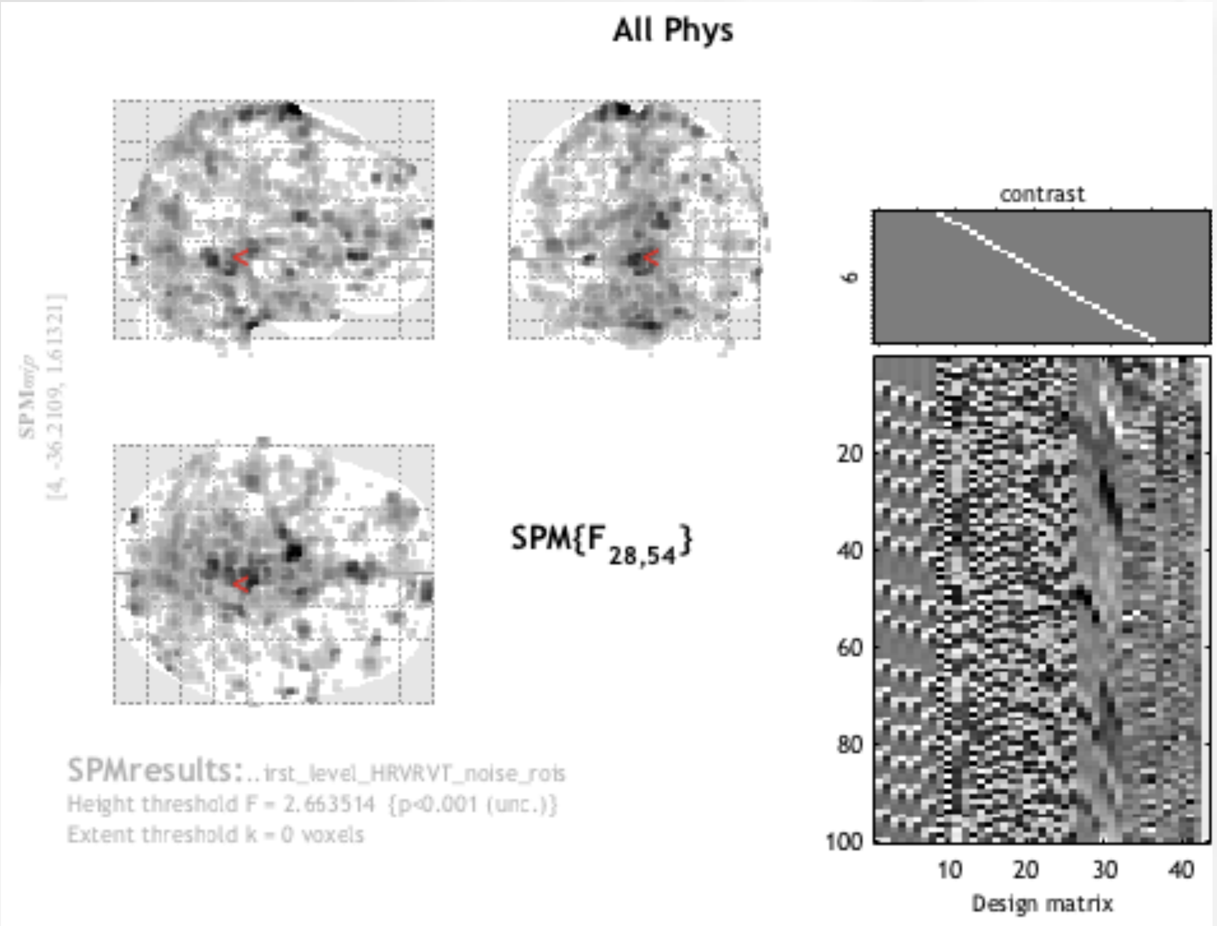


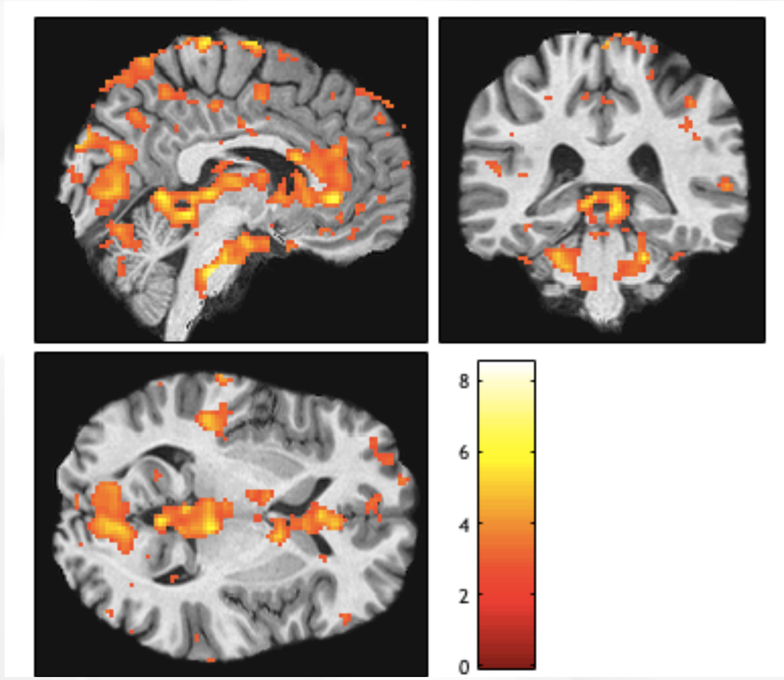


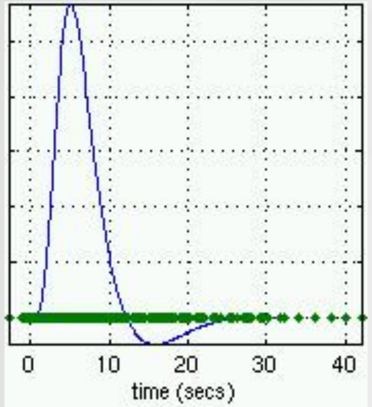
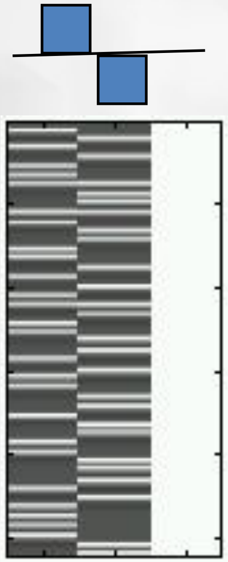
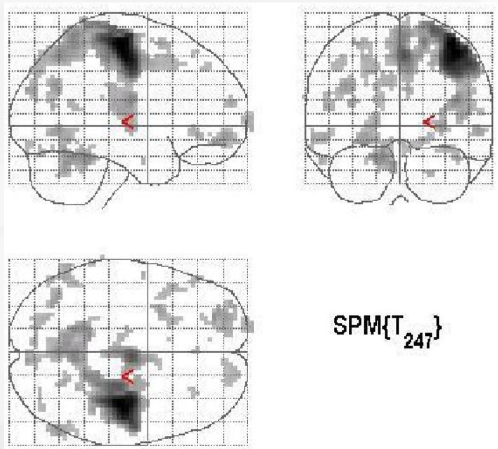
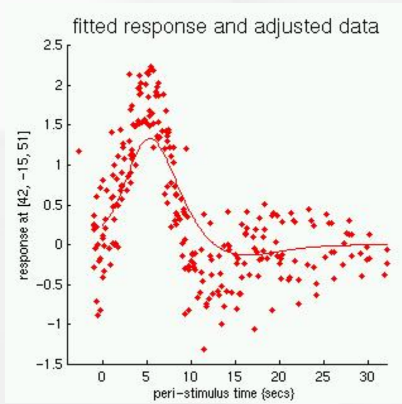
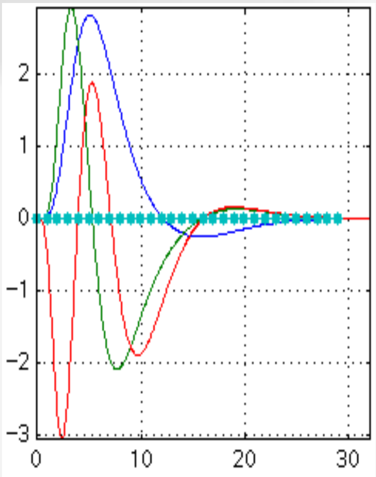
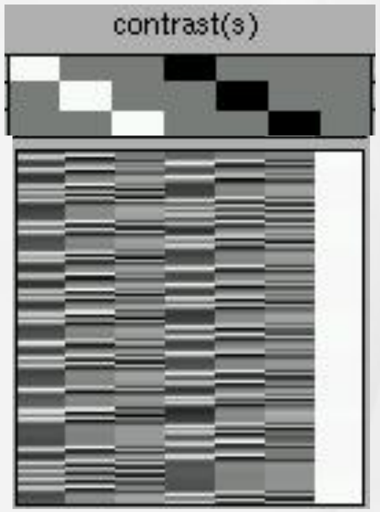
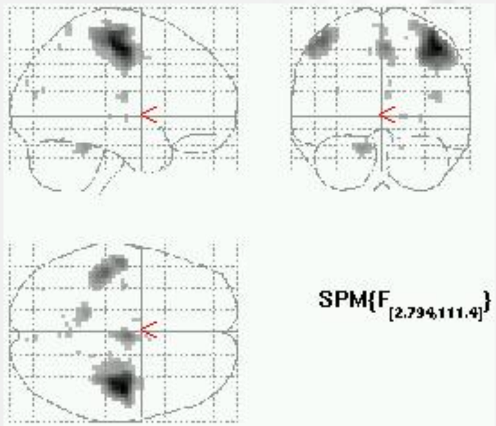
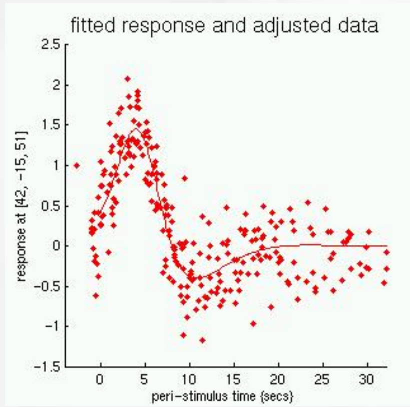




F-test example: physiological-noise related effects





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

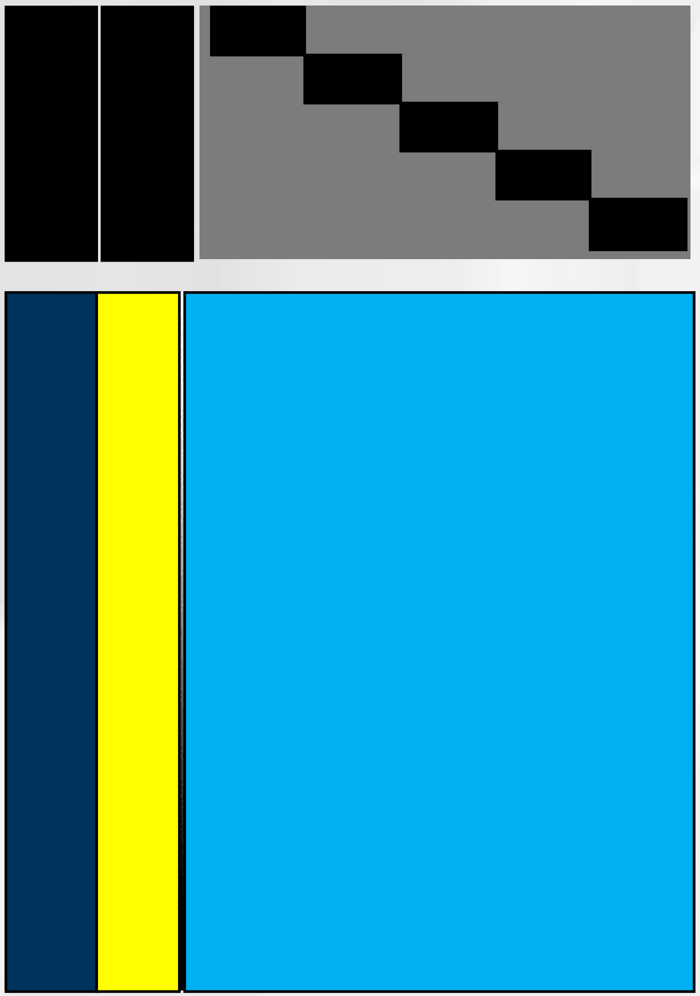
- Model and fit the data using the General Linear Model (GLM)
- T- and F-tests
  - What do they measure exactly?
- Multicollinearity

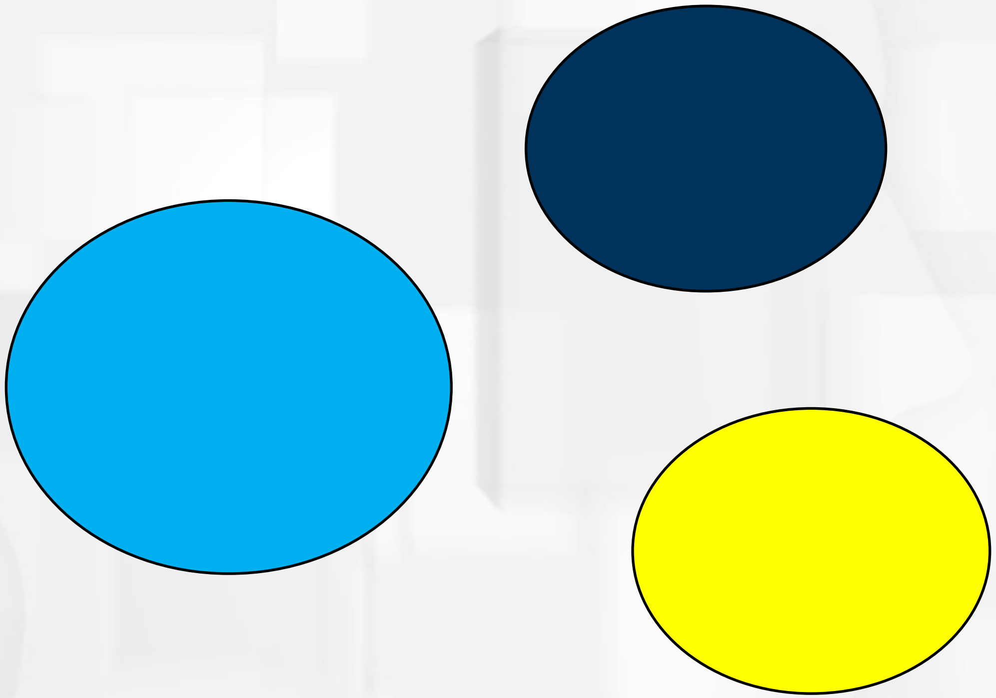
What is multicollinearity?

- 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.

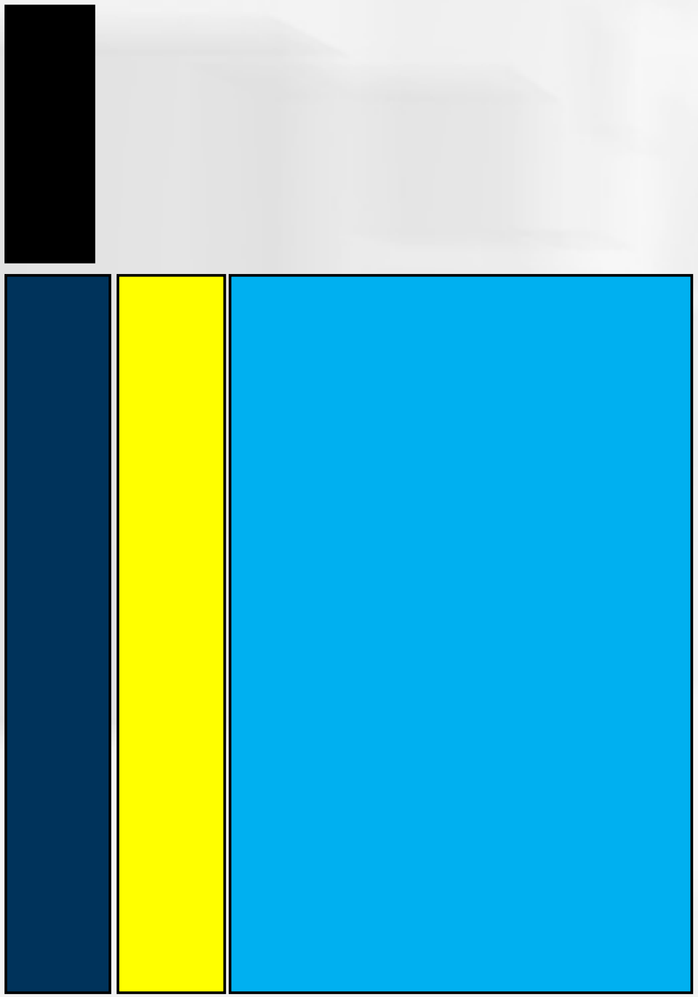
- How do we counteract it?

« Additional variance » : Again

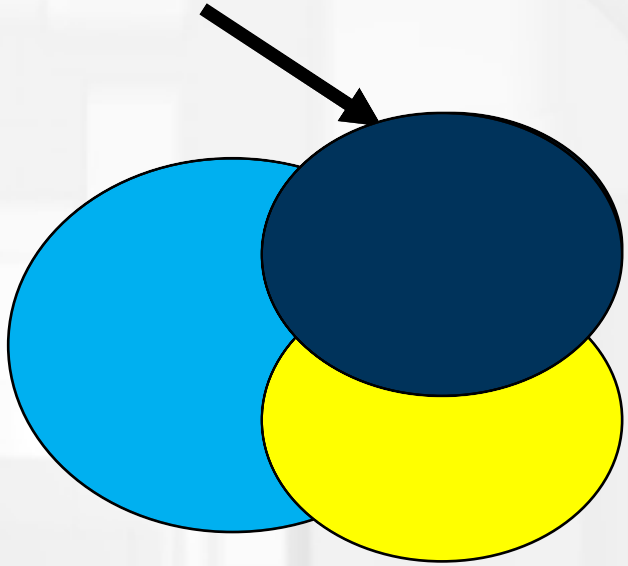




No correlation between cyan, blue and yellow



Testing for the blue

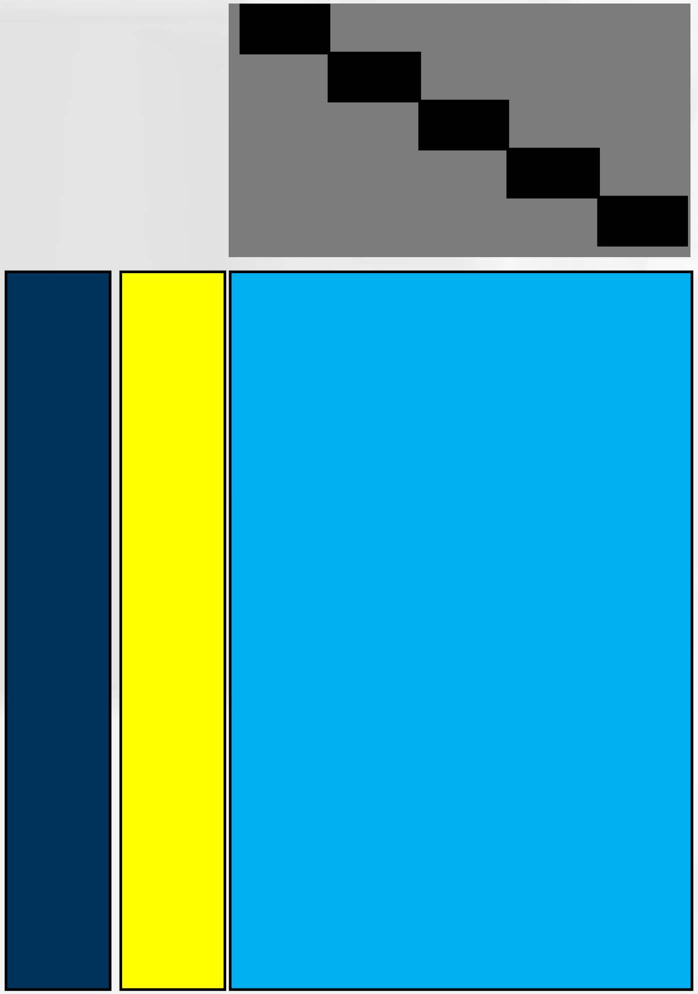


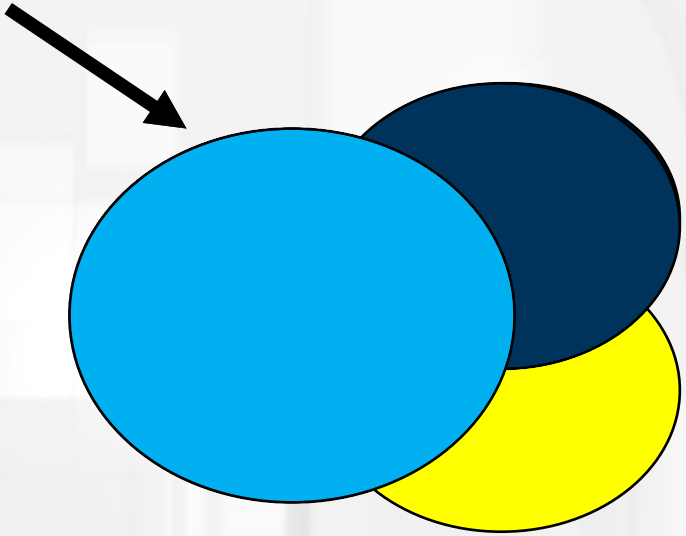
correlated regressors, for example

blue: subject age

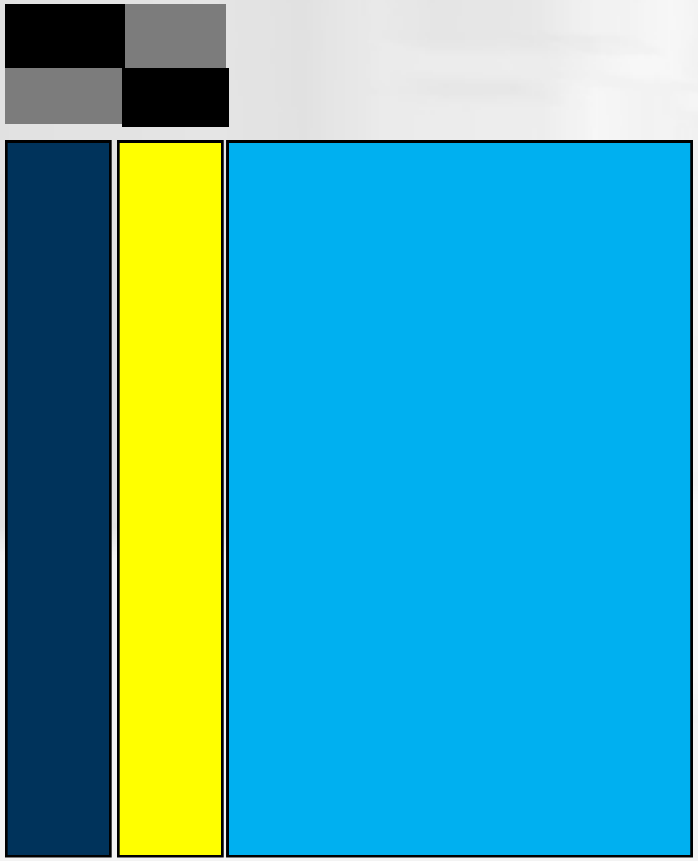
yellow: subject score

Testing for the cyan

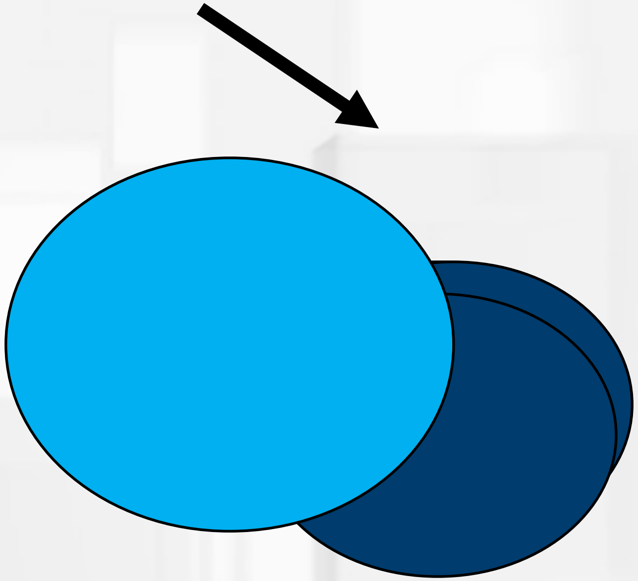




correlated contrasts

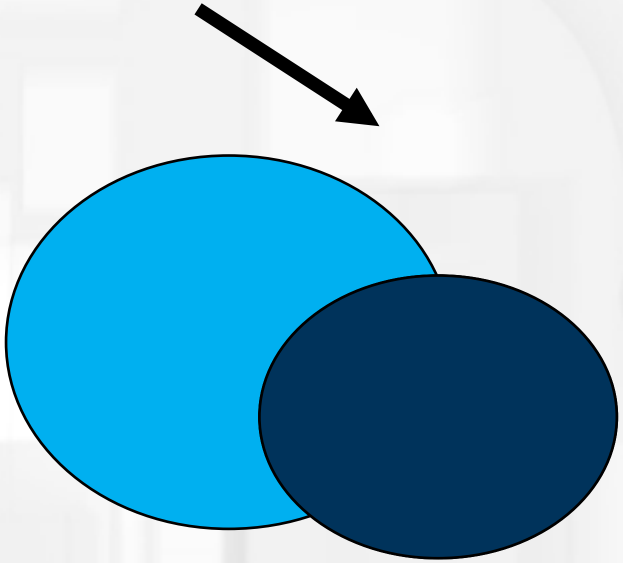


Testing for the blue and yellow



If significant ? Could be G or Y !

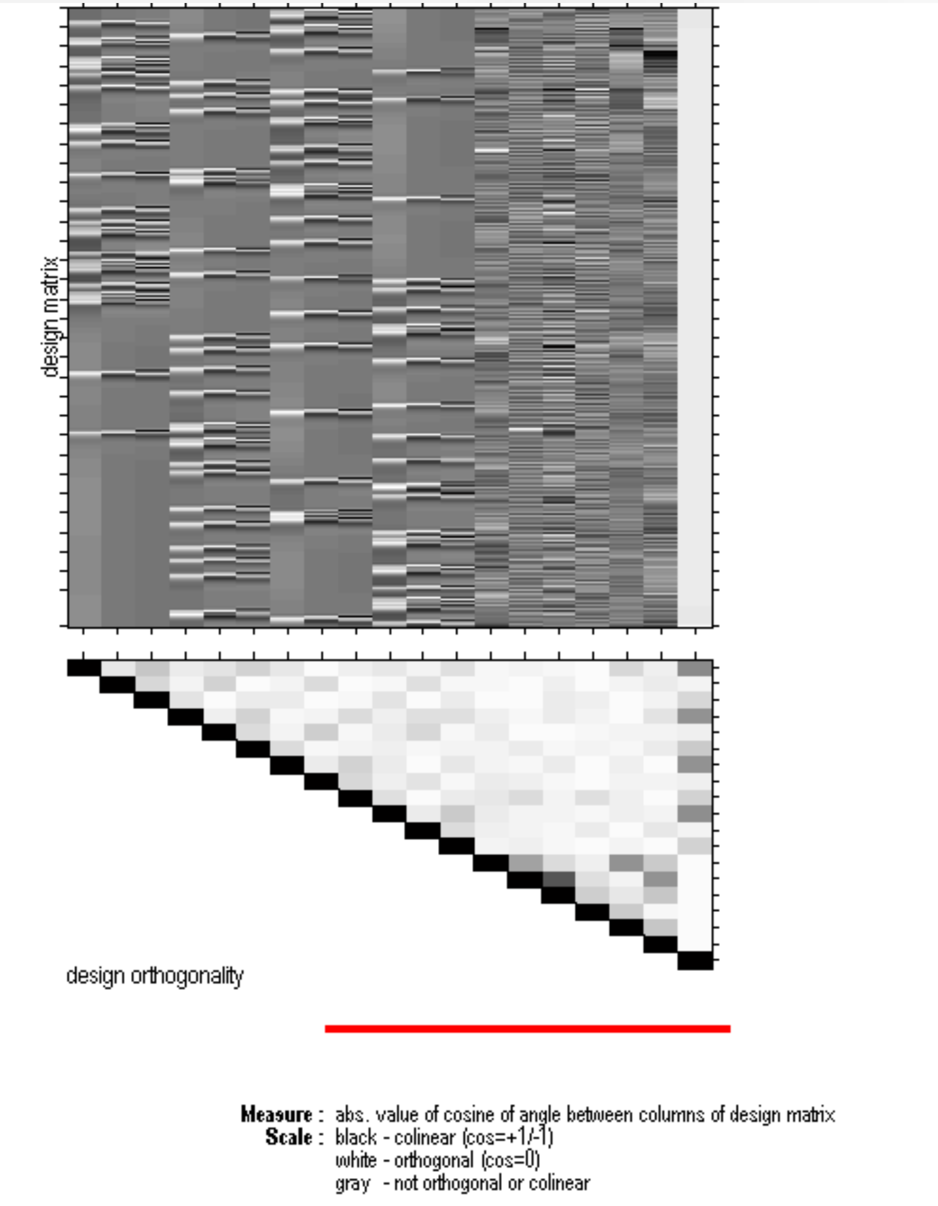
Testing for the blue



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

Design orthogonality

GLM | T-Test | F-Test | **Multicollinearity**

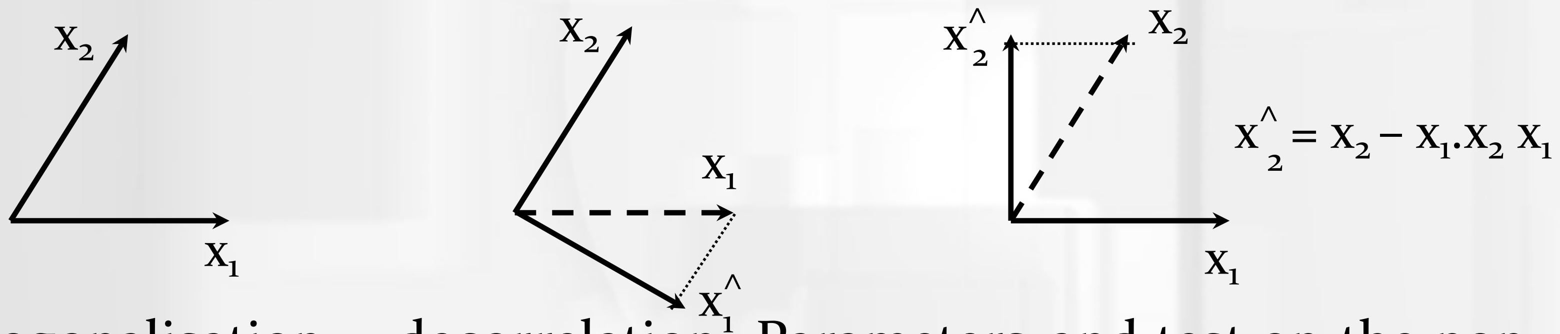


For each pair of columns of the design matrix, the orthogonality matrix depicts the magnitude of the **cosine of the angle** between them, with the range 0 to 1 mapped from white to black.

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

Multicollinearity: Take Home

- We implicitly test for an **additional** effect only. When testing for the first regressor, we are effectively removing the part of the signal that can be accounted for by the second regressor:
  ⇨ ***implicit orthogonalisation***.



- 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.
  - ⇨ change regressors (i.e. design) instead, e.g. factorial designs.
  - ⇨ 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

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*:

$$e(\hat{\sigma}^2, c, X) = (\hat{\sigma}^2 c^T (X^T X)^{-1} c)^{-1}$$

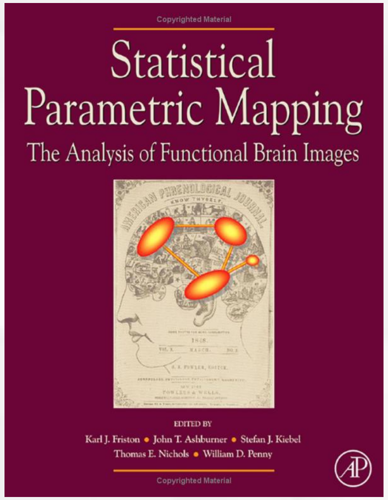
Noise variance ($\hat{\sigma}^2$) — Design variance ($(X^T X)$)

- 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:



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.