Classical inference and design efficiency

Zurich SPM Course 2015

Jakob Heinzle
heinzle@biomed.ee.ethz.ch

Translational Neuromodeling Unit (TNU)
Institute for Biomedical Engineering (IBT)
University and ETH Zürich

Many thanks to K. E. Stephan, G. Flandin and others for material

Translational Neuromodeling Unit

University of Zurich[UZH]

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Overview

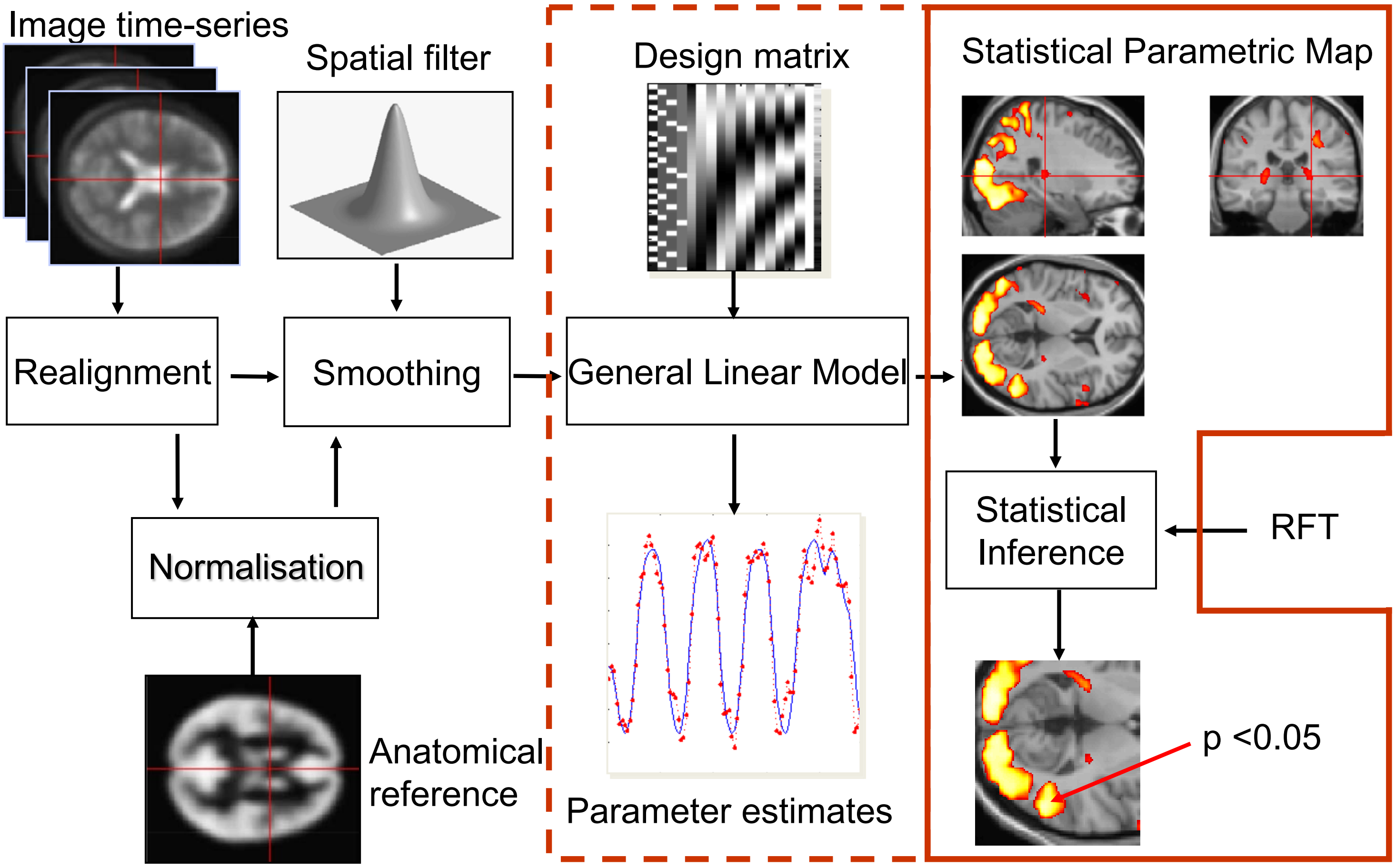
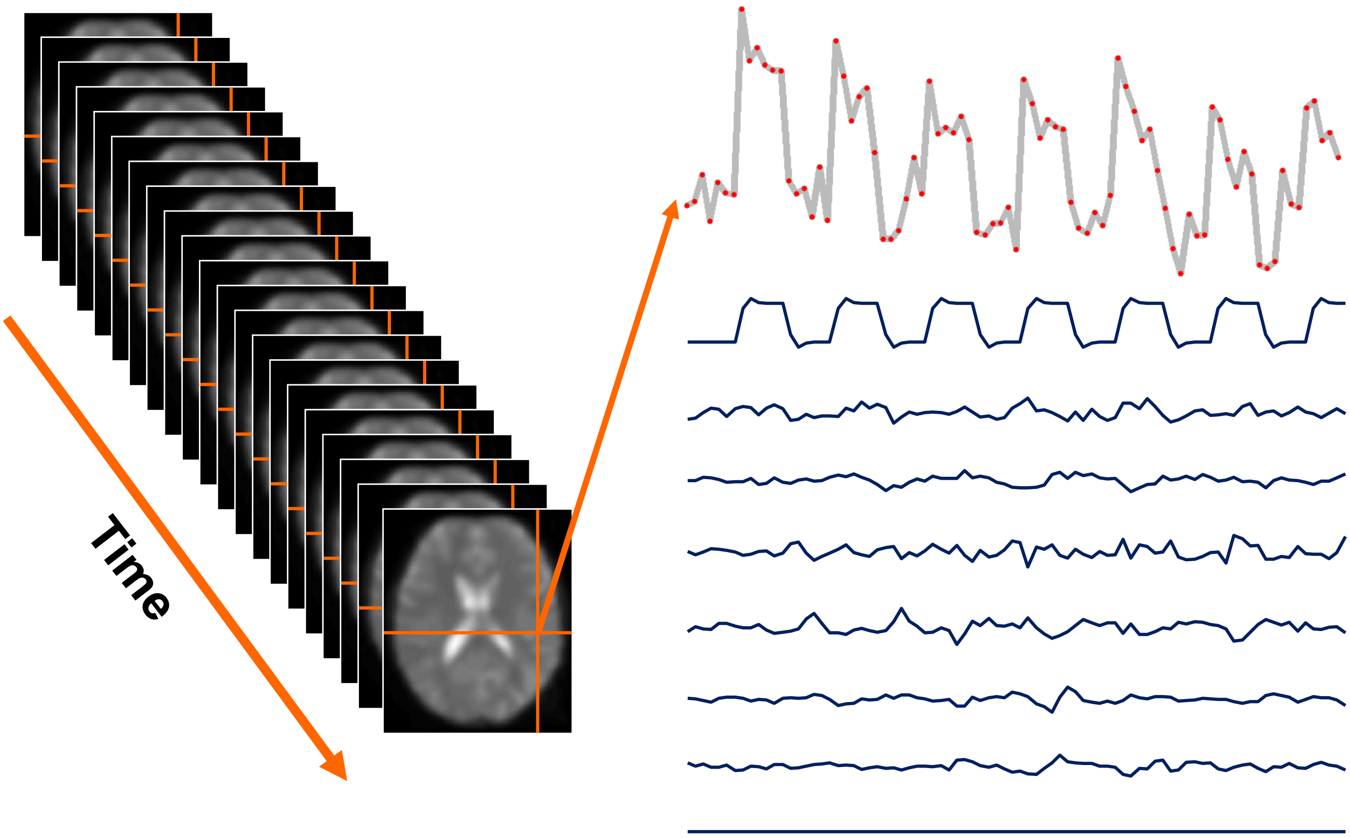


Image time-series
Spatial filter
Design matrix
Statistical Parametric Map
Realignment
Smoothing
General Linear Model
Normalisation
Statistical Inference
RFT
Anatomical reference
Parameter estimates
p <0.05

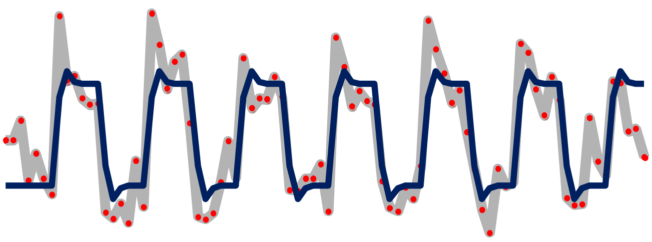
T-test

What are the values we want to make inference on?
Brief repetition of GLM

A mass-univariate approach

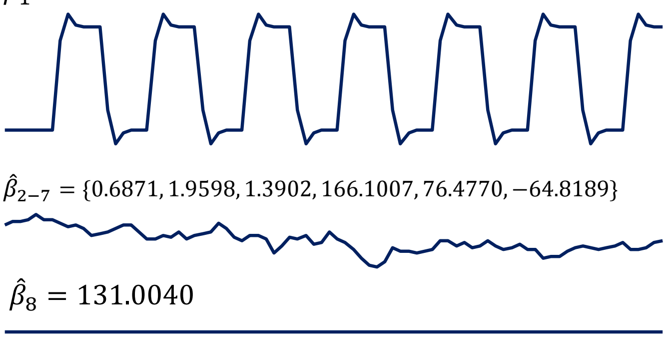


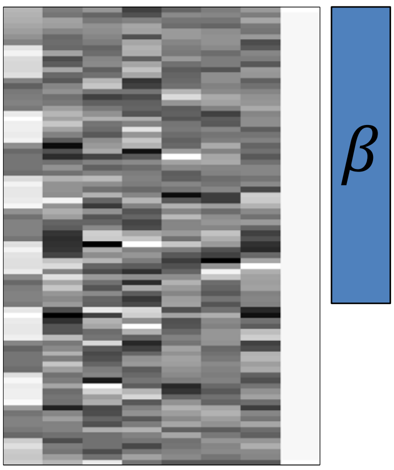
Estimation of the parameters



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

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



$y =$  β $+\varepsilon$

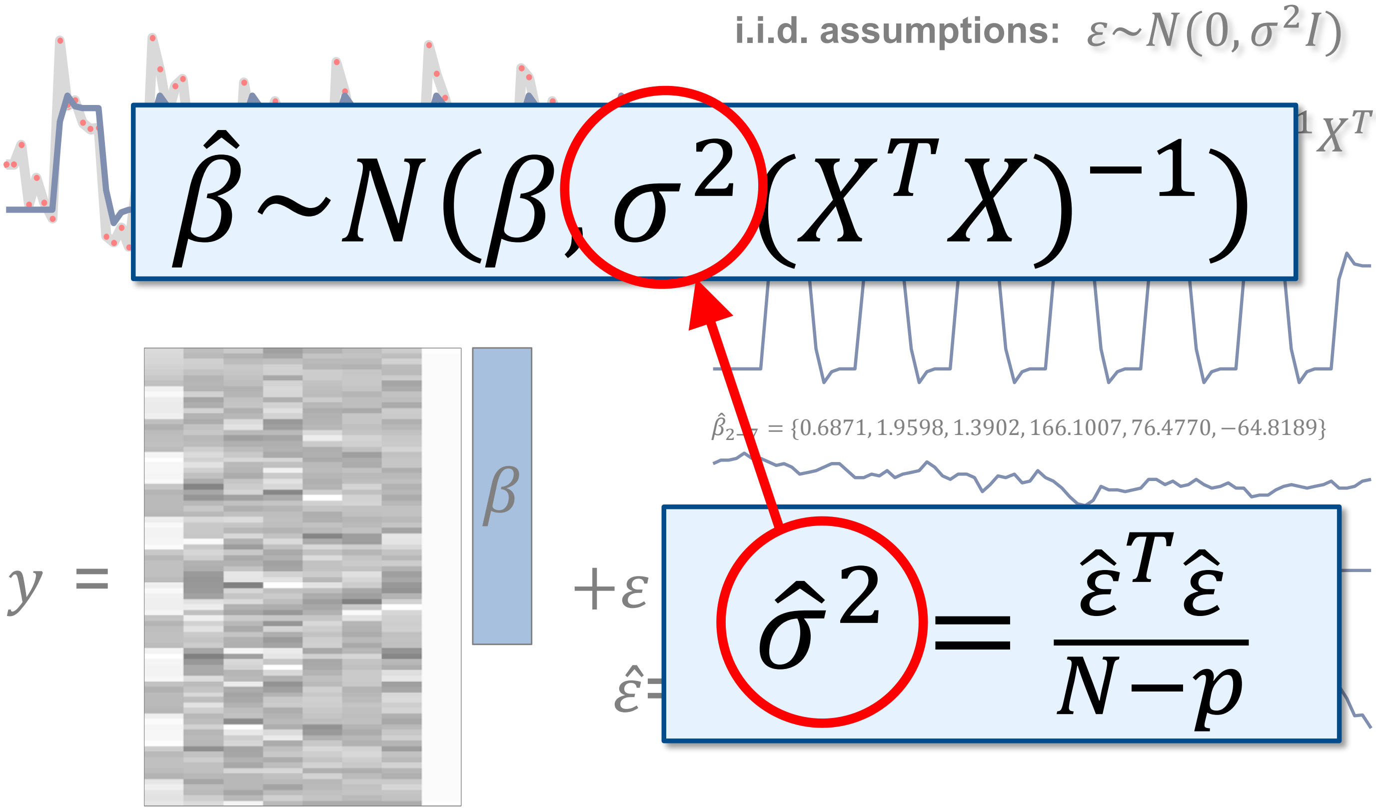
$\hat{\varepsilon} =$ 

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

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

Distribution of parameter estimates

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



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

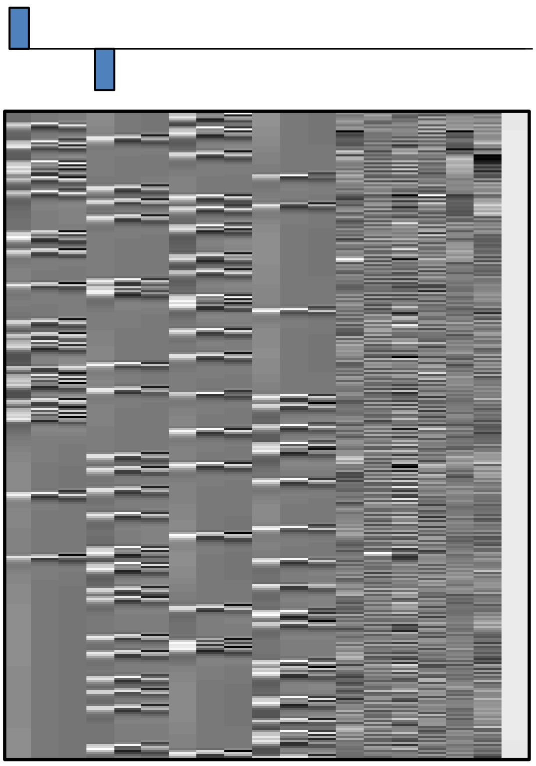
$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{N-p}$$

T-test

How can I test whether
a (combination of) regressor
has a significant effect
for explaining the data?

Contrasts

[100-1000000000000000]



- A contrast selects a specific effect of interest.
  - ⇨ A contrast c is a vector of length p.
  - ⇨ $c^T\beta$ is a linear combination of regression coefficients β.

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

$$c^T\beta = \mathbf{1} \times \beta_1 + \mathbf{0} \times \beta_2 + \mathbf{0} \times \beta_3 + \mathbf{0} \times \beta_4 + \cdots$$
$$= \boldsymbol{\beta_1}$$

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

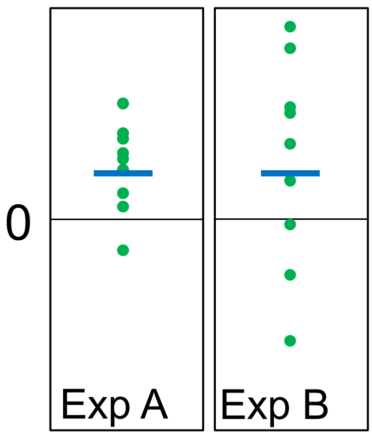
$$c^T\beta = \mathbf{1} \times \beta_1 + \mathbf{0} \times \beta_2 + \mathbf{0} \times \beta_3 + \mathbf{-1} \times \beta_4 + \cdots$$
$$= \boldsymbol{\beta_1 - \beta_4}$$

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

Hypothesis Testing - Introduction

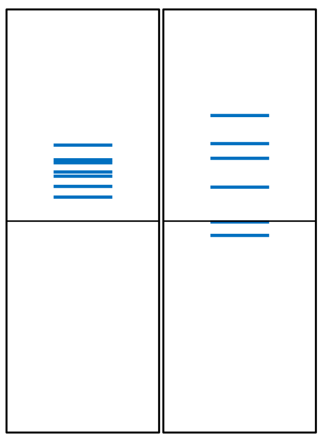
Is the mean of a measurement different from zero?

Mean of several measurements



$\mu_1, \sigma_1 \quad \mu_2, \sigma_2$

Many experiments

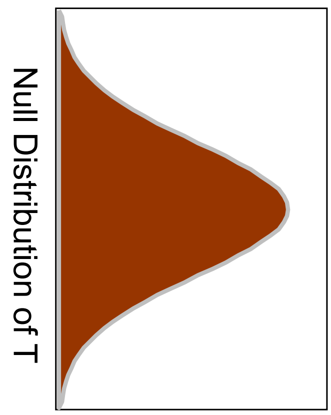


$$\sigma_\mu = {\sigma_{1,2}}/{\sqrt{n}}$$

Ratio of effect vs. noise
→ t-statistic

$$T = \frac{\mu}{\sigma / \sqrt{n}}$$

Null distribution



What distribution of T would we get for $\mu = 0$?

Hypothesis Testing

To test an hypothesis, we construct “test statistics”.

- **Null Hypothesis H_0**

  Typically what we want to disprove (no effect).

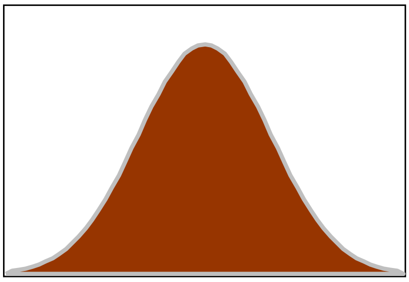
  ⇨ The *Alternative Hypothesis H_A* expresses outcome of interest.

- **Test Statistic T**

  The test statistic summarises evidence about H_0.

  Typically, test statistic is small in magnitude when the hypothesis H_0 is true and large when false.

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

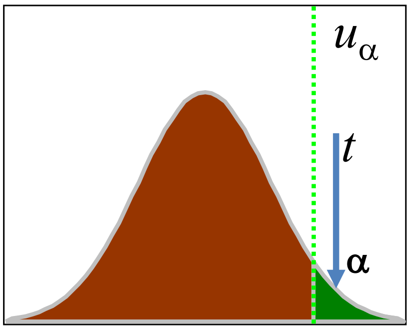


Null Distribution of T

Hypothesis Testing

- ***Significance level α:***
  Acceptable *false positive rate α*.
  ⇨ threshold u_α
  Threshold u_α controls the false positive rate

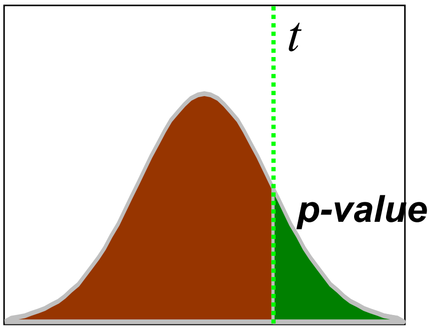
$$\alpha = p(T > u_\alpha \mid H_0)$$



Null Distribution of T

- **Conclusion about the hypothesis:**
  We reject the null hypothesis in favour of the alternative hypothesis if $t > u_\alpha$
- ***p-value***:
  A *p-value* summarises evidence against H_0.
  This is the chance of observing a value more extreme than *t* under the null hypothesis.

$$p(T > t|H_0)$$

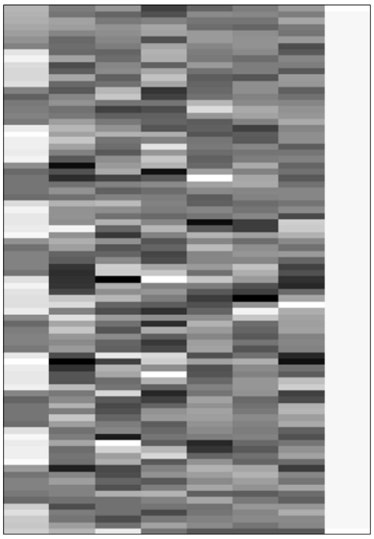


Null Distribution of T

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

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

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



Question:

effect of interest > 0 ?
=
amplitude > 0 ?
=
$\beta_1 = c^T\beta > 0$?

Null hypothesis:

$H_0: c^T\beta = 0$

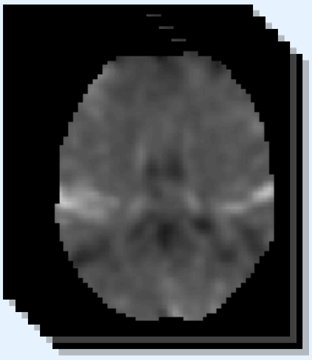
Test statistic:

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

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

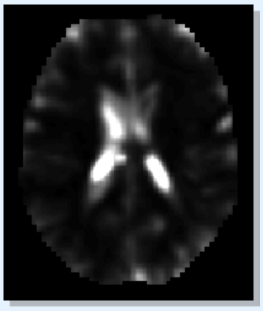
T-contrast in SPM

- For a given contrast c:



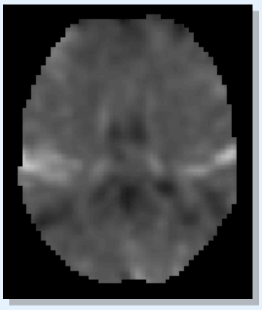
beta_???? images

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



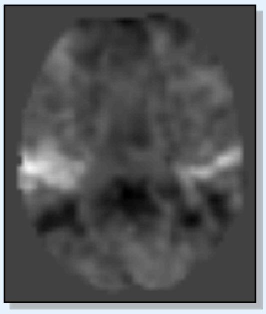
ResMS image

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



con_???? image

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



spmT_???? image

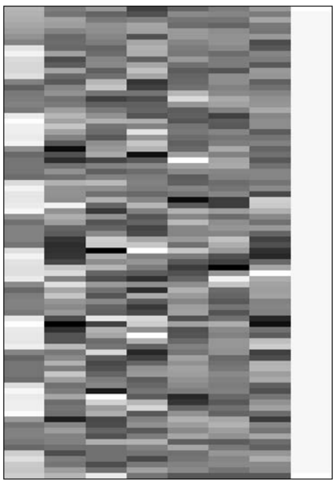
SPM{t}

T-test: a simple example

- Passive word listening versus rest

c^T = 1 0 0 0 0 0 0 0

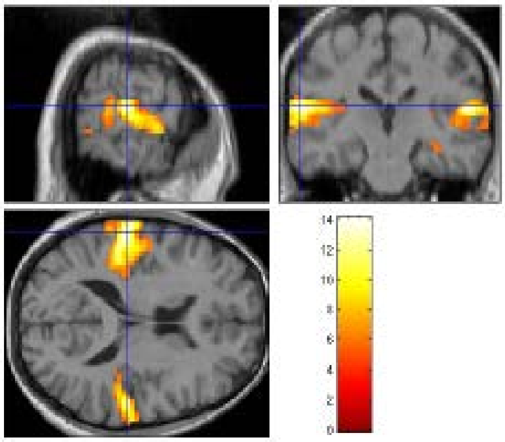
$\beta_1\ \beta_2\ \beta_3\ \beta_4\ \beta_5\ \ldots$



Q: activation during listening ?

Null hypothesis: $\beta_1 = 0$

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



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

	(Z≡)	$p_{uncorrected}$	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-test: summary

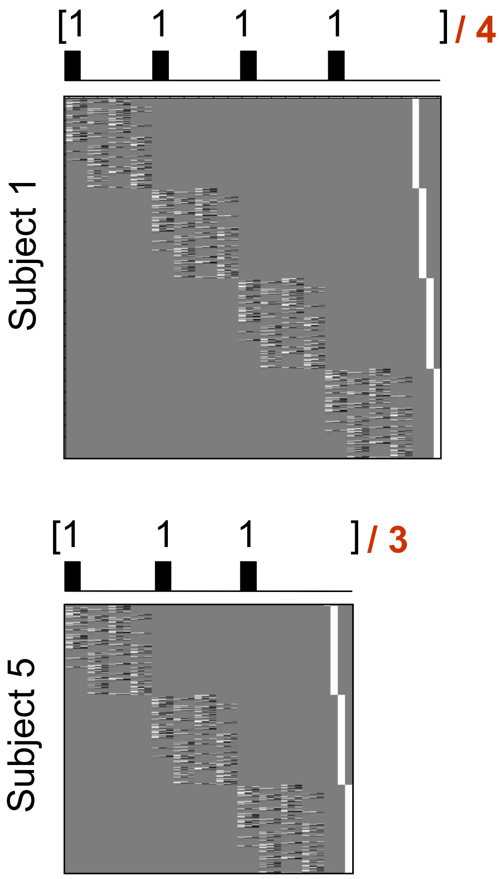
T-test is a *signal-to-noise* measure (ratio of estimate to standard deviation of estimate).

- Alternative hypothesis:

$$H_0: \; c^T \beta = 0 \qquad \text{vs} \qquad H_A: \; c^T \beta > 0$$

- *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



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

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

Scaling issues – a x c

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

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

Multiplying the contrast with a scalar does not change the t-value?

Scaling issues – b x X

$$T_b = \frac{c^T \hat{\beta}_b}{\sqrt{\mathrm{var}(c^T \hat{\beta}_b)}} = \frac{c^T \hat{\beta}_b}{\sqrt{\hat{\sigma}^2 c^T \left(bX^T bX\right)^{-1} c}}$$

$$\widehat{\beta_b} = (bX^T bX)^{-1} bX^T y = \hat{\beta}/b$$

$$T_b = \frac{c^T \hat{\beta} / b}{b^{-1}\sqrt{\hat{\sigma}^2 c^T \left(X^T X\right)^{-1} c}} = T$$

Multiplying the design matrix with a scalar does not change the t-value?

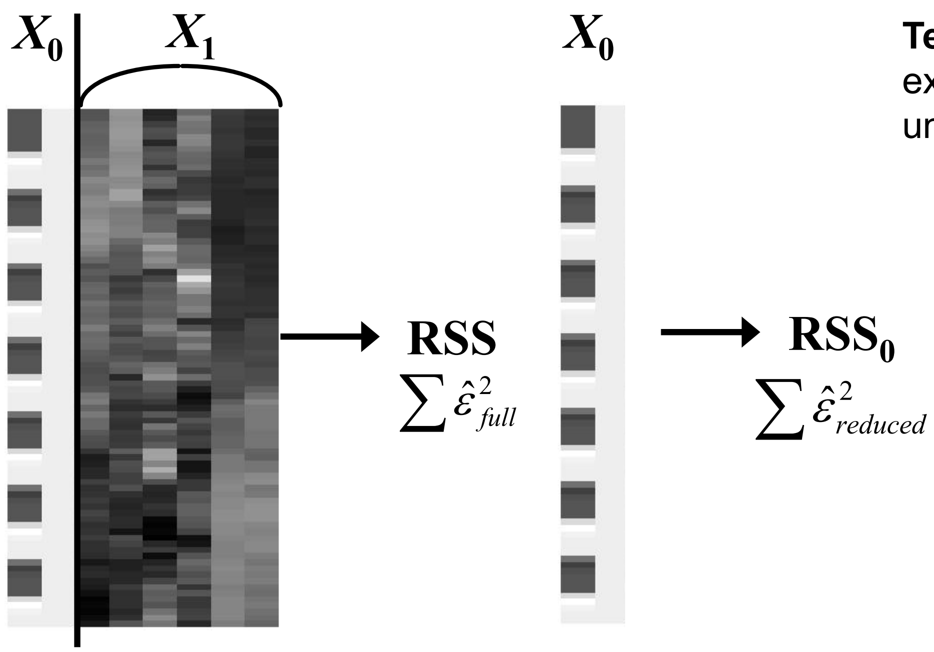
F-Test

How can I test whether (parts of) my design matrix explain any variation at all?

F-test - the extra-sum-of-squares principle

Model comparison:

Null Hypothesis H0: True model is X0 (reduced model)



Full model ? or Reduced model?

Test statistic: ratio of explained variability and unexplained variability (error)

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

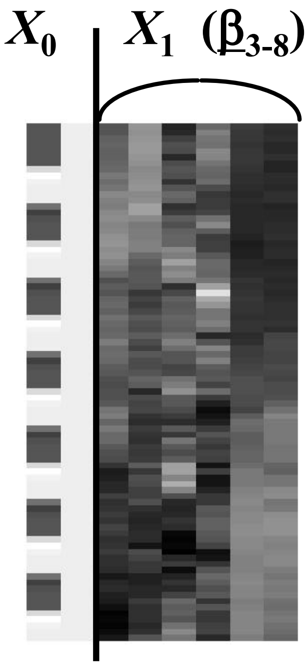
$$F \propto \frac{ESS}{RSS} \sim F_{\nu_1,\nu_2}$$

$\nu_1 = \text{rank}(X) - \text{rank}(X_0)$
$\nu_2 = N - \text{rank}(X)$

F-test - multidimensional contrasts – SPM{*F*}

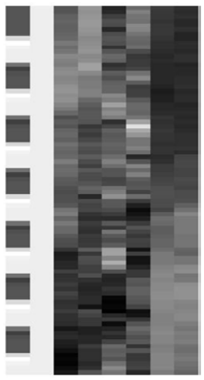
Test multiple linear hypotheses:

Null Hypothesis H0: $\beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$ → $c^T\beta = 0$

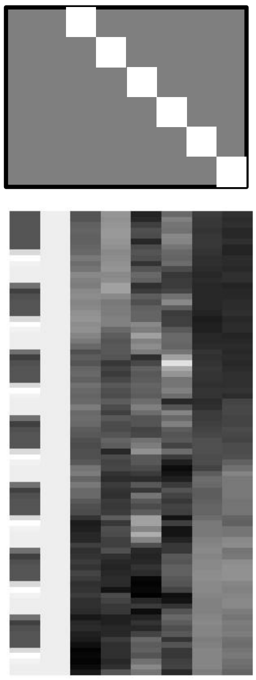


Full model ?

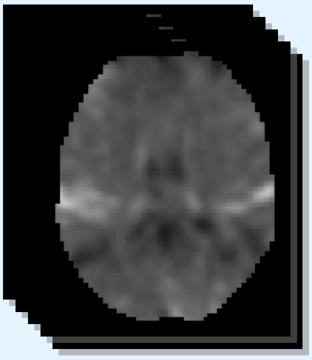
$$c^T = \begin{matrix} 0&0&1&0&0&0&0&0 \\ 0&0&0&1&0&0&0&0 \\ 0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&0&1 \end{matrix}$$



Is any of $\beta_{3\text{-}8}$ not equal 0?

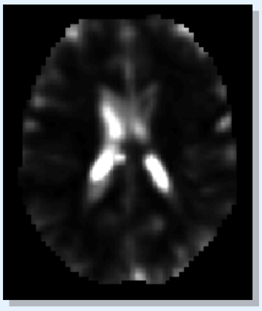


F-contrast in SPM



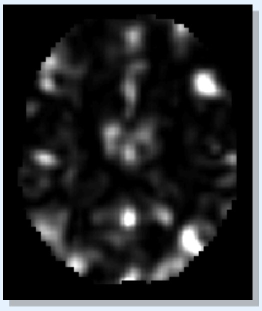
beta_???? images

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



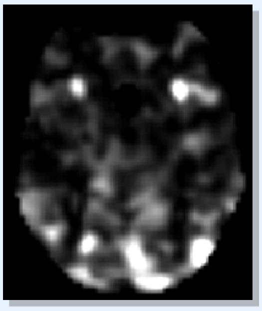
ResMS image

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



ess_???? images

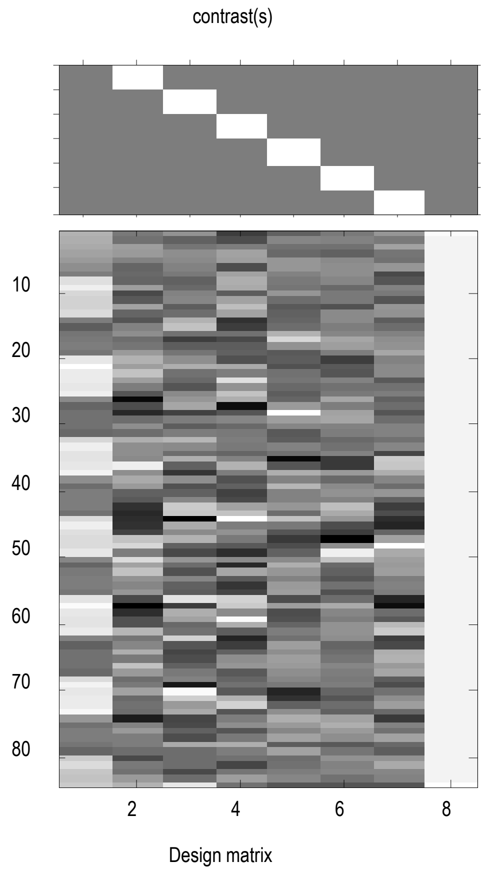
$$(RSS_0 \ - \ RSS)$$

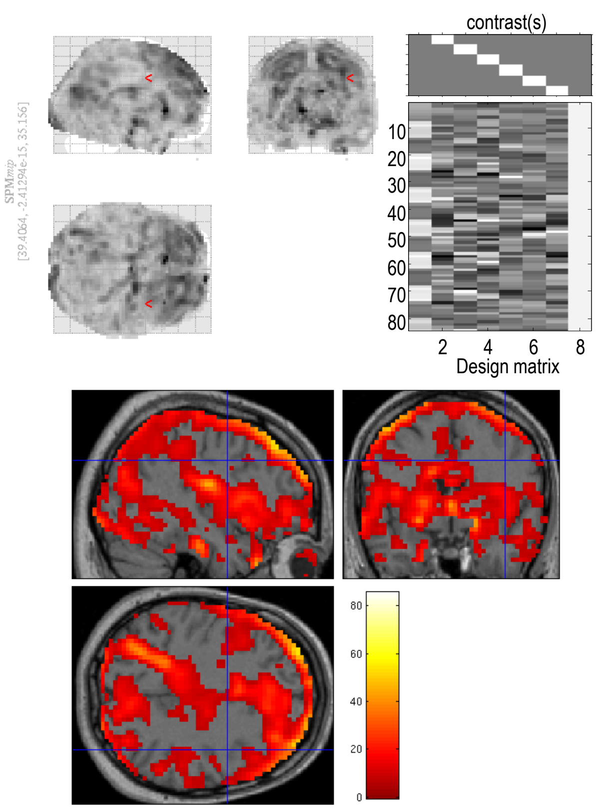


spmF_???? images

SPM{F}

F-test example: movement related effects





F-test: summary

F-tests can be viewed as testing for the additional variance explained by a larger model wrt a simpler (***nested***) model ⇨ ***model comparison***.

- F tests a weighted **sum of squares** of one or several combinations of the regression coefficients β.
- In practice, we don't have to explicitly separate X into $[X_1 X_2]$ thanks to **multidimensional contrasts**.
- Hypotheses:

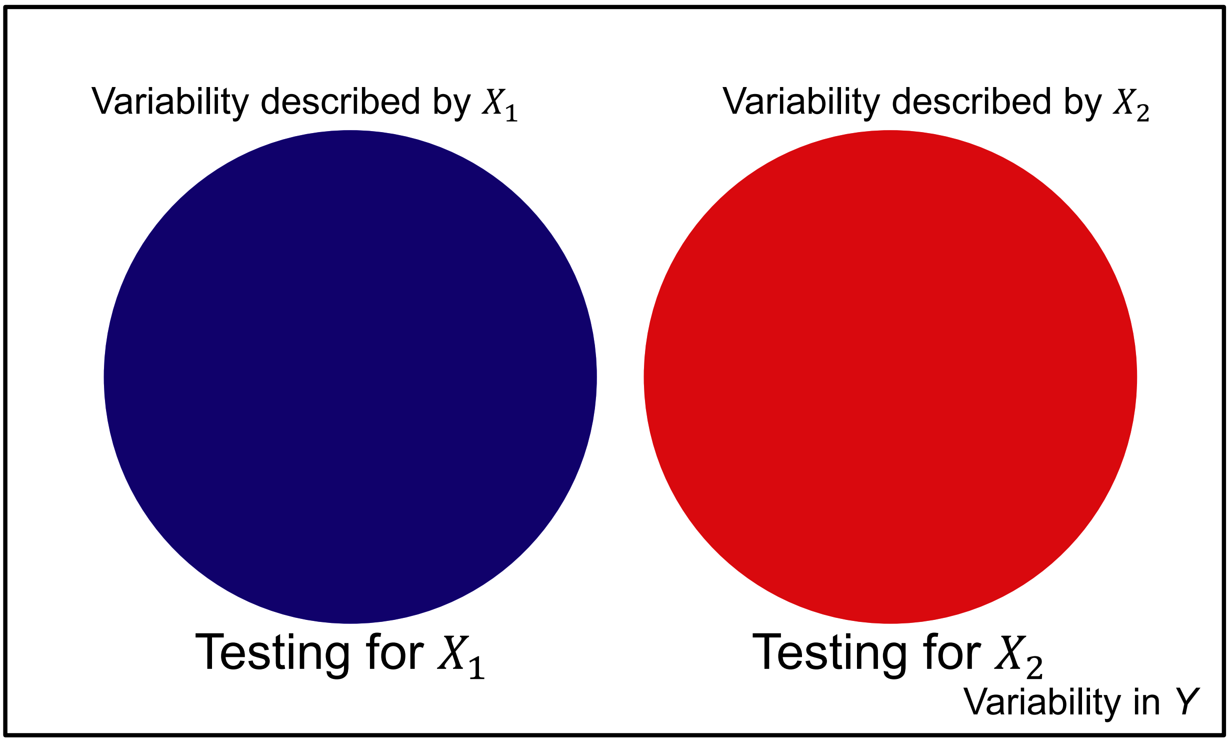
$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \begin{array}{l} \text{Null Hypothesis } H_0 : \ \beta_1 = \beta_2 = \beta_3 = 0 \\ \text{Alternative Hypothesis } H_A : \ \text{at least one } \beta_k \neq 0 \end{array}$$

- In testing uni-dimensional contrast with an *F*-test, for example $\beta_1 - \beta_2$, the result will be the same as testing $\beta_2 - \beta_1$. It will be exactly the square of the *t*-test, testing for both positive and negative effects.

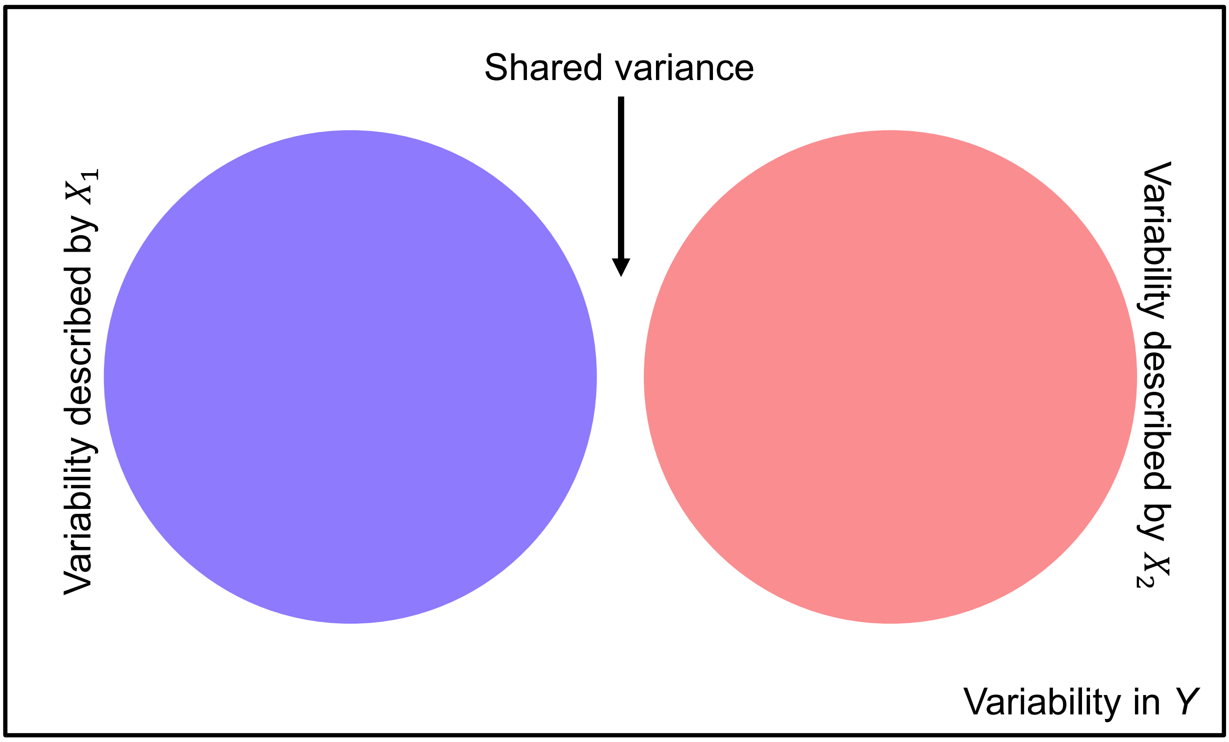
Orthogonal regressors

What's (not) the problem if I use a design with correlated regressors?

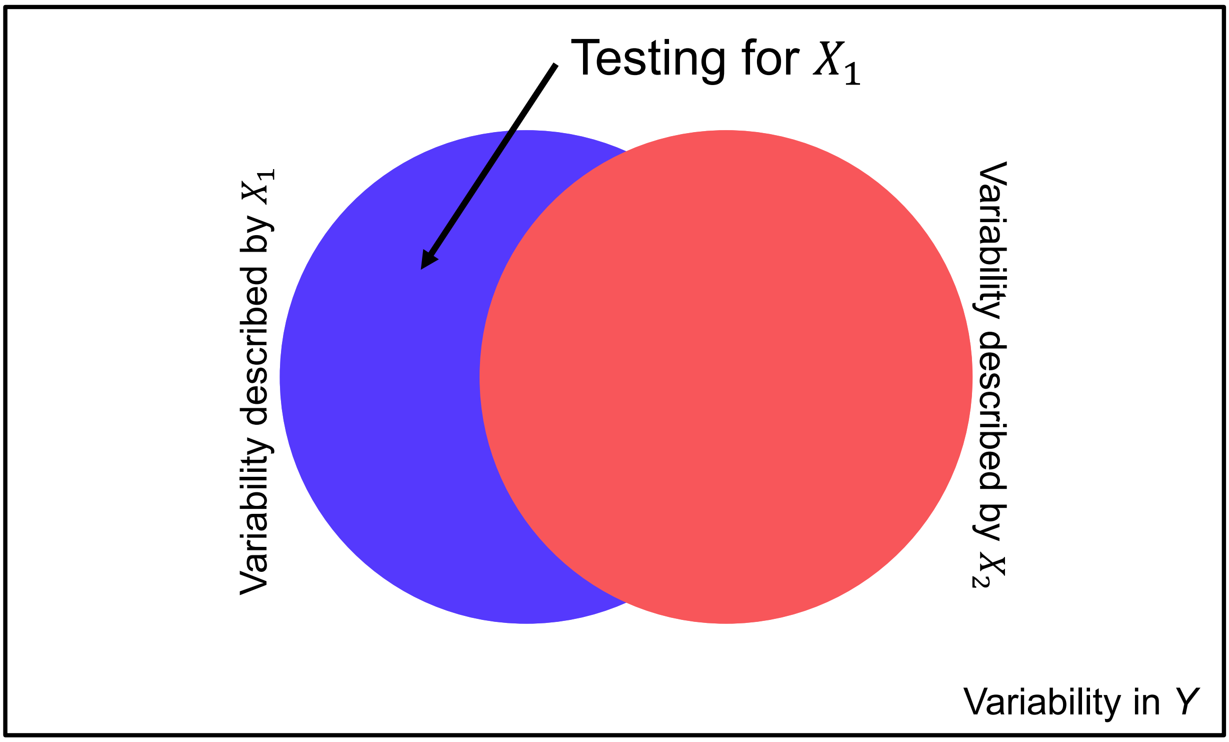
Orthogonal regressors



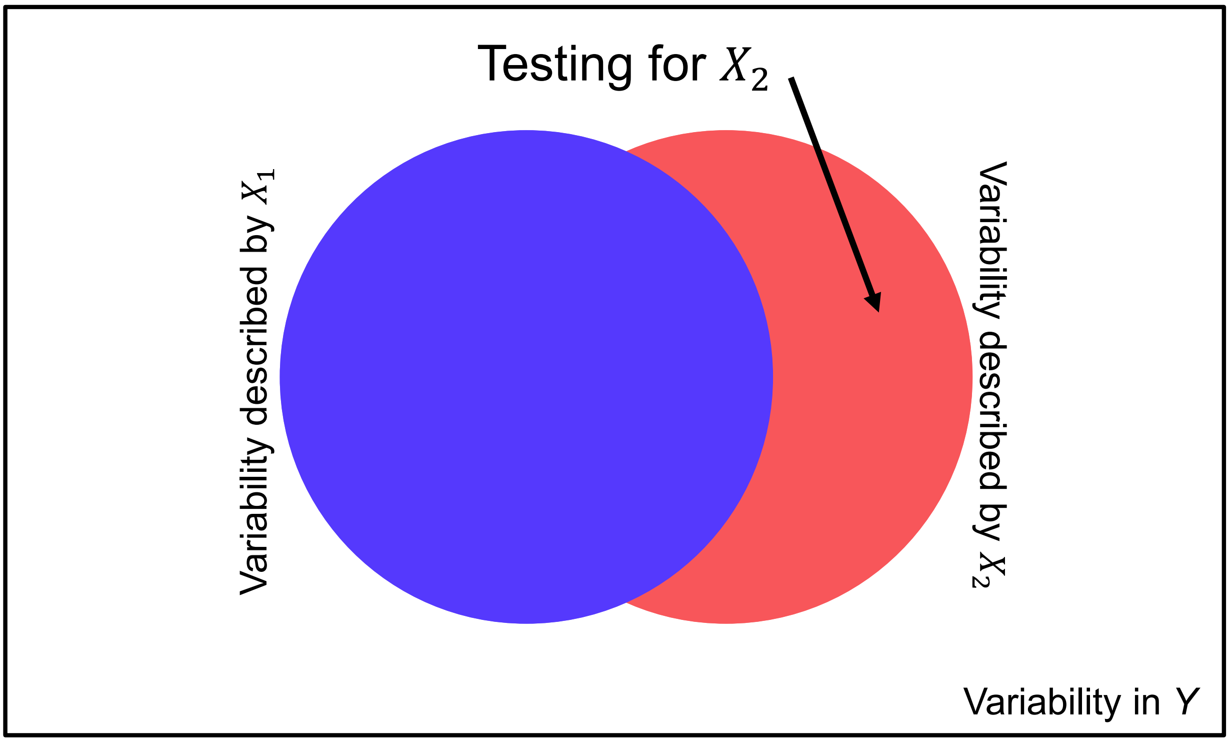
Correlated regressors



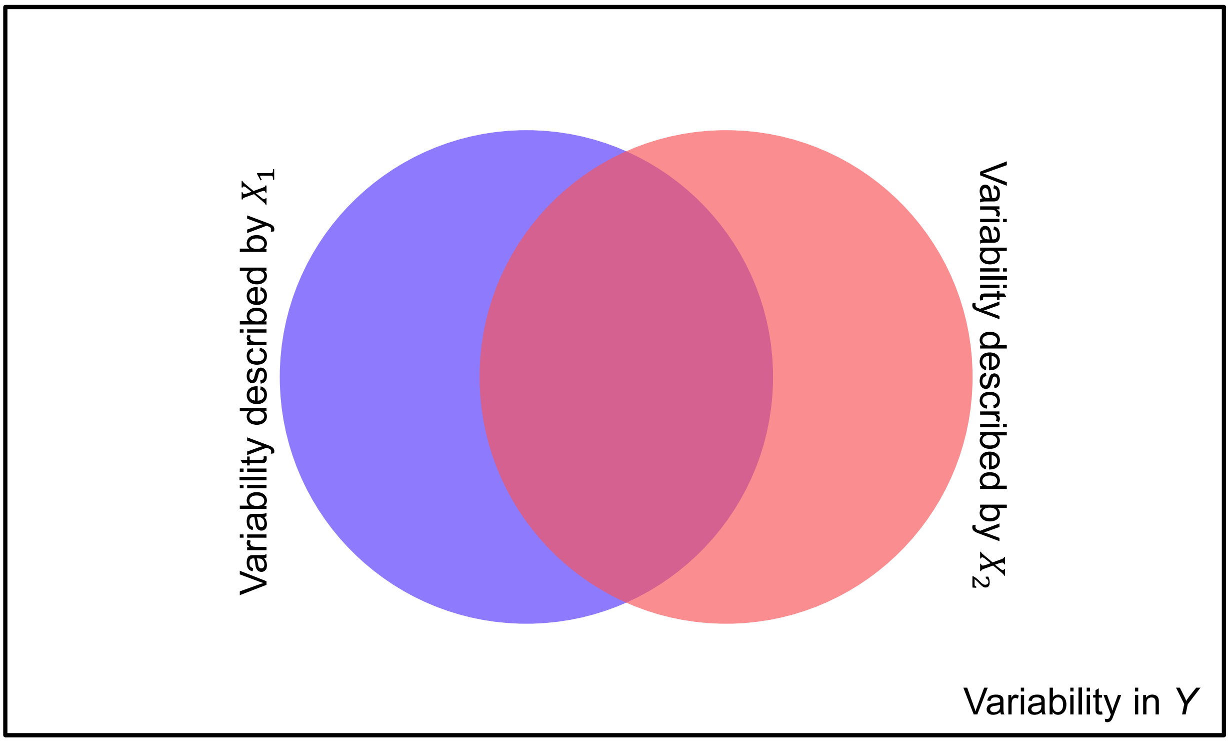
Correlated regressors



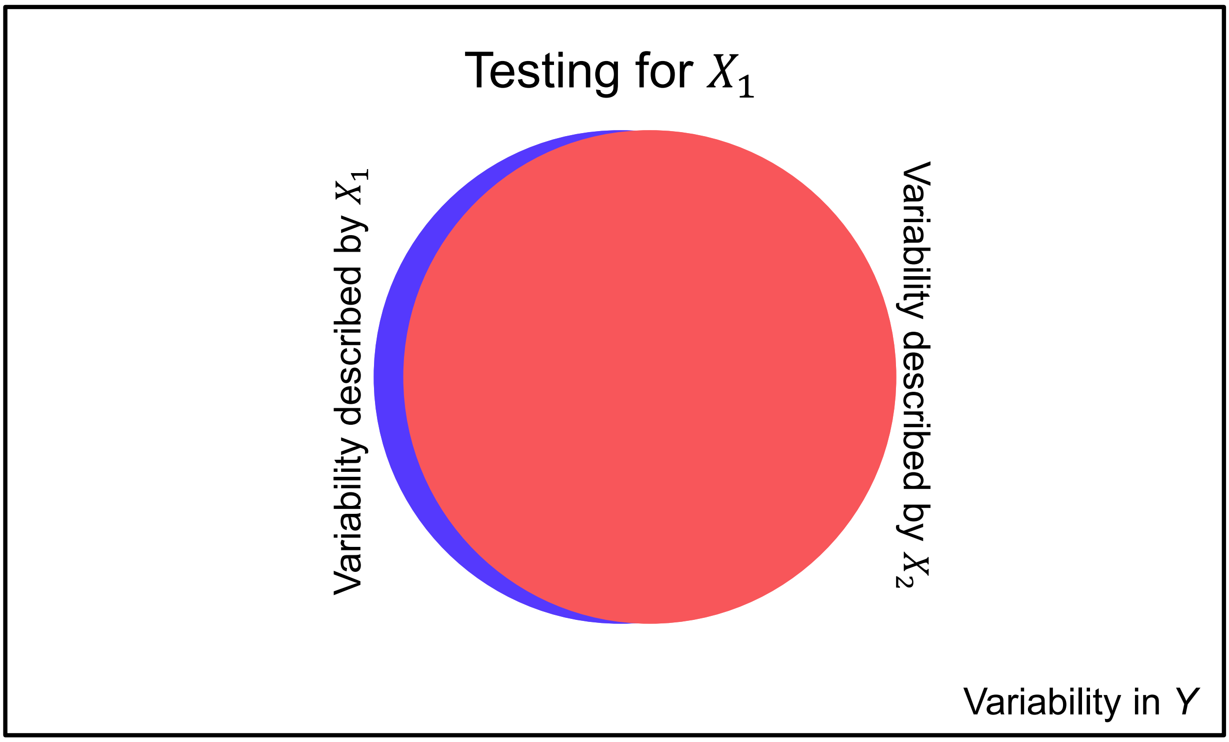
Correlated regressors



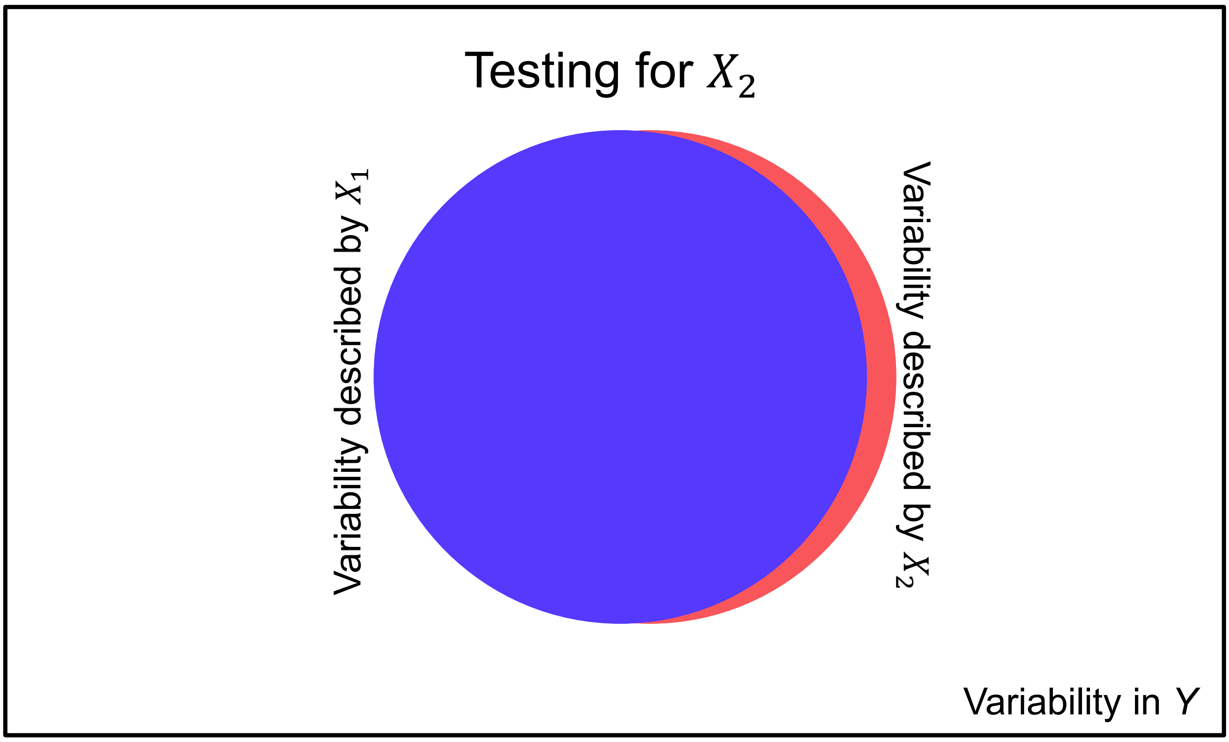
Correlated regressors



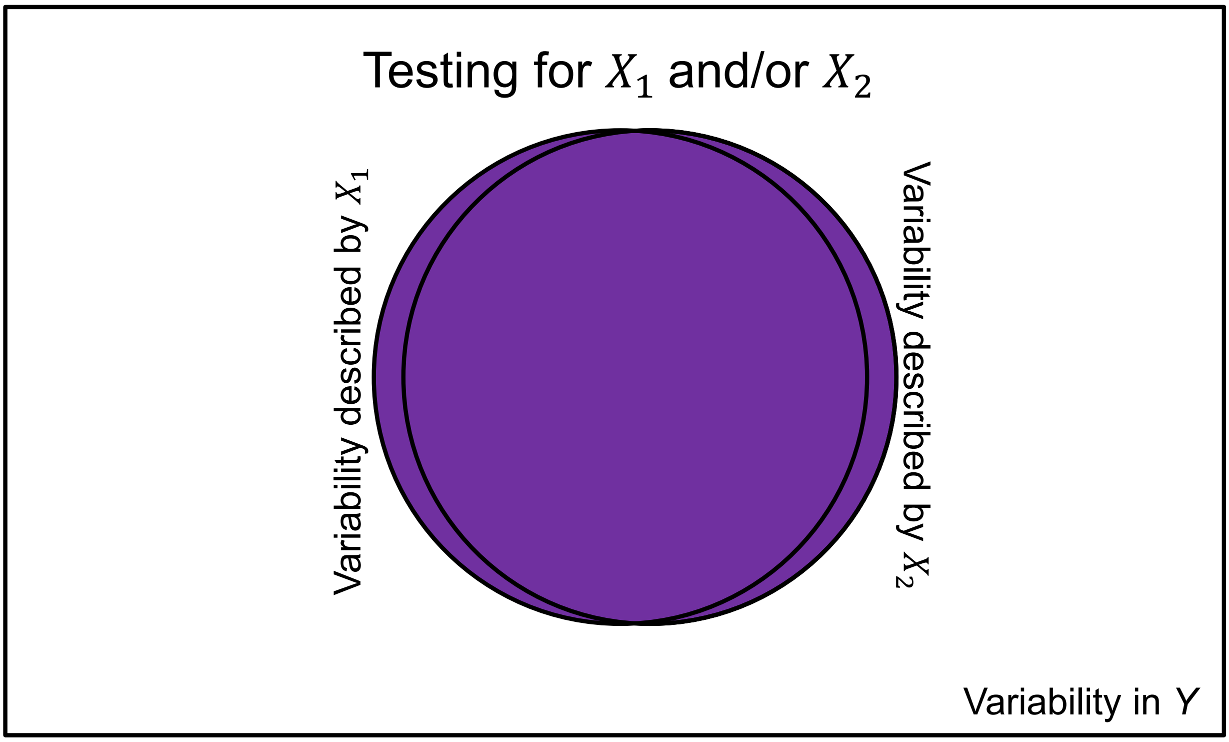
Correlated regressors



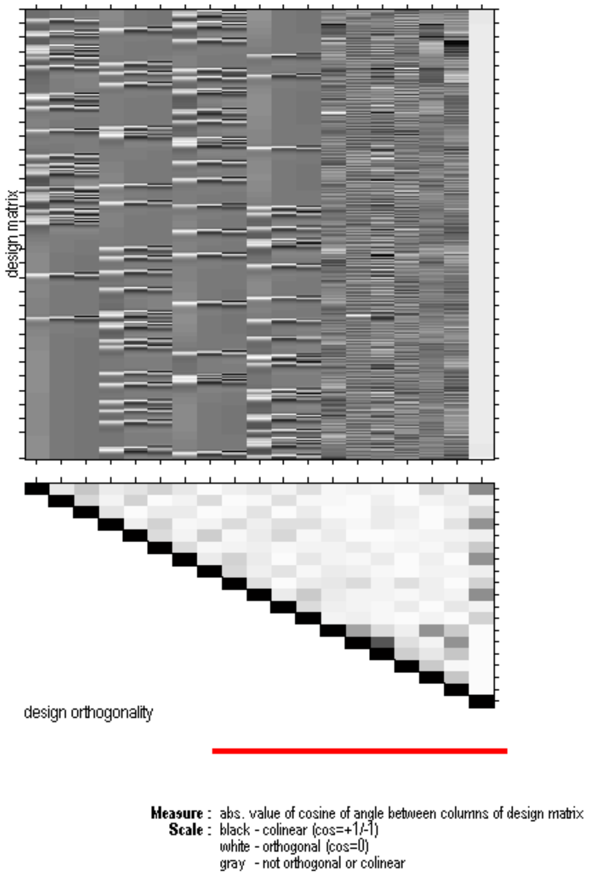
Correlated regressors



Correlated regressors



Design orthogonality

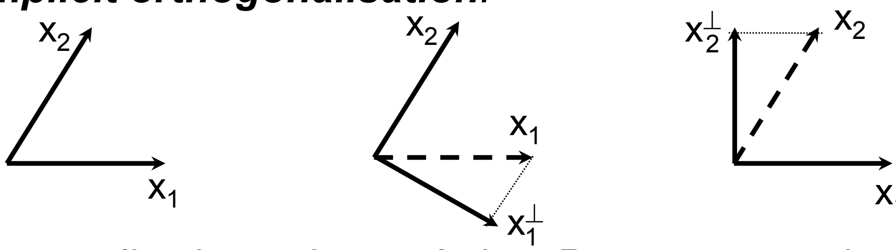


For each pair of columns of the design matrix, the orthogonality matrix depicts the magnitude of the **cosine of the angle** between them, with the range 0 to 1 mapped from white to black.

- If both vectors have **zero mean** then the cosine of the angle between the vectors is the same as the **correlation** between the two variates.

Correlated regressors: summary

- 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***.



$$x_2^{\perp} = x_2 - x_1.x_2\ x_1$$

- 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's the contrast you are testing which should be as decorrelated as possible from the rest of the design matrix

Design efficiency

How can I make my experimental design as good (powerful) as possible?

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*:

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

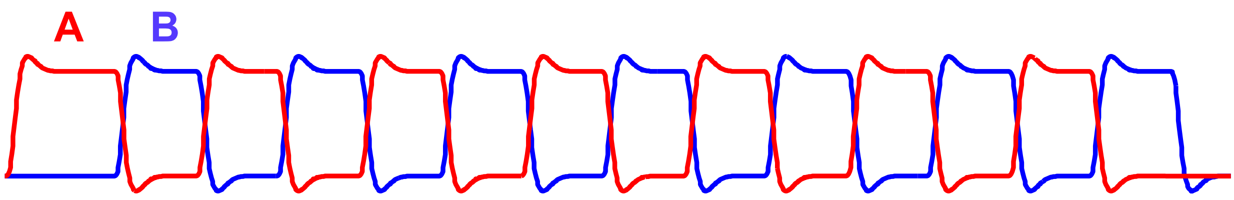
Noise variance ($\hat{\sigma}^2$) — Design variance ($X^T X$)

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

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

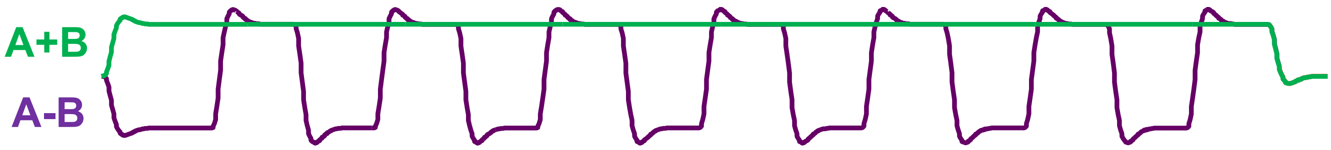
- This is a relative measure: all we can really say is that one design is more efficient than another (for a given contrast).

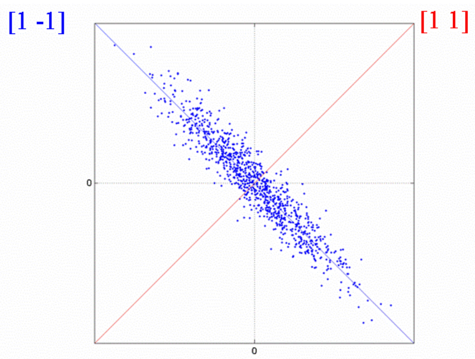
Design efficiency



$$X^T X = \begin{pmatrix} 1 & -0.9 \\ -0.9 & 1 \end{pmatrix}$$

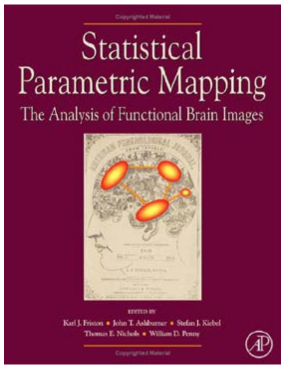
$$c = [1\ 0]^T: \quad e(c, X) = 18.1$$
$$c = [0.5\ 0.5]^T: \quad e(c, X) = 19.0$$
$$c = [1\ -1]^T: \quad e(c, X) = 95.2$$





- ❑ High correlation between regressors leads to low sensitivity to each regressor alone.
- ❑ We can still estimate efficiently the difference between them.

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.