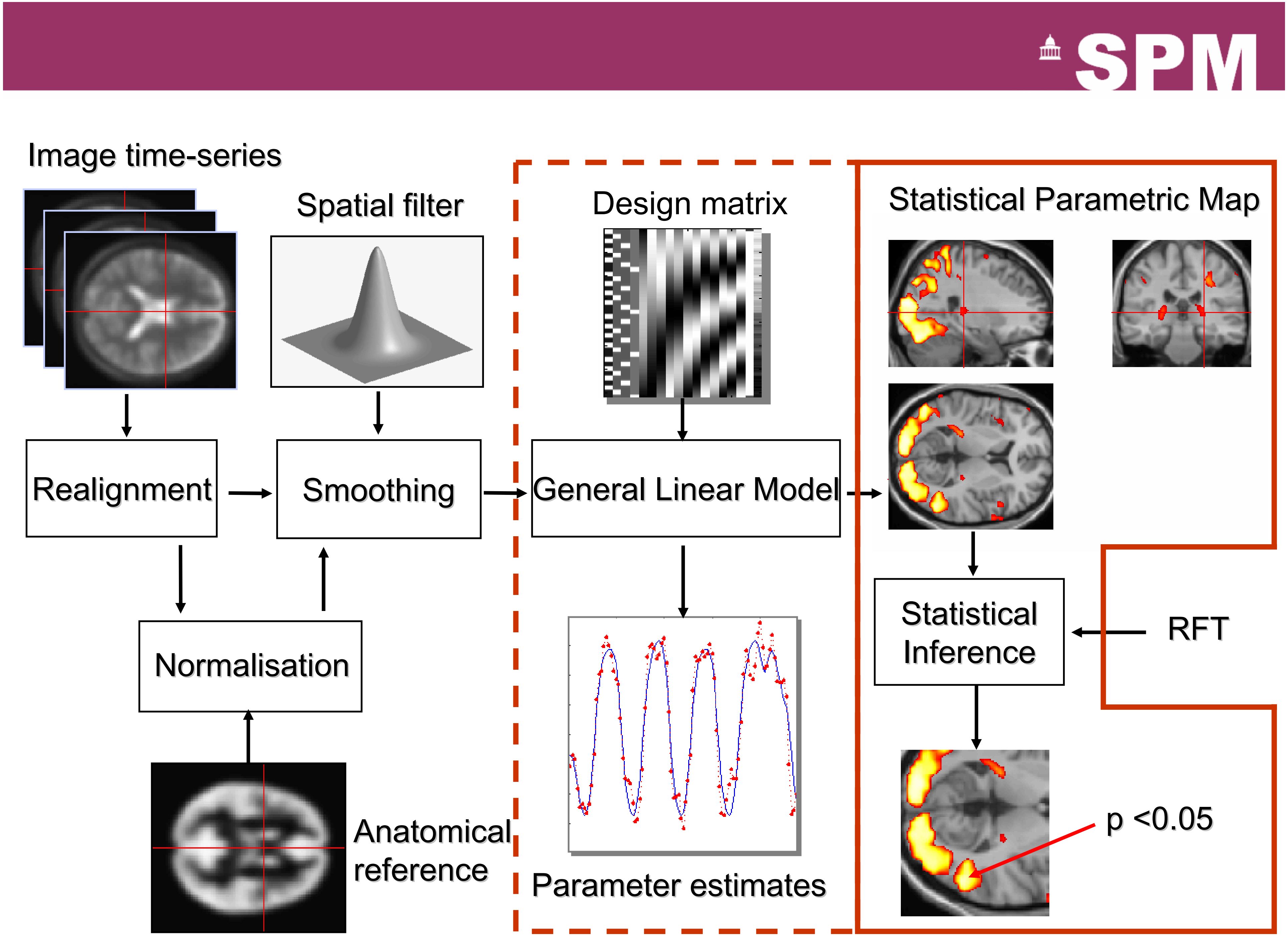
SPM

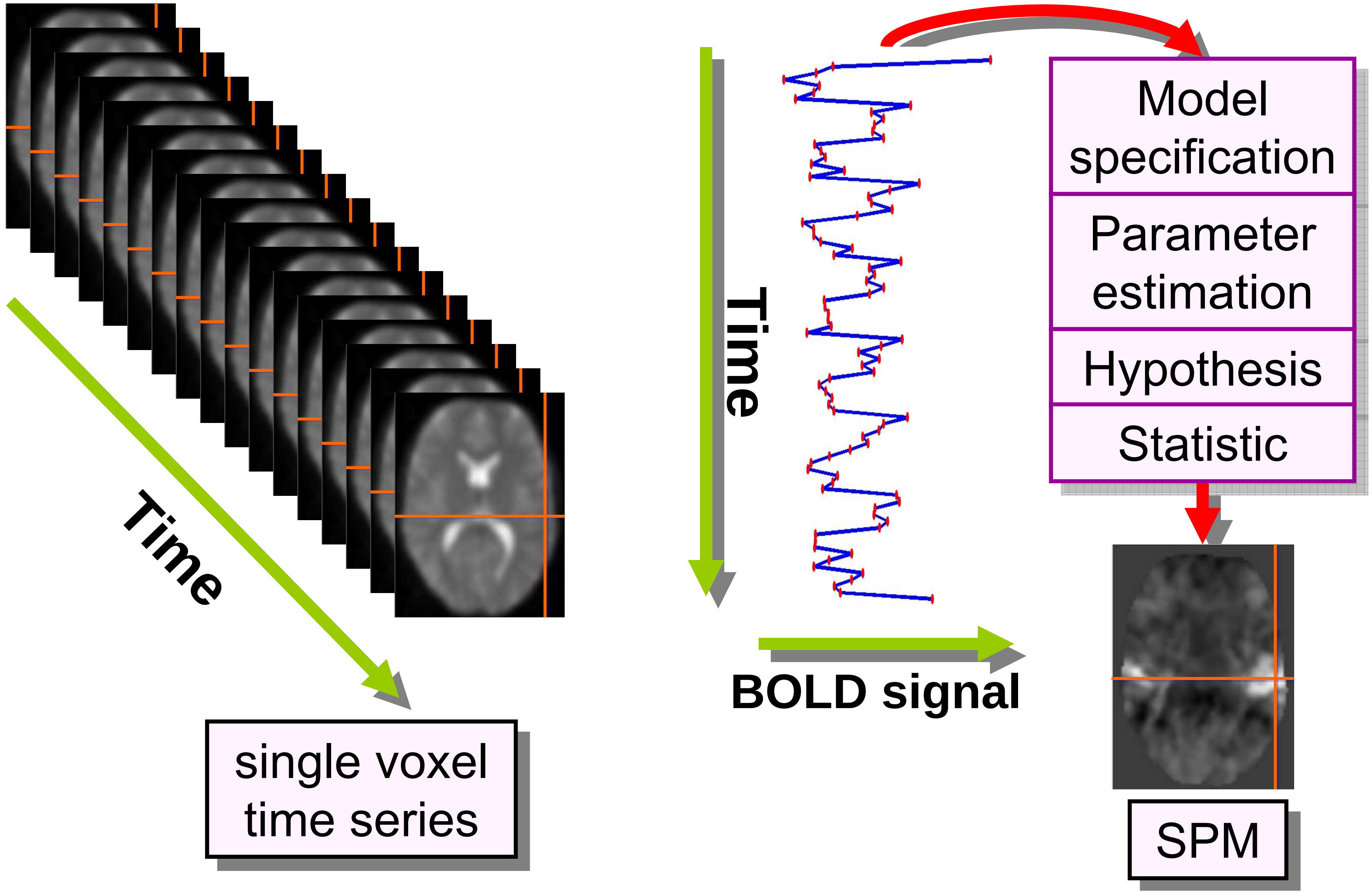
Statistical Inference

Jean Daunizeau
Wellcome Trust Centre for Neuroimaging
University College London

SPM Course
Edinburgh, April 2010


SPM
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

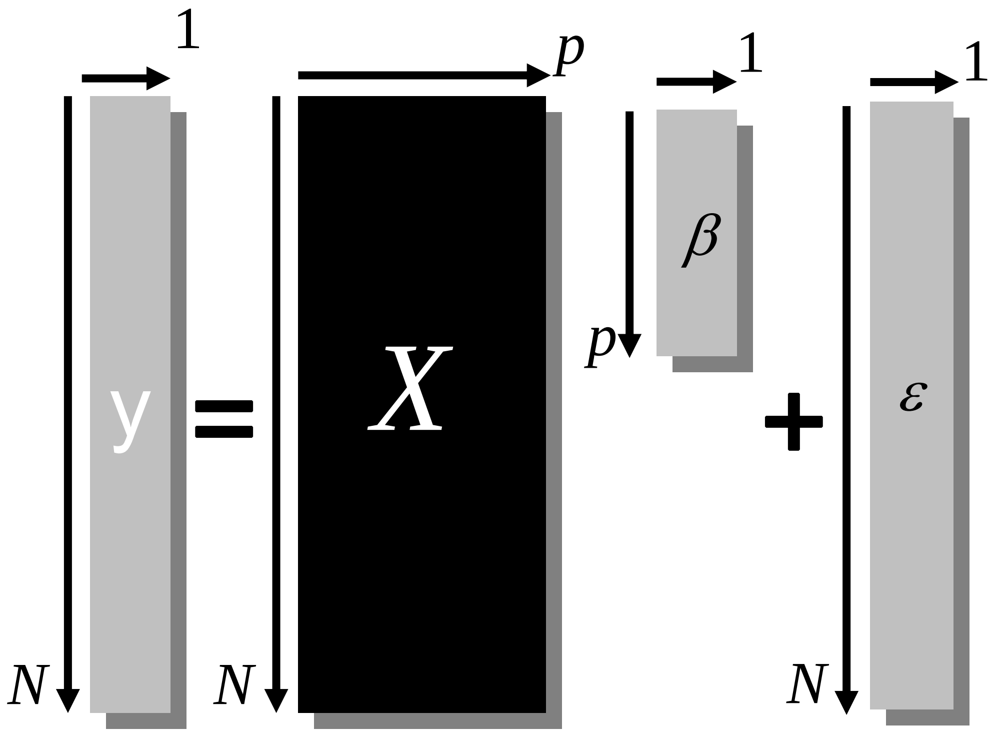
Voxel-wise time series analysis



Overview

- Model specification and parameters estimation
- Hypothesis testing
- Contrasts
  - *T*-tests
  - *F*-tests
- Contrast estimability
- Correlation between regressors
  - Example(s)
- Design efficiency

Model Specification: The General Linear Model



N: number of scans, p: number of regressors

$$y = X\beta + \varepsilon$$

Sphericity assumption:
Independent and identically distributed (i.i.d.) error terms

$$\varepsilon \sim N(0, \sigma^2 I)$$

- The General Linear Model is an equation that expresses the observed response variable in terms of a linear combination of explanatory variables *X* plus a well behaved error term. Each column of the design matrix corresponds to an effect one has built into the experiment or that may confound the results.

Parameter Estimation: Ordinary Least Squares

- Find $\hat{\beta}$ that minimises $\boxed{\|y - X\beta\|^2 = \varepsilon^T \varepsilon}$

- The Ordinary Least Estimates are:

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

- Under i.i.d. assumptions, the Ordinary Least Squares estimates are Maximum Likelihood.

$$\varepsilon \sim N(0, \sigma^2 I) \longrightarrow Y \sim N(X\beta, \sigma^2 I)$$

$$\varepsilon \sim N(0, \sigma^2 I) \longrightarrow \hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$$

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

Hypothesis Testing

To test an hypothesis, we construct "test statistics".

- **The Null Hypothesis H_0**

  Typically what we want to disprove (no effect).

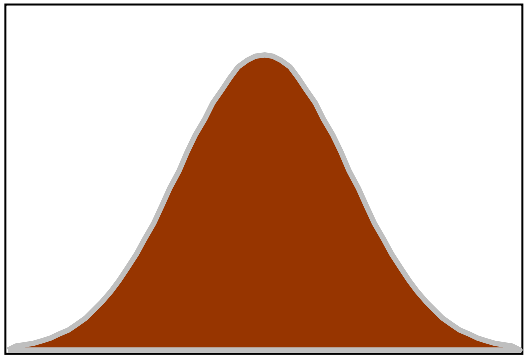
  ⇨ The Alternative Hypothesis H_A expresses outcome of interest.

- **The 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

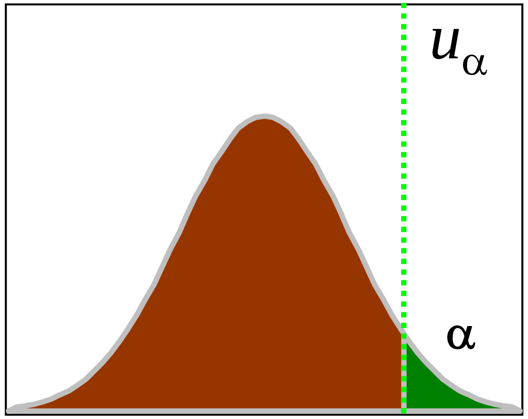
- ***Significance level α:***

  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

Observation of test statistic t, a realisation of T

- **The conclusion about the hypothesis:**

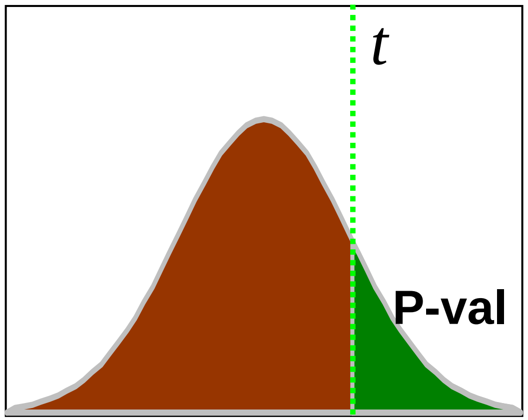
  We reject the null hypothesis in favour of the alternative hypothesis if $t > u_\alpha$

- ***P*-value:**

  A *p*-value summarises evidence against H_0.

  This is the chance of observing value more extreme than *t* under the null hypothesis.

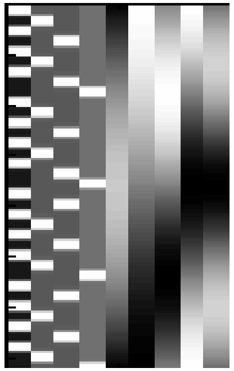
$$\boxed{p(T > t \mid H_0)}$$



Null Distribution of T

Contrasts

- We are usually not interested in the whole β vector.
- 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.



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

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

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

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

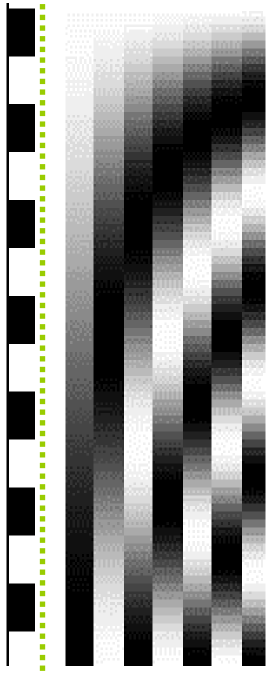
- Under i.i.d assumptions:

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

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

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

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

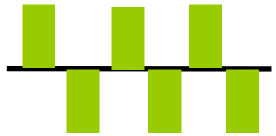


Question: box-car amplitude > 0 ?

=

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

Null hypothesis: H_0: $c^T\beta=0$



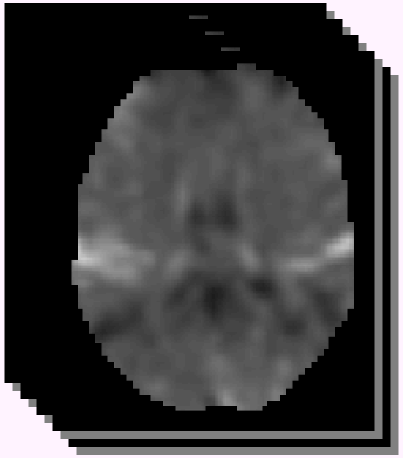
Test statistic:

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

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

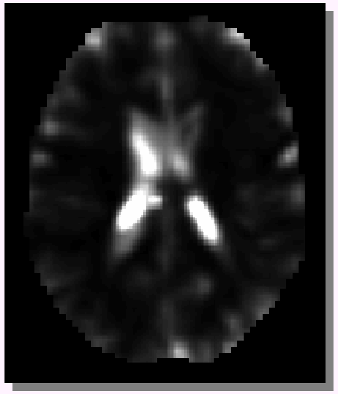
T-contrast in SPM

- For a given contrast *c*:



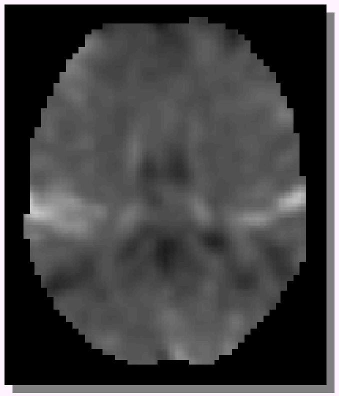
beta_???? images

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



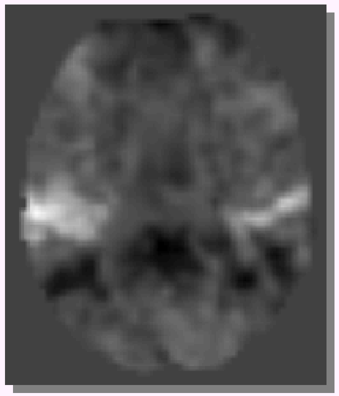
ResMS image

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



con_???? image

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



spmT_???? image

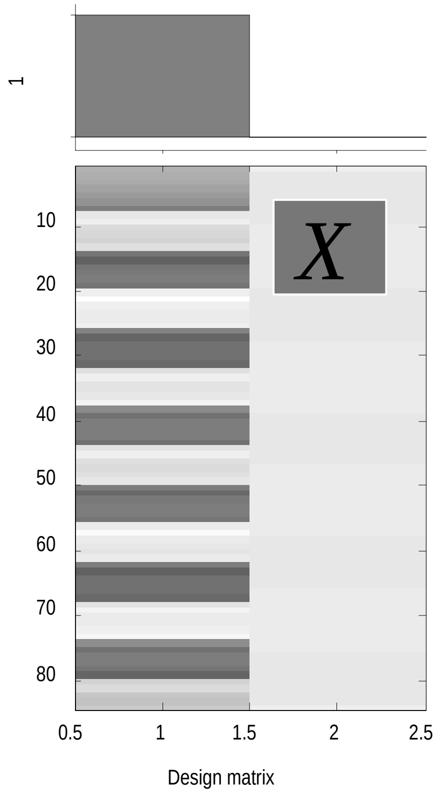
SPM{*t*}

T-test: a simple example

- Passive word listening versus rest

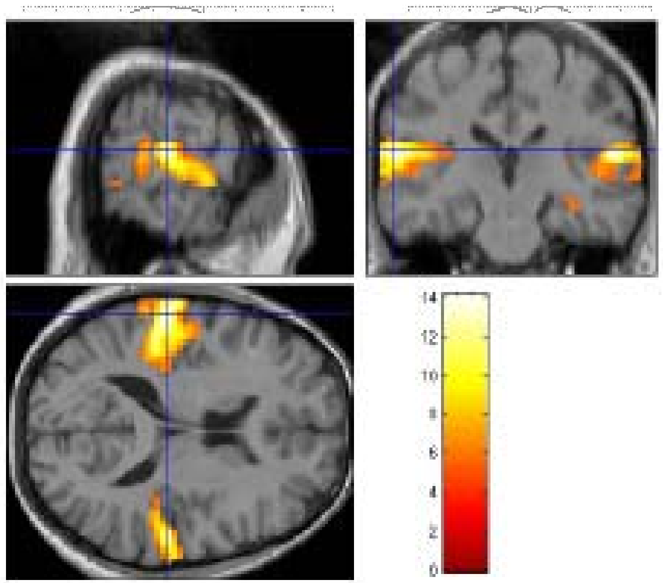
$c^T = [\ 1 \quad 0\ ]$

Q: activation during listening ?



Null hypothesis: $\beta_1 = 0$

$$t = \frac{c^T \hat{\beta}}{Std(c^T \hat{\beta})}$$



SPMresults:
Height threshold T = 3.2057 {p<0.001}

Statistics: p-values adjusted for search volume

voxel-level T	(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
6.19	5.53	0.000	-30	-33	-18
5.96	5.36	0.000	36	-27	9
5.84	5.27	0.000	-45	42	9
5.44	4.97	0.000	48	27	24
5.32	4.87	0.000	36	-27	42

***T*-test:** a few remarks

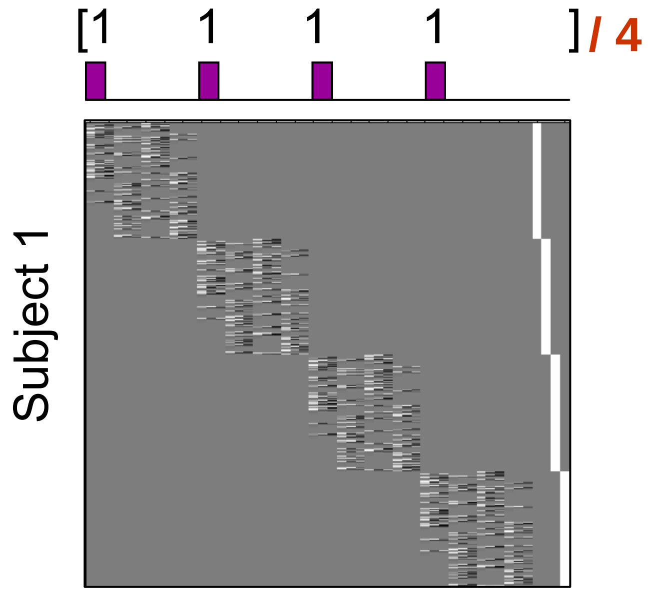
- *T*-test is a signal-to-noise measure (ratio of estimate to standard deviation of estimate).

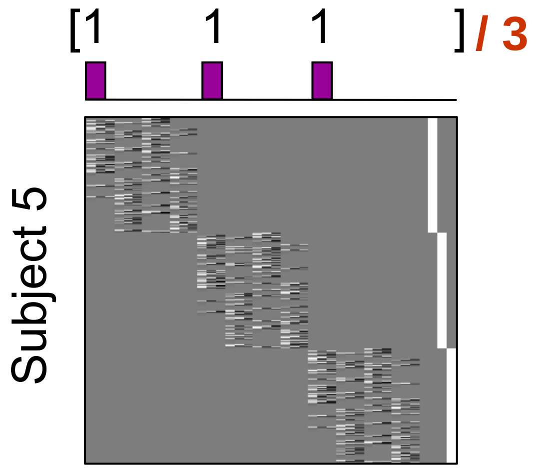
- *T*-contrasts are simple combinations of the betas; the T-statistic does not depend on the scaling of the regressors or the scaling of the contrast.

- Unilateral test:

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

Scaling issue





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

- 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):

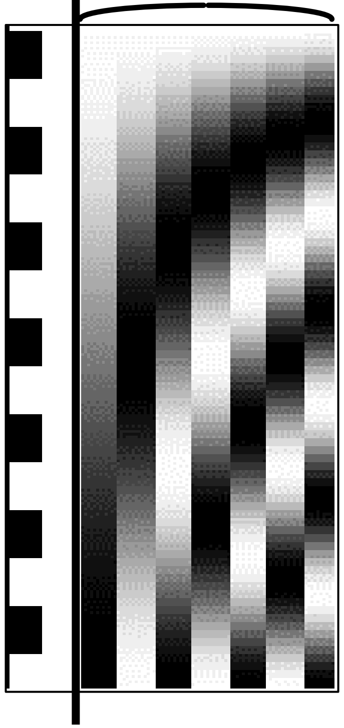
sum ≠ average

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

- Model comparison: *Full vs. Reduced model?*

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

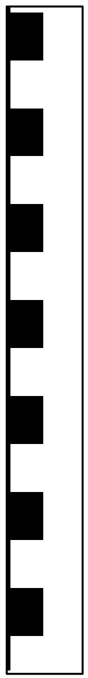
X_0 X_1



Full model ?

→ **RSS** $\sum \hat{\varepsilon}^2_{full}$

X_0



Or Reduced model?

→ **RSS$_0$** $\sum \hat{\varepsilon}^2_{reduced}$

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

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

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

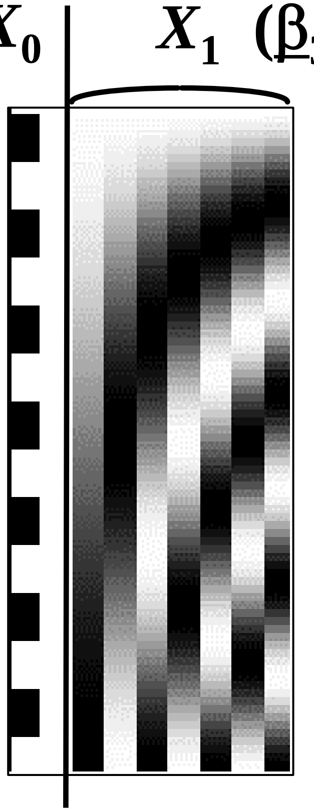
$\nu_1 = \mathrm{rank}(X) - \mathrm{rank}(X_0)$
$\nu_2 = N - \mathrm{rank}(X)$

F-test - multidimensional contrasts – SPM{F}

- Tests multiple linear hypotheses:

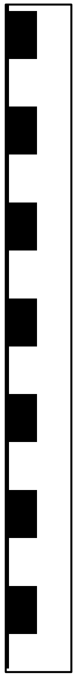
H₀: True model is X_0 | **H₀**: $\beta_3 = \beta_4 = \ldots = \beta_9 = 0$ | **test H₀** : $c^T\beta = 0$?

X_0 X_1 ($\beta_{3\text{-}9}$)



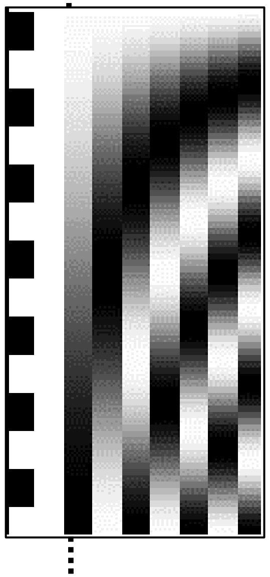
Full model?

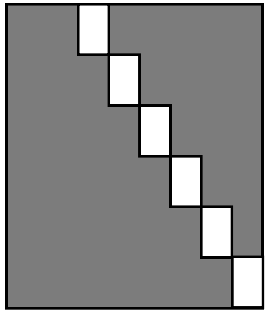
X_0

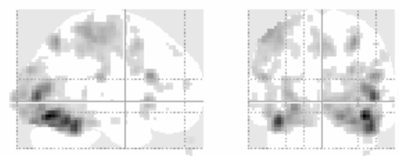


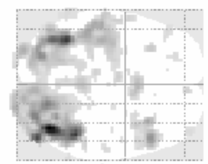
Reduced model?

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



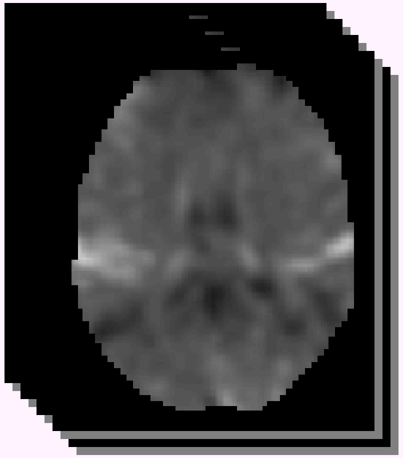






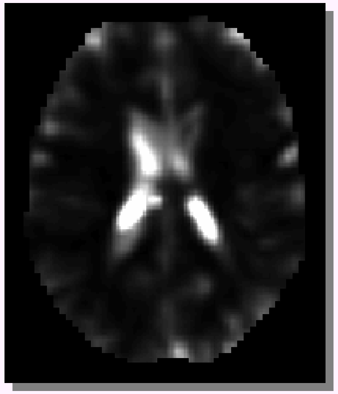
SPM{$F_{6,322}$}

F-contrast in SPM



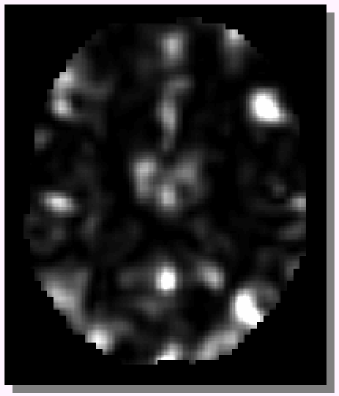
beta_???? images

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



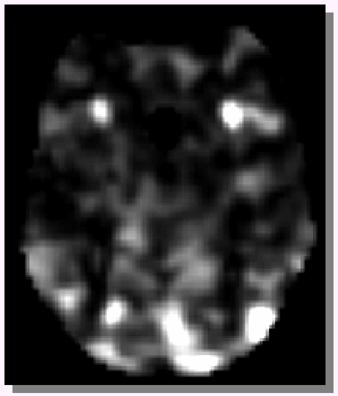
ResMS image

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



ess_???? images

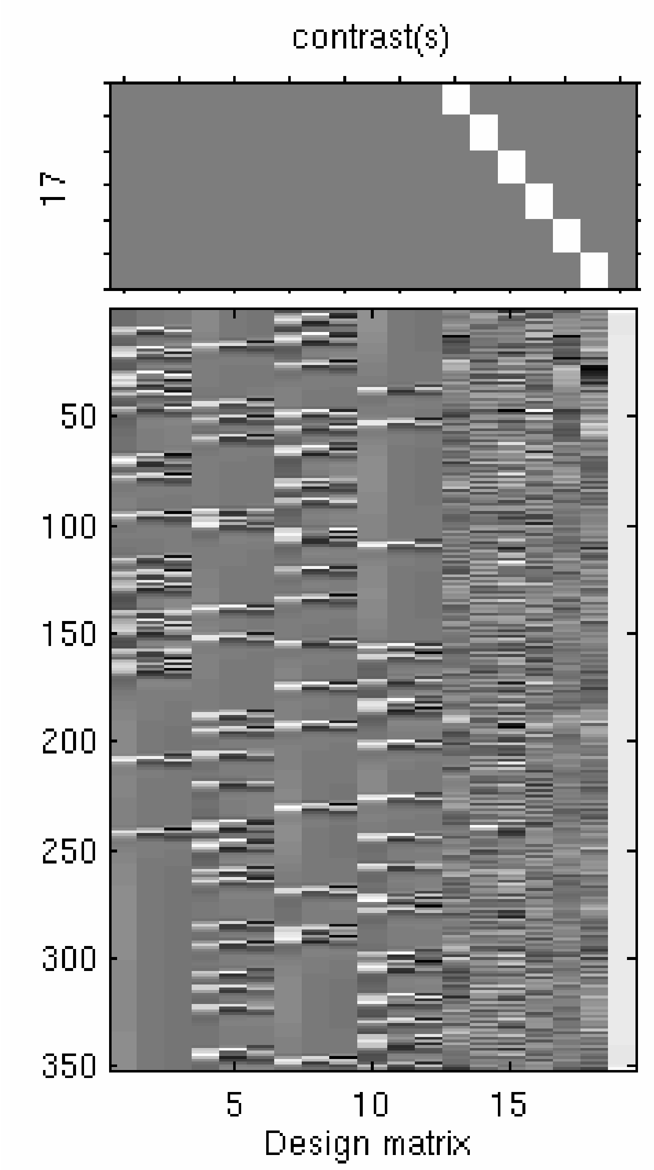
$$(RSS_0 \ - \ RSS)$$

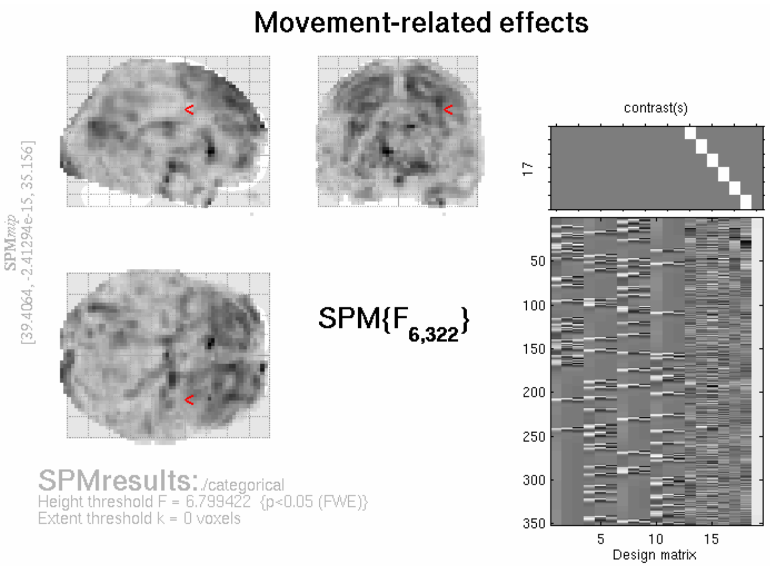


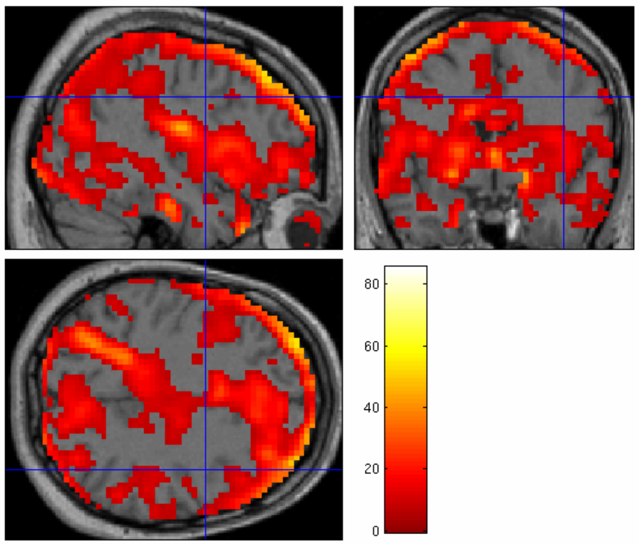
spmF_???? images

SPM{F}

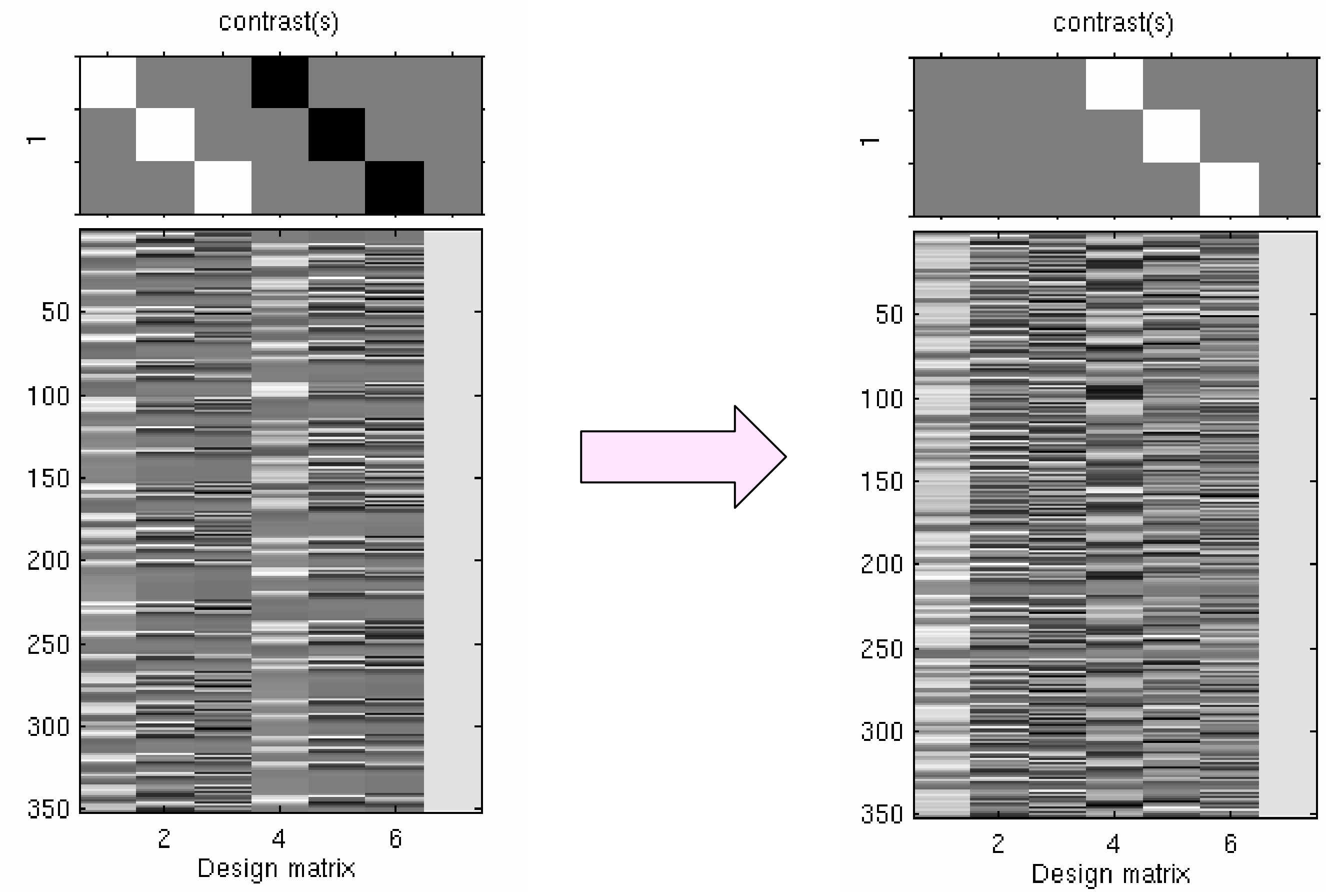
F-test example: movement related effects







Multidimensional contrasts



Think of it as constructing 3 regressors from the 3 differences and complement this new design matrix such that data can be fitted in the same exact way (same error, same fitted data).

F-test: a few remarks

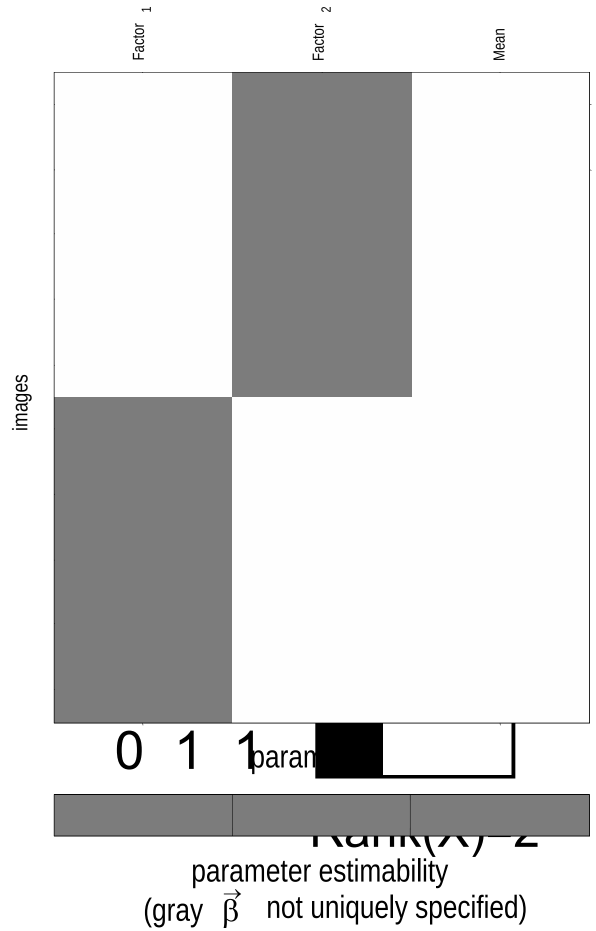
- F-tests can be viewed as testing for the additional variance explained by a larger model wrt a simpler (nested) model ⇨ Model comparison.
- F tests a weighted sum of squares of one or several combinations of the regression coefficients β.
- In practice, we don't have to explicitly separate X into $[X_1 X_2]$ 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.

Estimability of a contrast

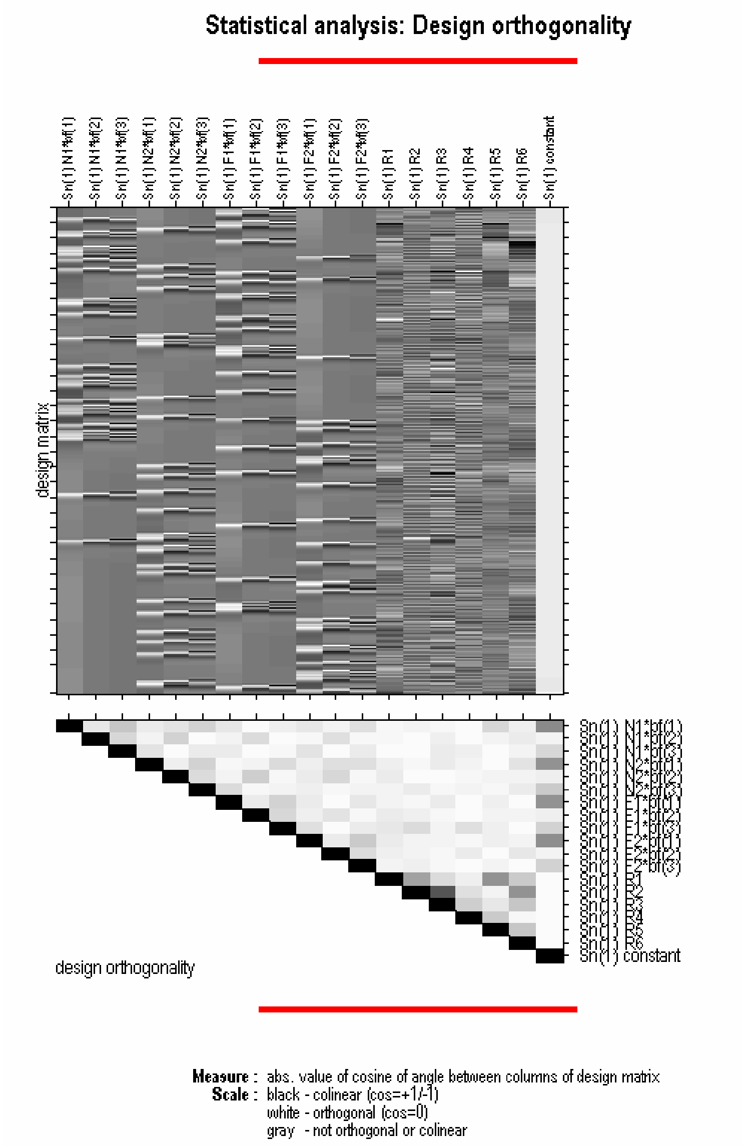
- If X is not of **full rank** then we can have $X\beta_1 = X\beta_2$ with $\beta_1 \neq \beta_2$ (different parameters).
- The parameters are **not** therefore 'unique', 'identifiable' or '**estimable**'.
- For such models, X^TX is not invertible so we must resort to generalised inverses (SPM uses the **pseudo-inverse**).



- Example:

[1 0 0], [0 1 0], [0 0 1] are not estimable.
[1 0 1], [0 1 1], [1 -1 0], [0.5 0.5 1] are estimable.

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.

- The cosine of the angle between two vectors *a* and *b* is obtained by:

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

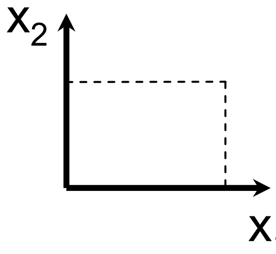
- If both vectors have **zero mean** then the cosine of the angle between the vectors is the same as the **correlation** between the two variates.

Multicollinearity

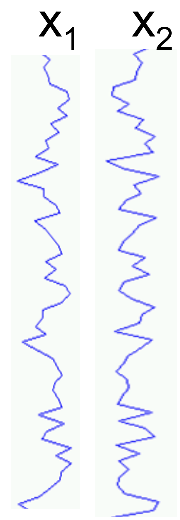
$$Var(c^T \hat{\beta}) = \sigma^2 c^T (X^T X)^{-1} c$$

- Orthogonal regressors (=*uncorrelated*):

  By varying each separately, one can predict the combined effect of varying them jointly.

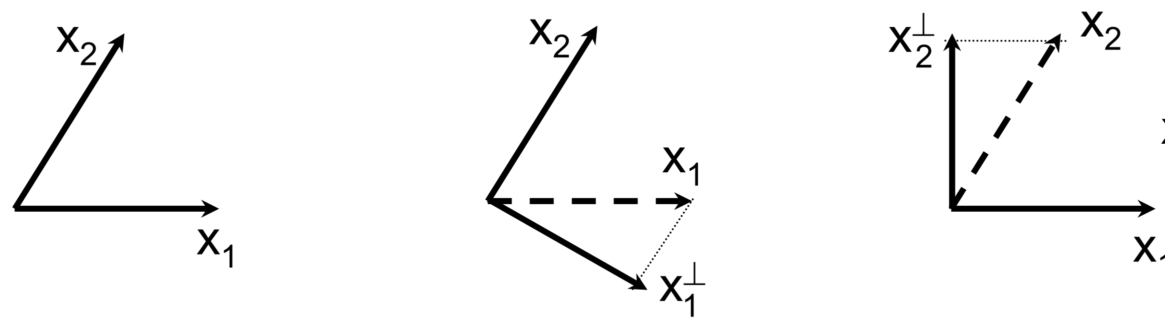


$(X^T X)^{-1}$ is diagonal



- Non-orthogonal regressors (=*correlated*):

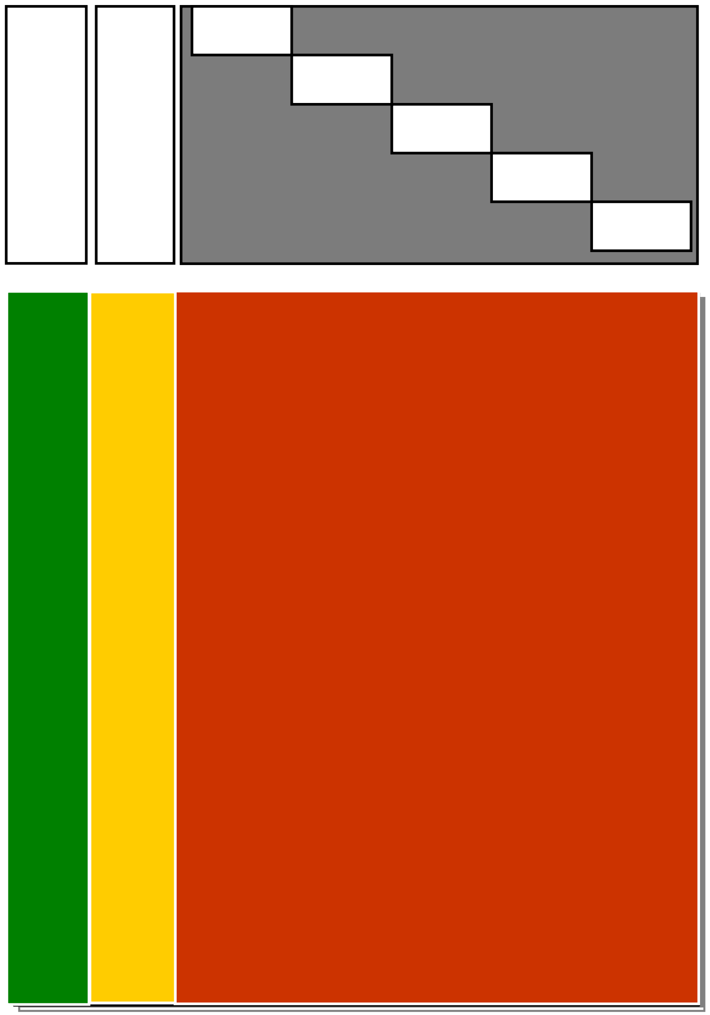
  When testing for the first regressor, we are effectively removing the part of the signal that can be accounted for by the second regressor ⇨ *implicit orthogonalisation.*

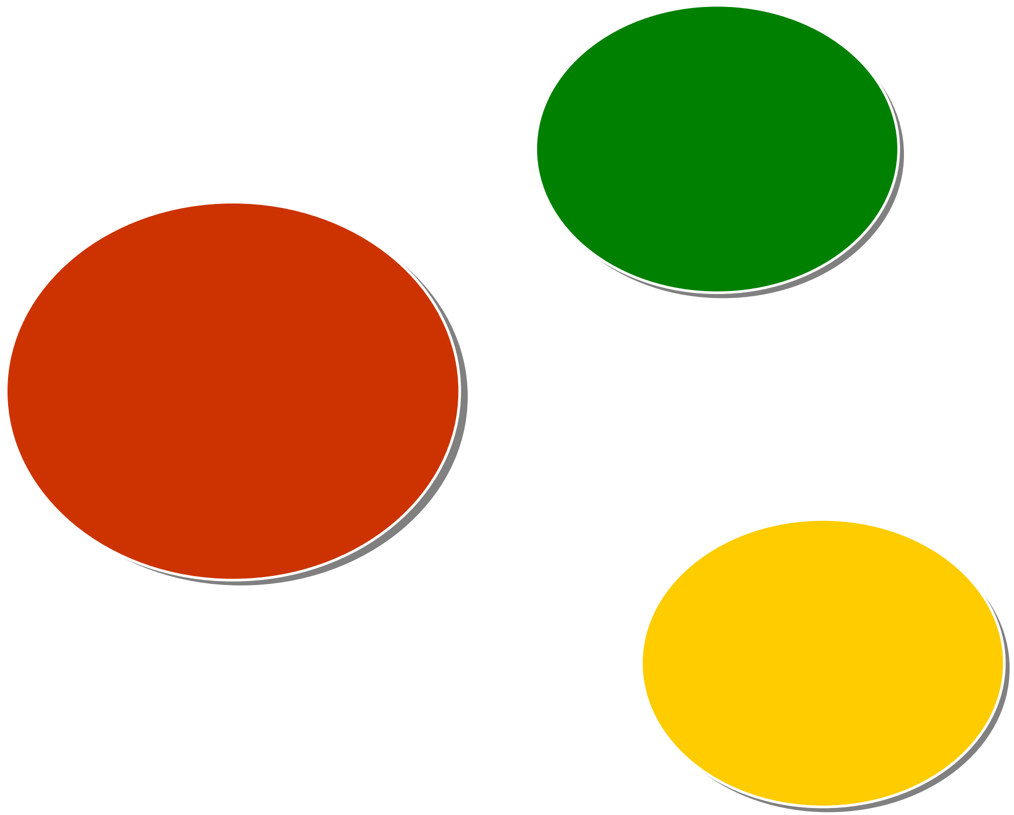


$x_2^\perp = x_2 - x_1.x_2\, x_1$

It does not reduce the predictive power or reliability of the model as a whole.

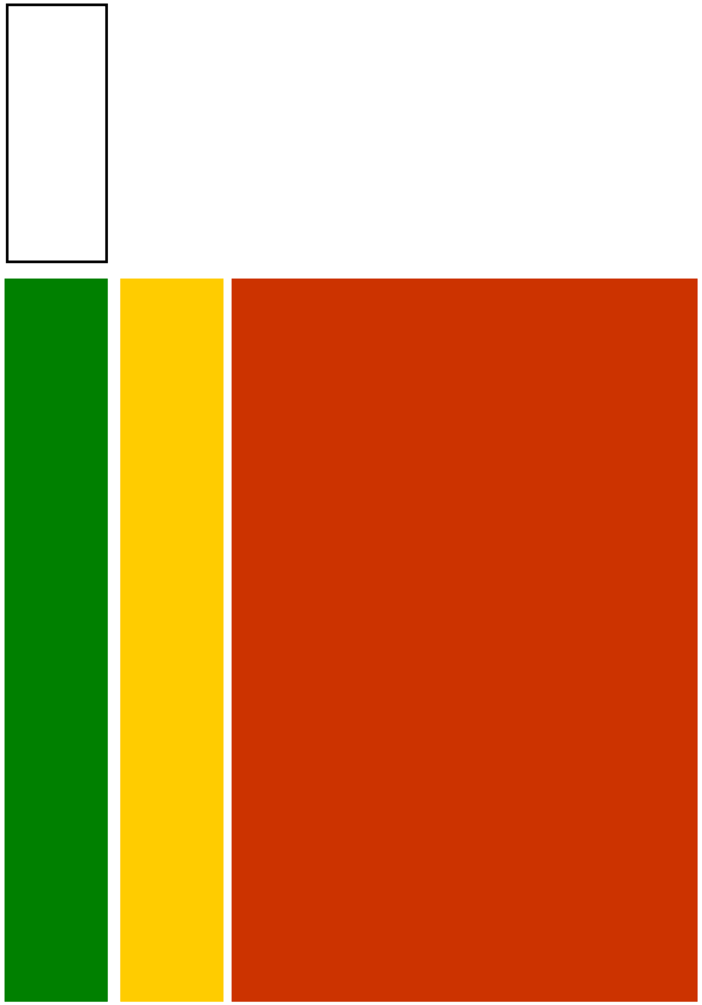
Shared variance



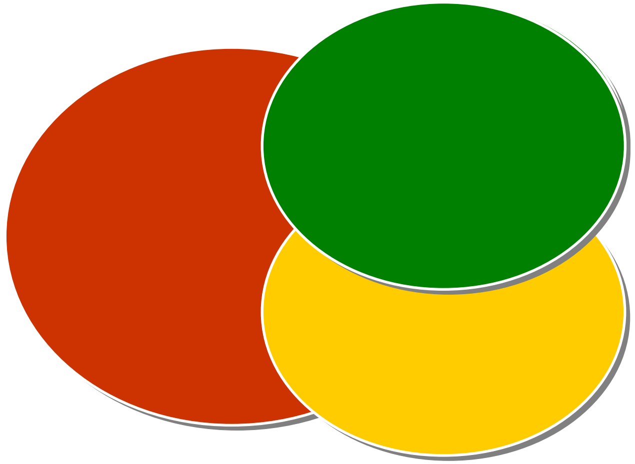


Orthogonal regressors.

Shared variance



Testing for the green:

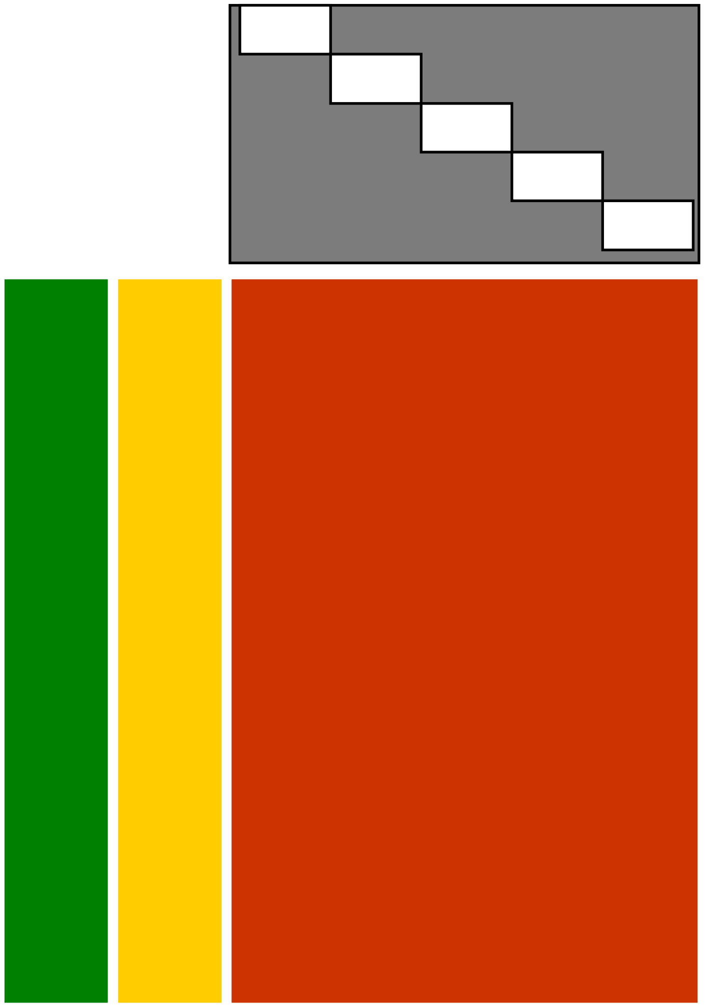


Correlated regressors, for example:

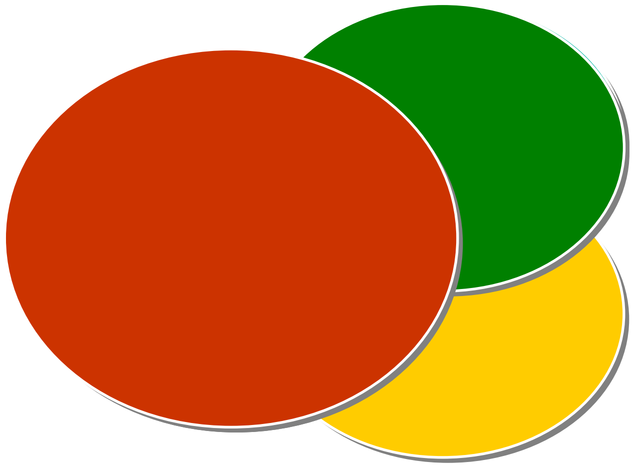
green: subject age

yellow: subject score

Shared variance

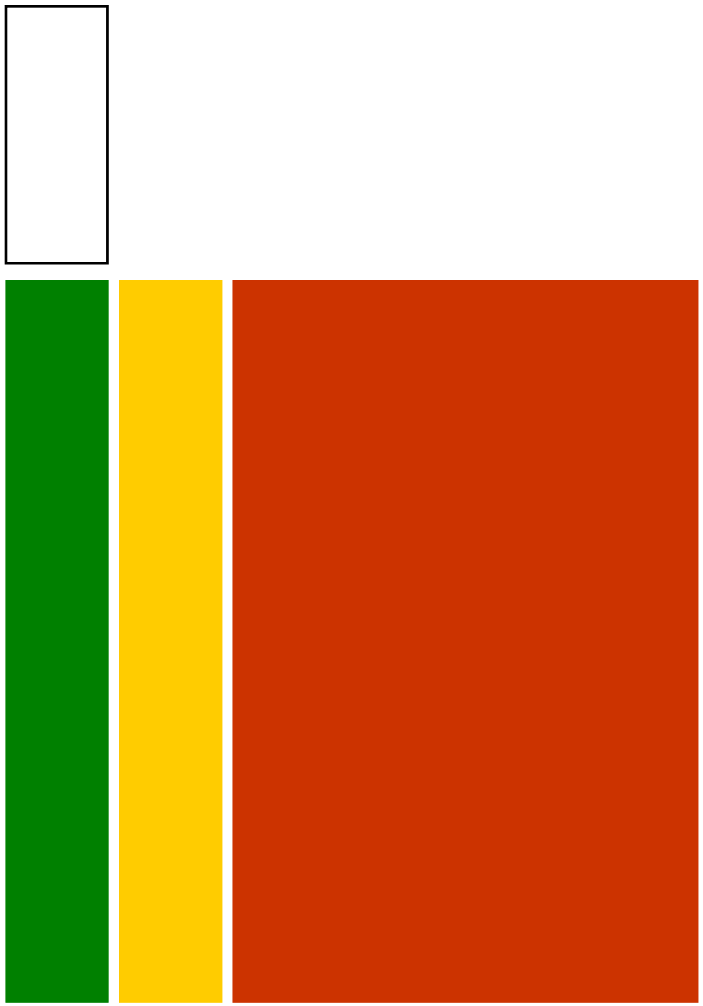


Testing for the red:

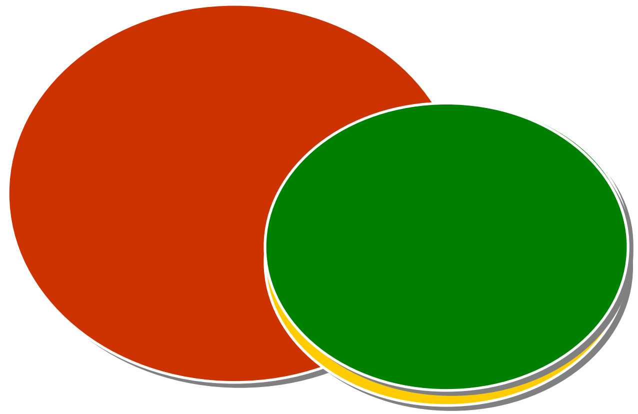


Correlated regressors.

Shared variance



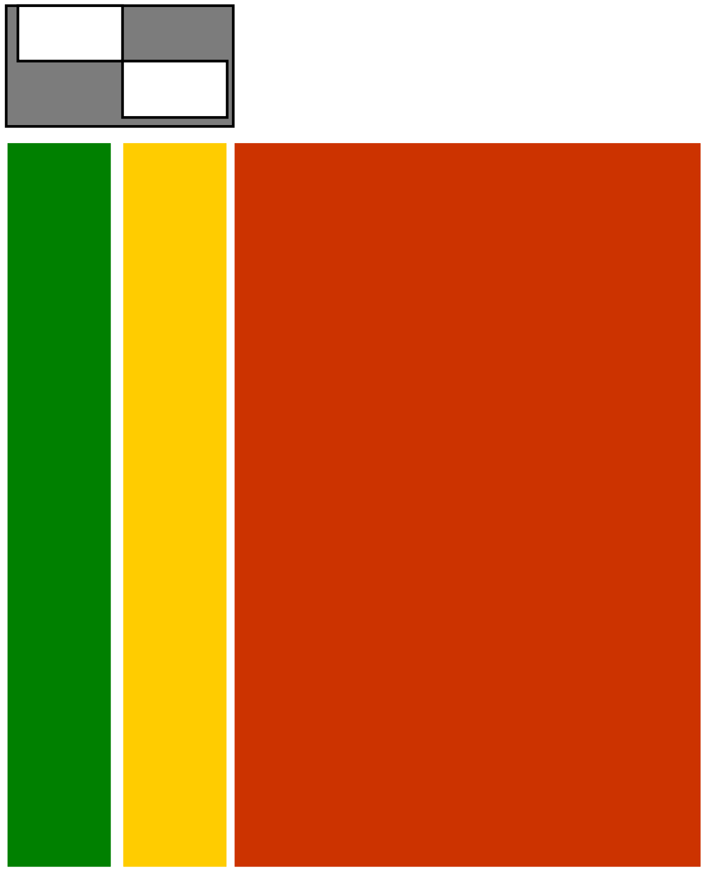
Testing for the green:

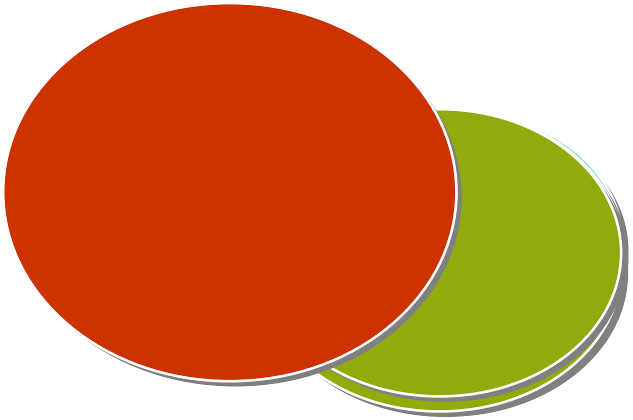


Highly correlated.
Entirely correlated ⇨ non estimable

Shared variance

Testing for the green and yellow





If significant, can be G and/or Y

A few remarks

- We implicitly test for an additional effect only, be careful if there is correlation
  - Orthogonalisation = decorrelation : not generally needed
  - Parameters and test on the non modified regressor change

- It is always simpler to have orthogonal regressors and therefore designs.

- In case of correlation, use F-tests to see the overall significance. There is generally no way to decide to which regressor the « common » part should be attributed to.

- 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

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

- 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:

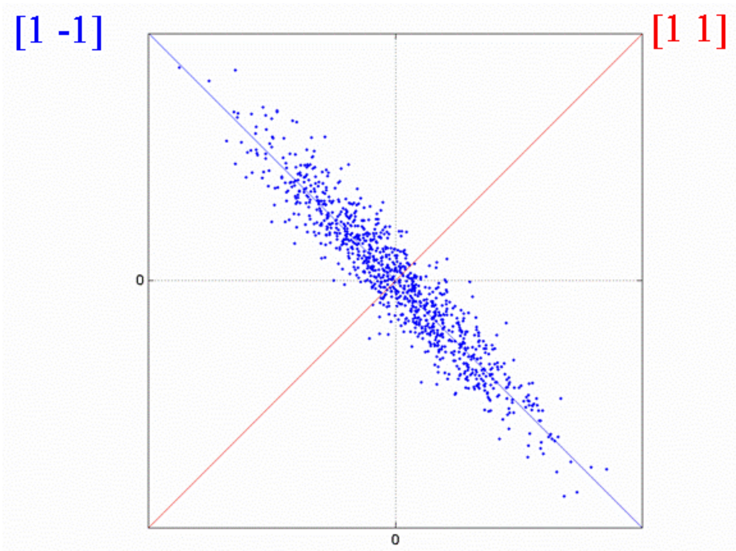
$$e(c, X) = (c^T (X^T X)^{-1} c)^{-1}$$

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

Design efficiency

- The efficiency of an estimator is a measure of how reliable it is and depends on error variance (the variance not modeled by explanatory variables in the design matrix) and the design variance (a function of the explanatory variables and the contrast tested).
- $X^T X$ represents covariance of regressors in design matrix; high covariance increases elements of $(X^T X)^{-1}$.
- High correlation between regressors leads to low sensitivity to each regressor alone.

$$\boxed{c^T (X^T X)^{-1} c}$$



$$\begin{bmatrix} 1 & -0.9 \\ -0.9 & 1 \end{bmatrix}$$

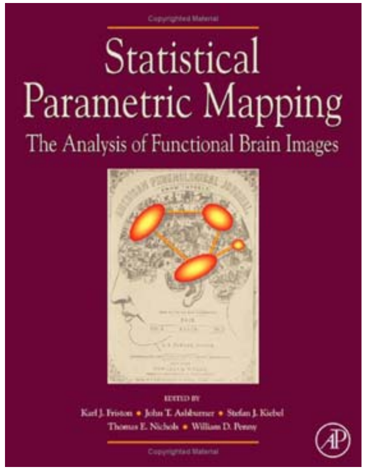
c^T=[1 0]: 5.26

c^T=[1 1]: 20

c^T=[1 -1]: 1.05

SPM

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.

With many thanks to G. Flandin, J.-B. Poline and Tom Nichols for slides.