Classical inference in neuroimaging

fMRI methods & models 2016

Justin Chumbley

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

Many thanks to Jacob Heinzle, 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

Last week:
- "Best" estimate
- Univariate

This week:
- Generalization out-of-sample
- Large-scale, multivariate

Big idea
- Problem with last week's "best fit": overfitting

In practise
- Tests, tests, tests (z, t, F)
- Tradeoffs (type-I/II)

Overview

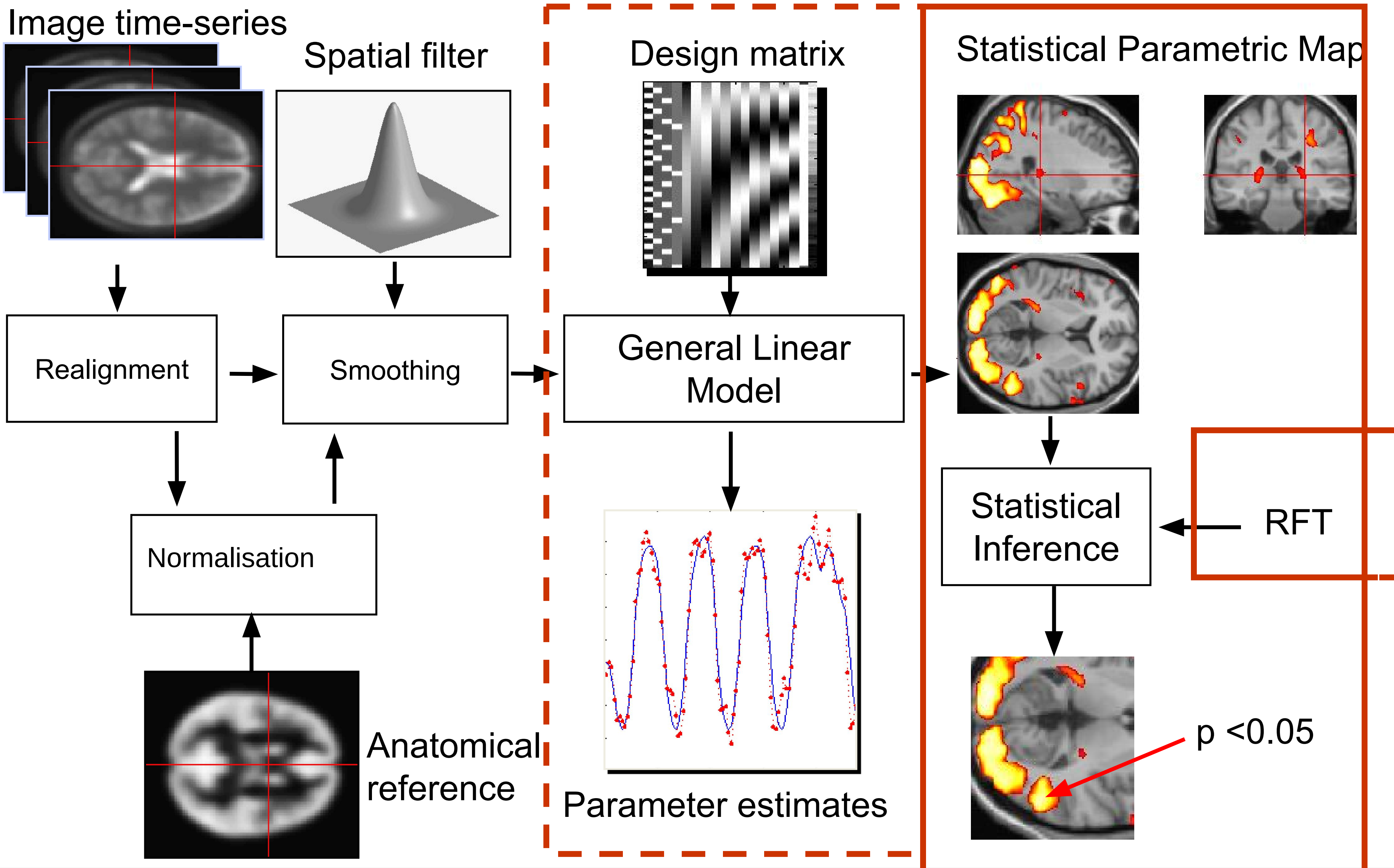
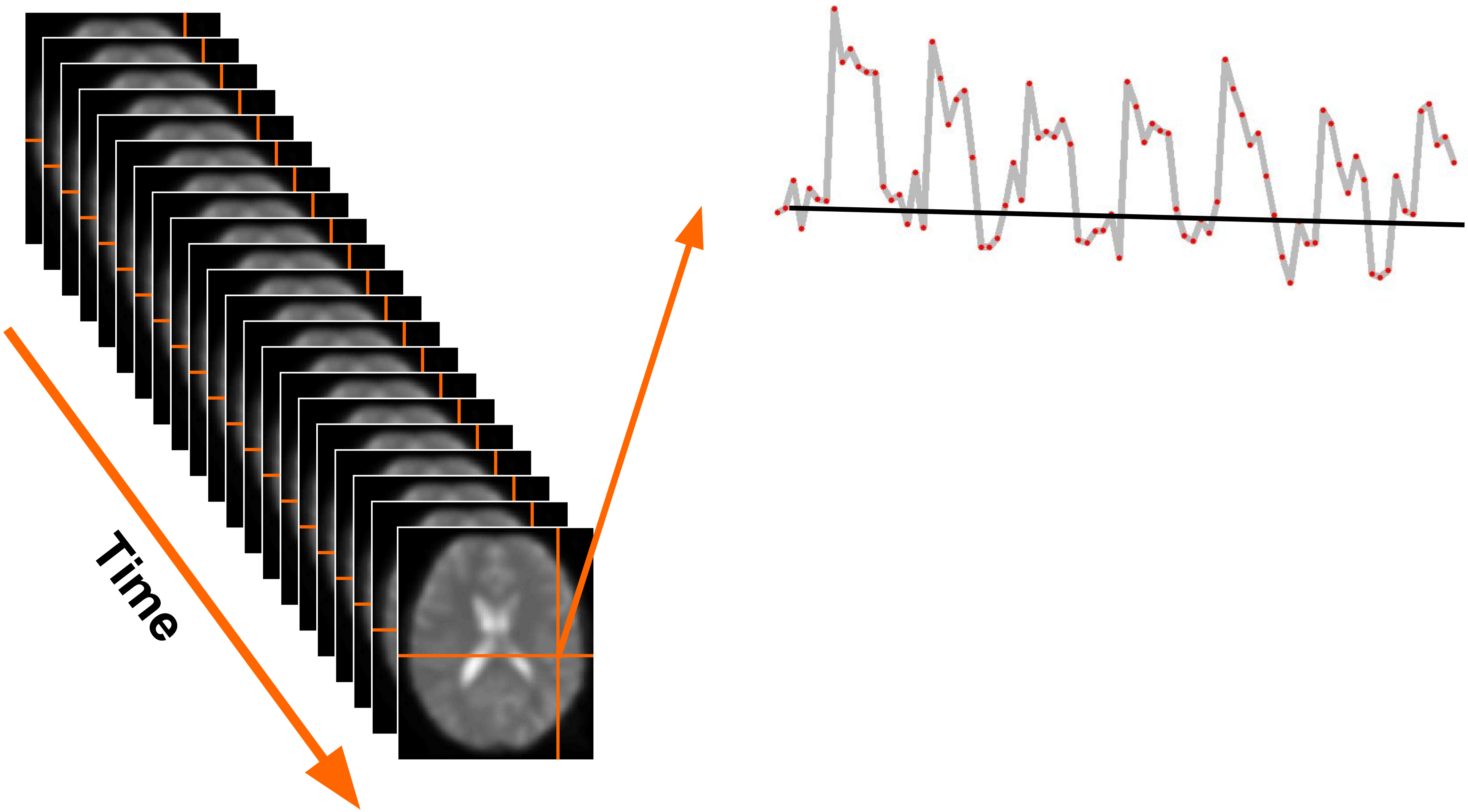
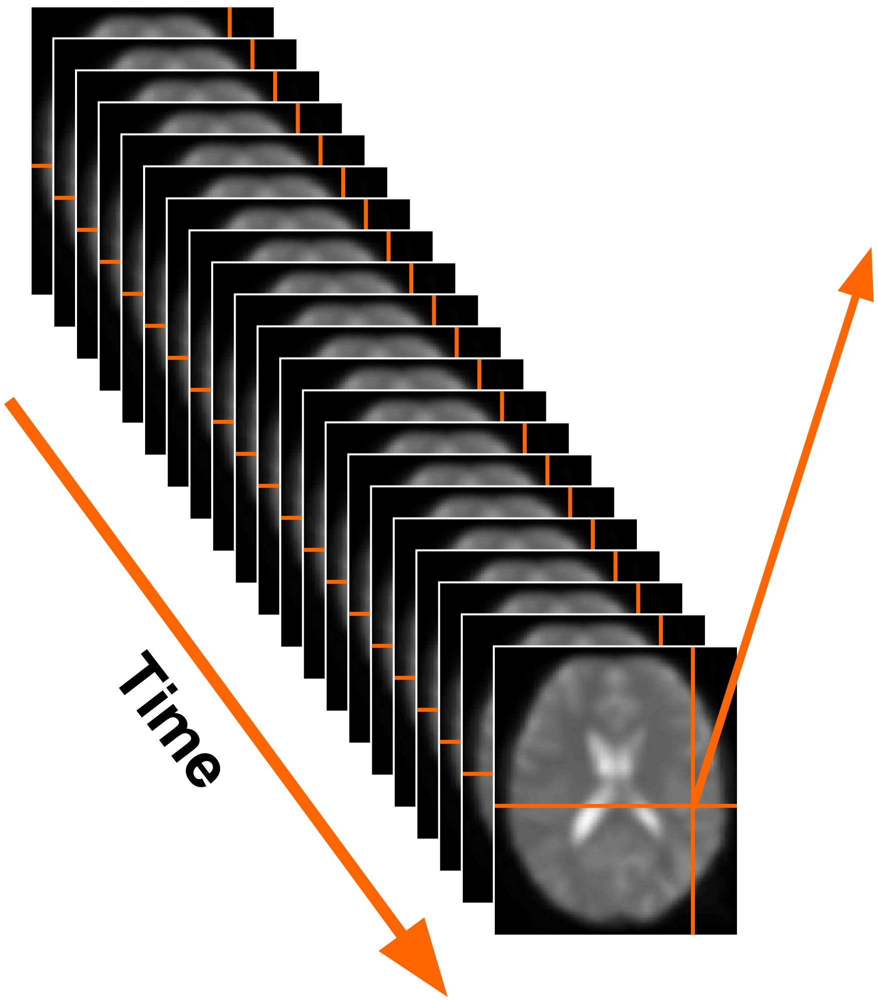


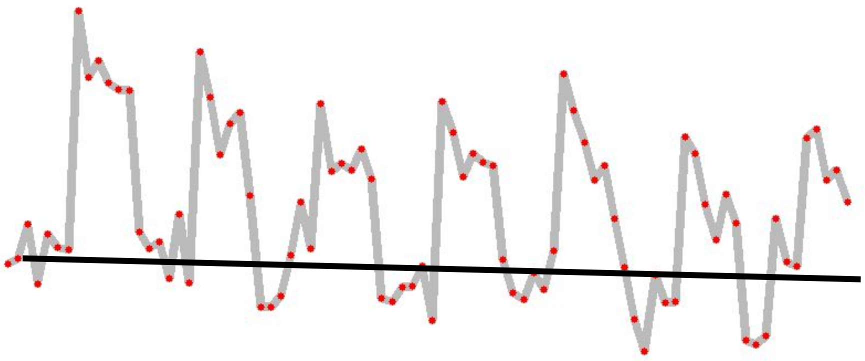
Image time-series
Spatial filter
Design matrix
Statistical Parametric Map
Realignment
Smoothing
General Linear Model
Statistical Inference
RFT
Normalisation
Anatomical reference
Parameter estimates
p <0.05

Univariate regression, single voxel



Univariate regression, single voxel



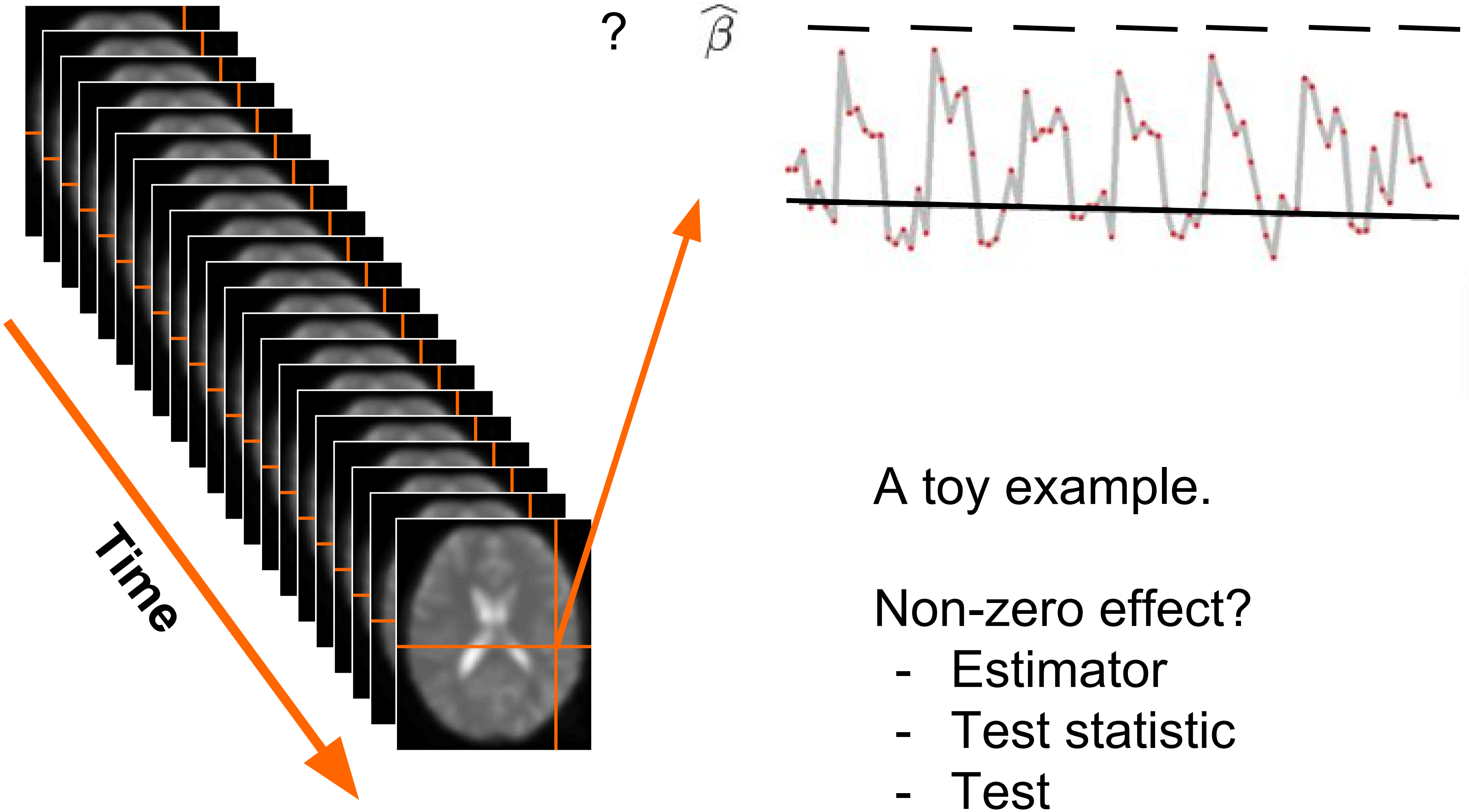


A toy example.

Non-zero effect?

- Estimator
- Test statistic
- Test

Univariate regression, single voxel

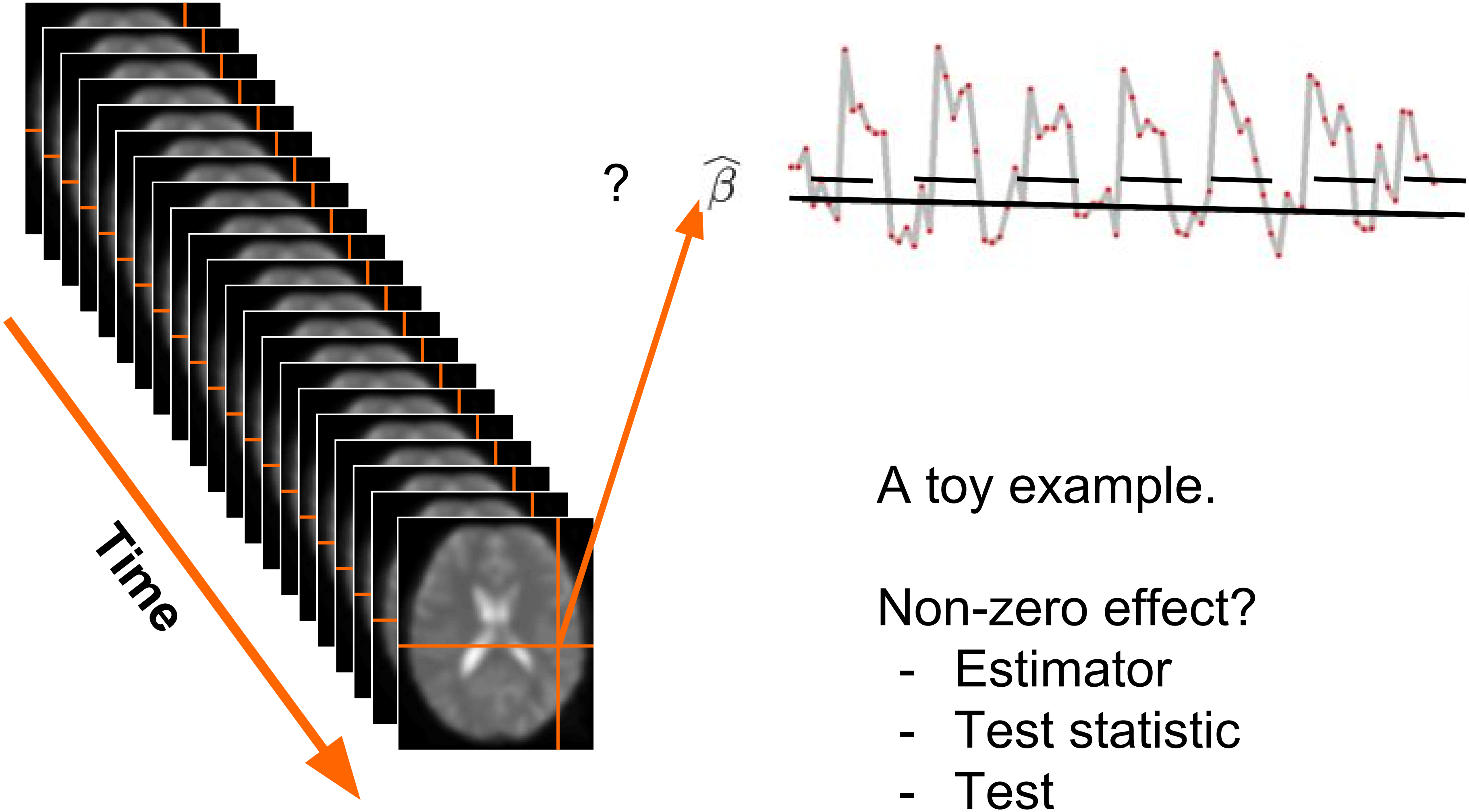


A toy example.

Non-zero effect?

- Estimator
- Test statistic
- Test

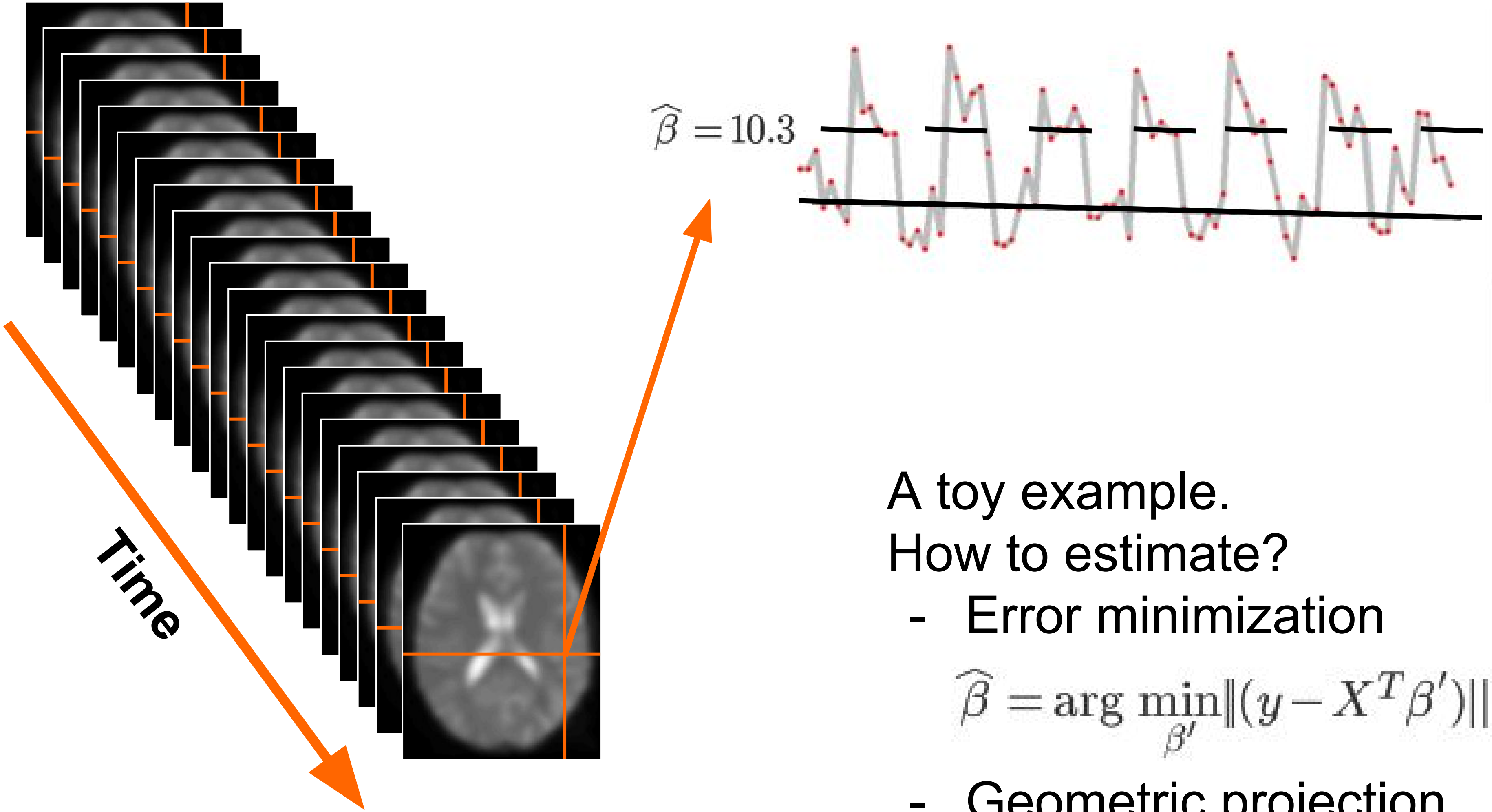
Univariate regression, single voxel



A toy example.

Non-zero effect?
- Estimator
- Test statistic
- Test

Univariate regression, single voxel



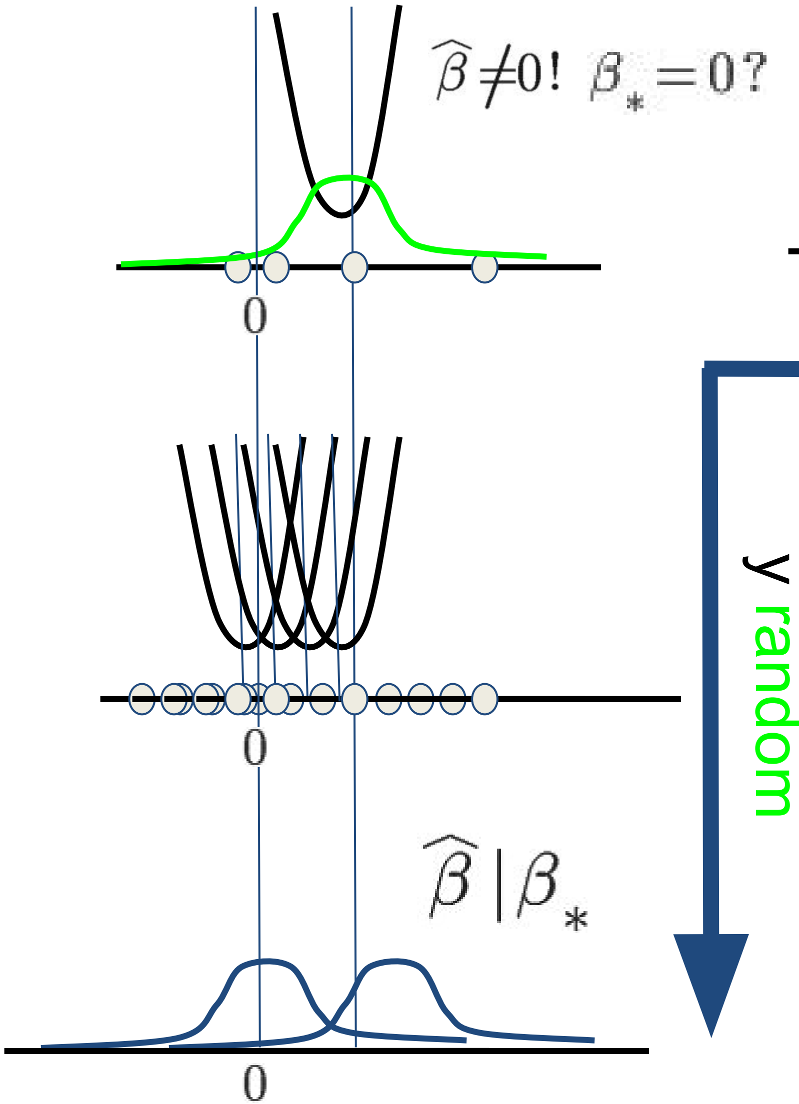
A toy example.
How to estimate?

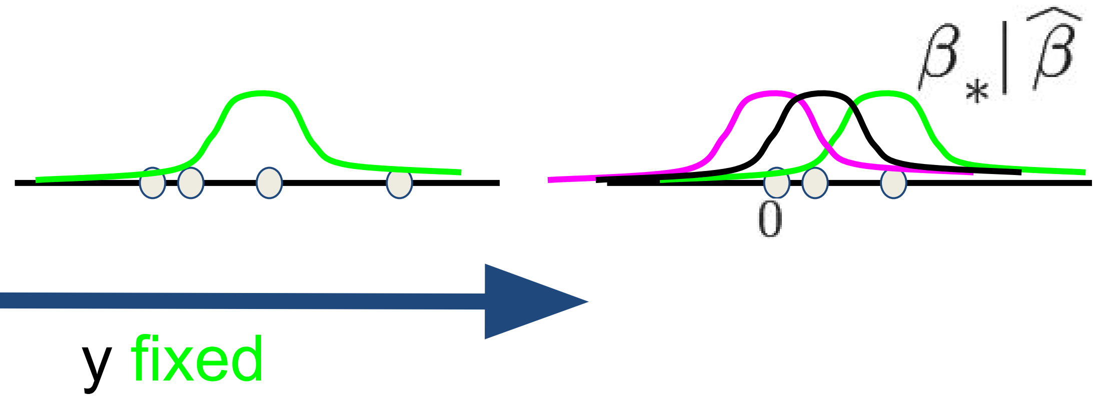
- Error minimization

$$\widehat{\beta} = \arg\min_{\beta'} \|(y - X^T\beta')\|_2^2$$

- Geometric projection
- Maximum likelihood

Univariate cartoon





Overfitting: $\widehat{\beta} \neq 0!$ with 100%,

Reduce overfitting?

- Type-I penalties
- Population prior penalties
- Problem in high dimensions
- Tradeoffs

Math univariate

$$\hat{\beta} = \arg\min_{\beta'} ||(y - X\beta')||_2^2 = (X^T X)^{-1} X^T y$$

$$Infer\ \beta\ from\ \hat{\beta}\ ?$$

$$Y = X\beta + \epsilon,\ \ \epsilon \sim N(0, \sigma^2 I)$$

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

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

$$Assume\ X = \mathbf{1}_n$$

$$\hat{\beta} = \beta + \frac{1}{n} \sum_{i=1}^{n} \epsilon_i$$

Systematic bias?

$$Assume\ \beta = 0$$

$$\frac{\hat{\beta}}{\sigma/\sqrt{n}} = \frac{\frac{1}{n}\sum_{i=1}^{n} \epsilon_i}{\sigma/\sqrt{n}} \sim N(0, 1)$$

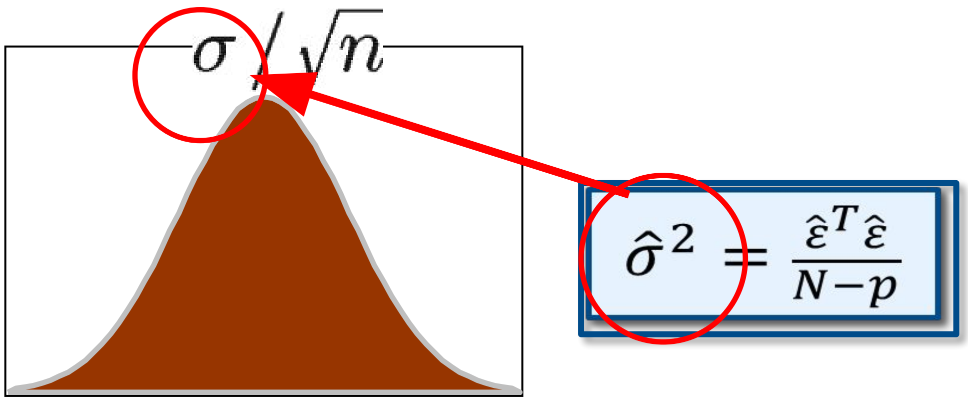
Random variability?

- We’d ideally eliminate bias and variance
- Typically tradeoffs (Examples: confounding bias, multiplicity bias)
- Conventional hierarchy of errors

Univariate hypothesis testing

H_0 predicts “small” T (relative to $\sigma / \sqrt{n}$).

Is T surprisingly big?



Null Distribution of “Test statistic” T

- Scale of ϵ depends on design X (number of replicates and explaining variation from independent causes)
- May be reduced.
- May be estimated via $\hat{\epsilon}$

Univariate hypothesis testing

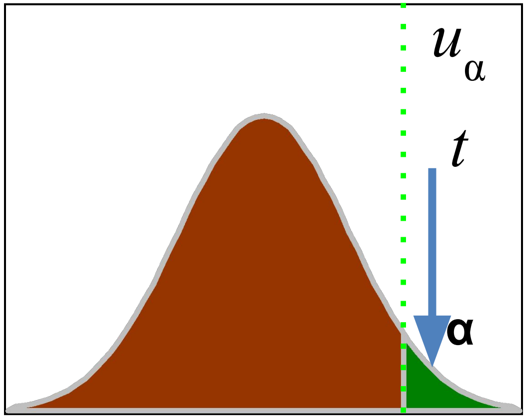
- ***Test construction***

Significance level α:

Acceptable α ⇨ threshold u_α

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

Procedure: if $t > u_\alpha$, then reject $\mathbf{H_0}$



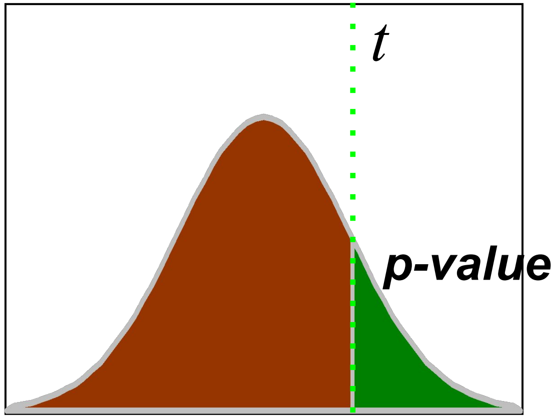
Null Distribution of T

- ***P-value***:

Summarises evidence against H_0.

Chance of observing a value more extreme than t under the null hypothesis.

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



Null Distribution of T

Univariate hypothesis testing

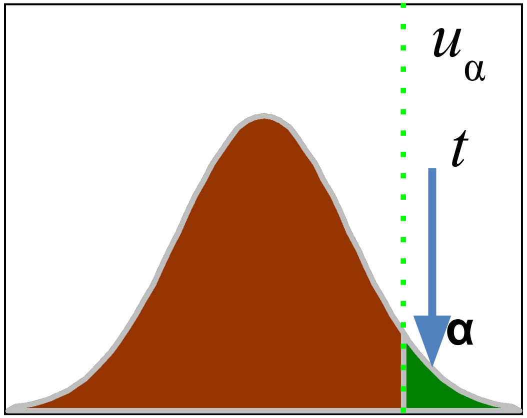
- ***Test construction***

Significance level α:

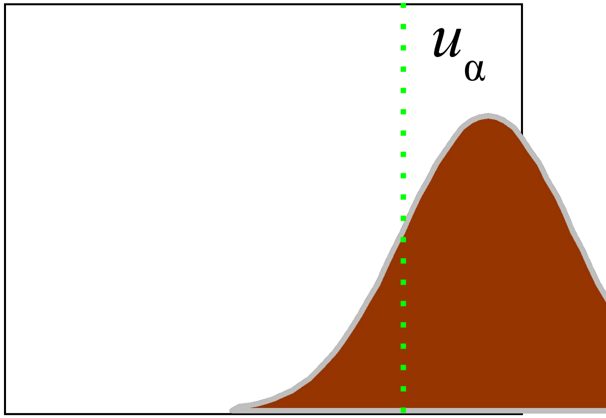
Acceptable $\alpha \Rightarrow$ threshold u_α

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

Procedure: if $t > u_\alpha$, then reject $\mathbf{H_0}$



Null Distribution of T

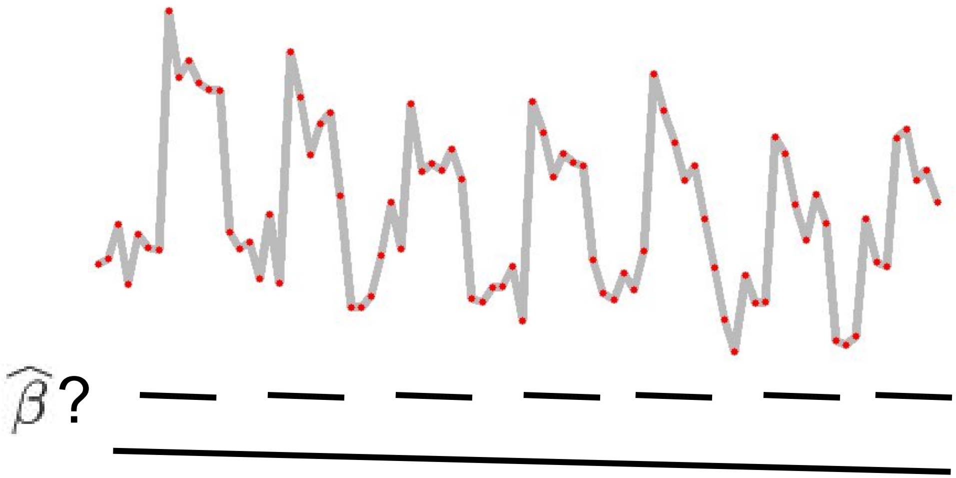


Alternative Distribution of T

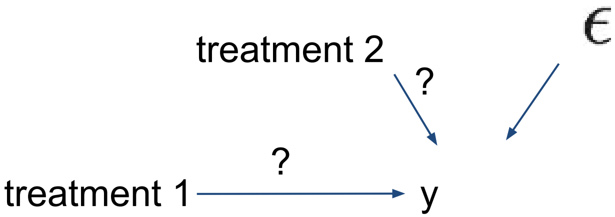
- ***Power of a test***

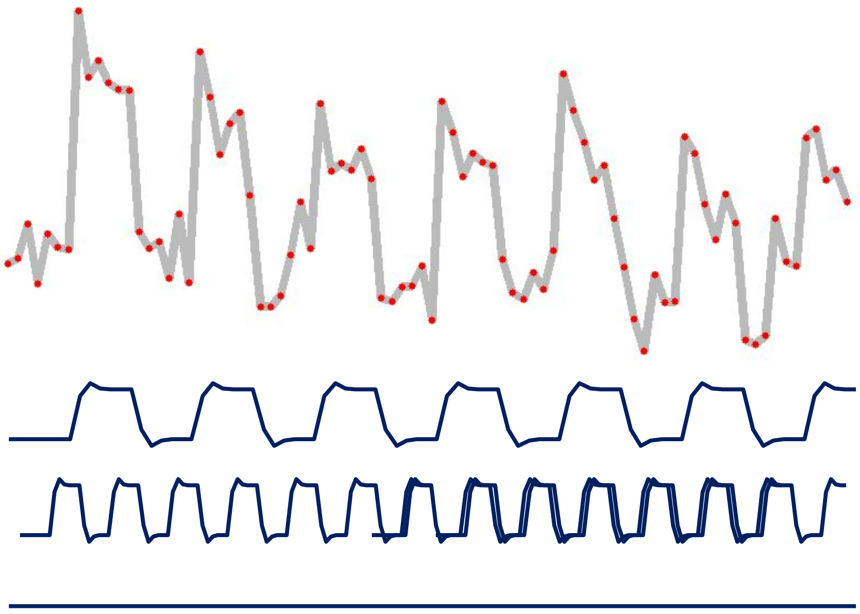
Depends on random variance

Univariate regression

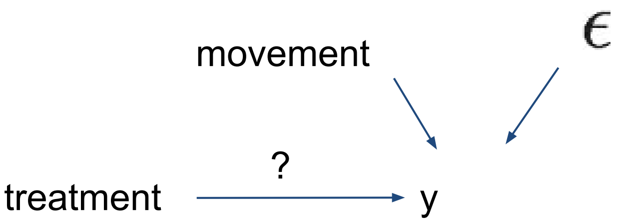


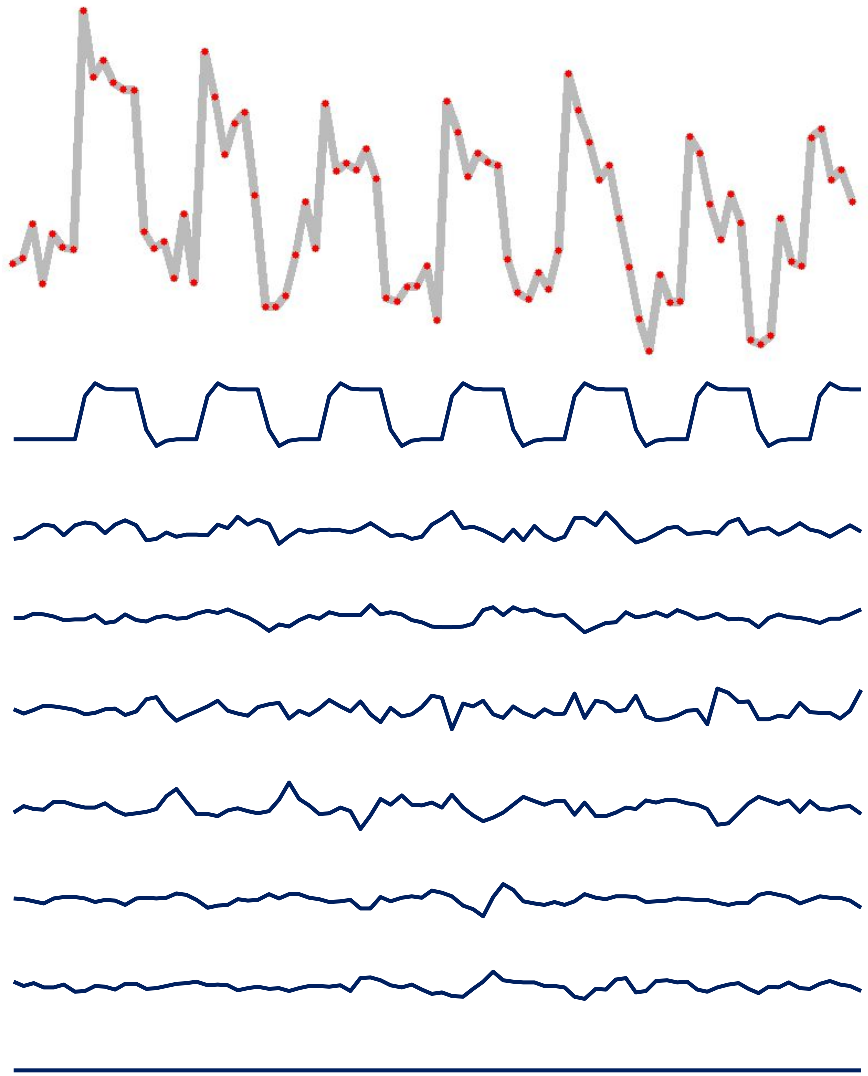
Why multiple predictors?
Problems they solve/create?





Independent treatments (orthogonal/decorrelated)



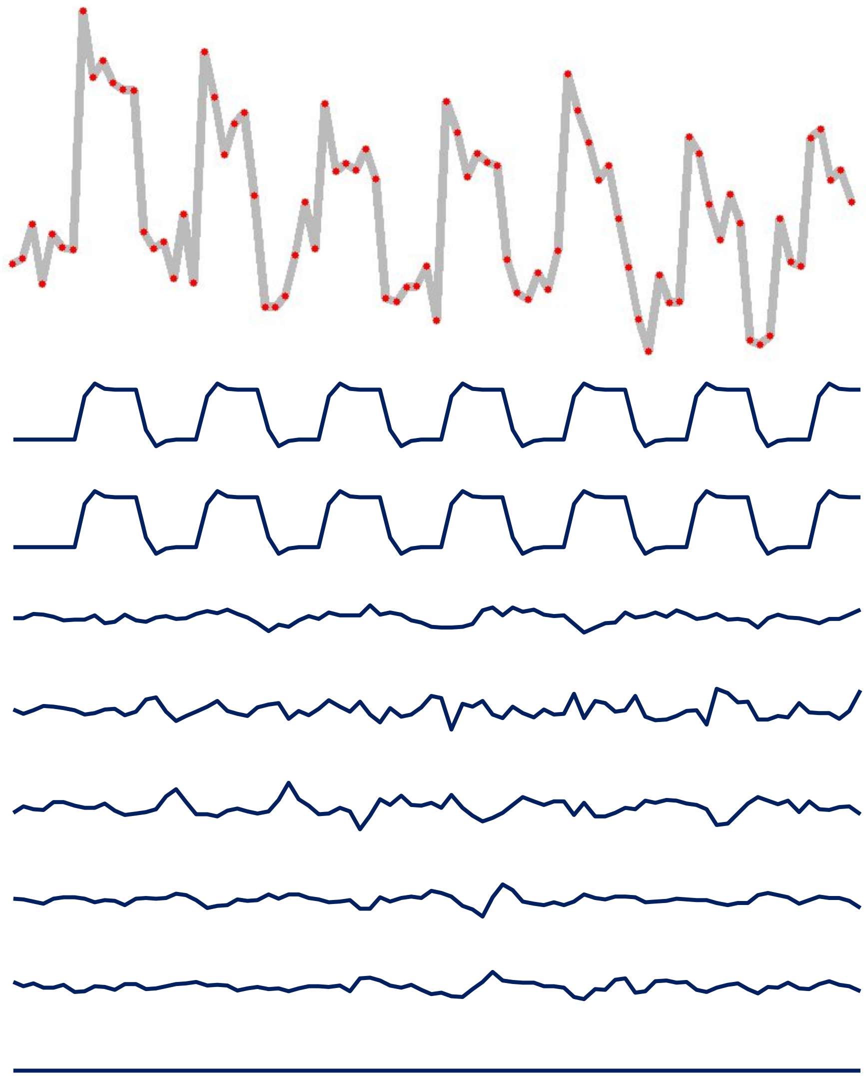


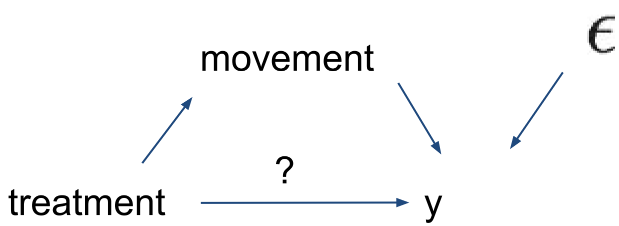
Reduce variance

$$\sigma^2(X'X)^{-1}\, via\, \sigma^2$$

so increase sensitivity, for given type-I.

Like prospective “blocking”.
We know movement is a cause .



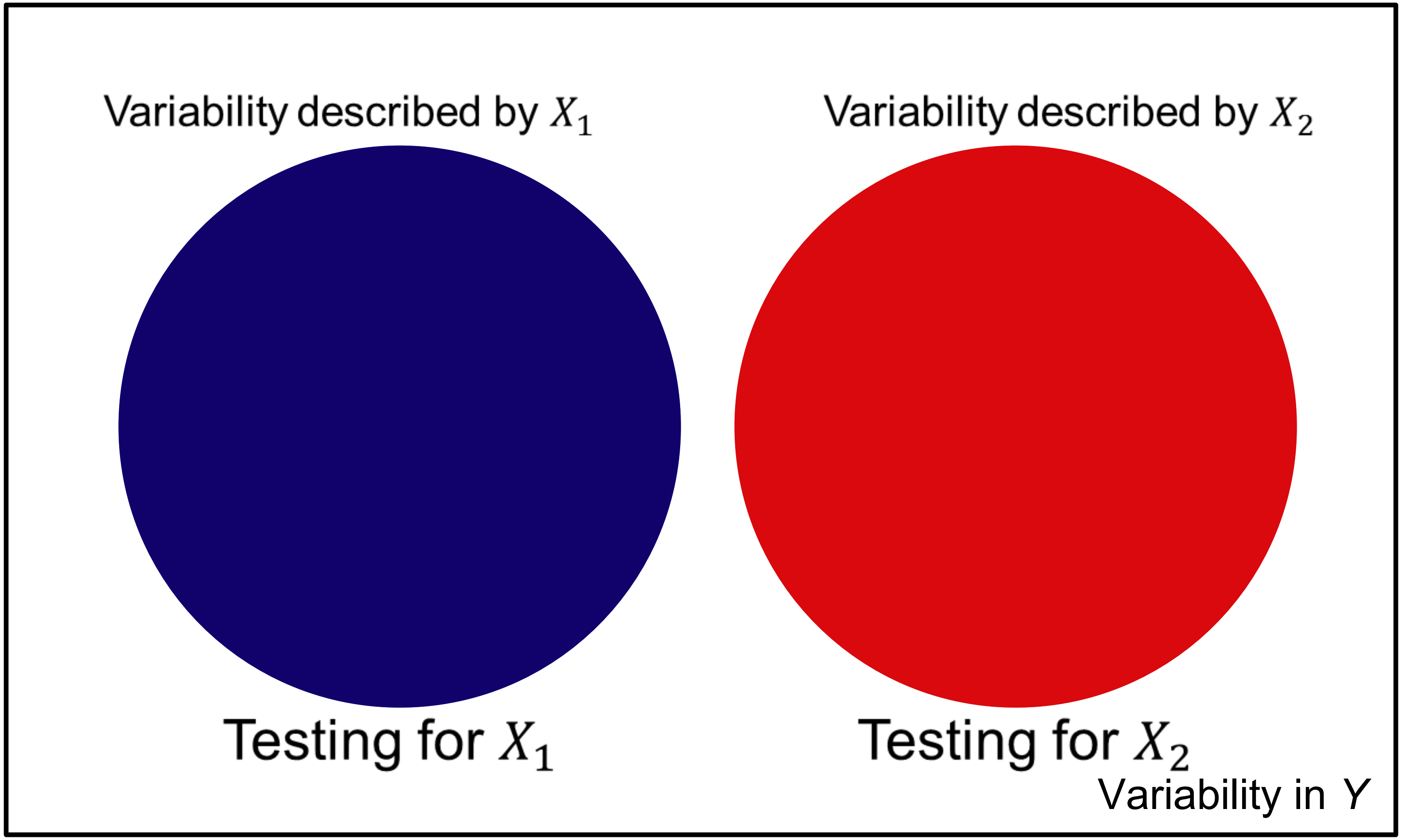


Tradeoff: reduce bias but reduce power

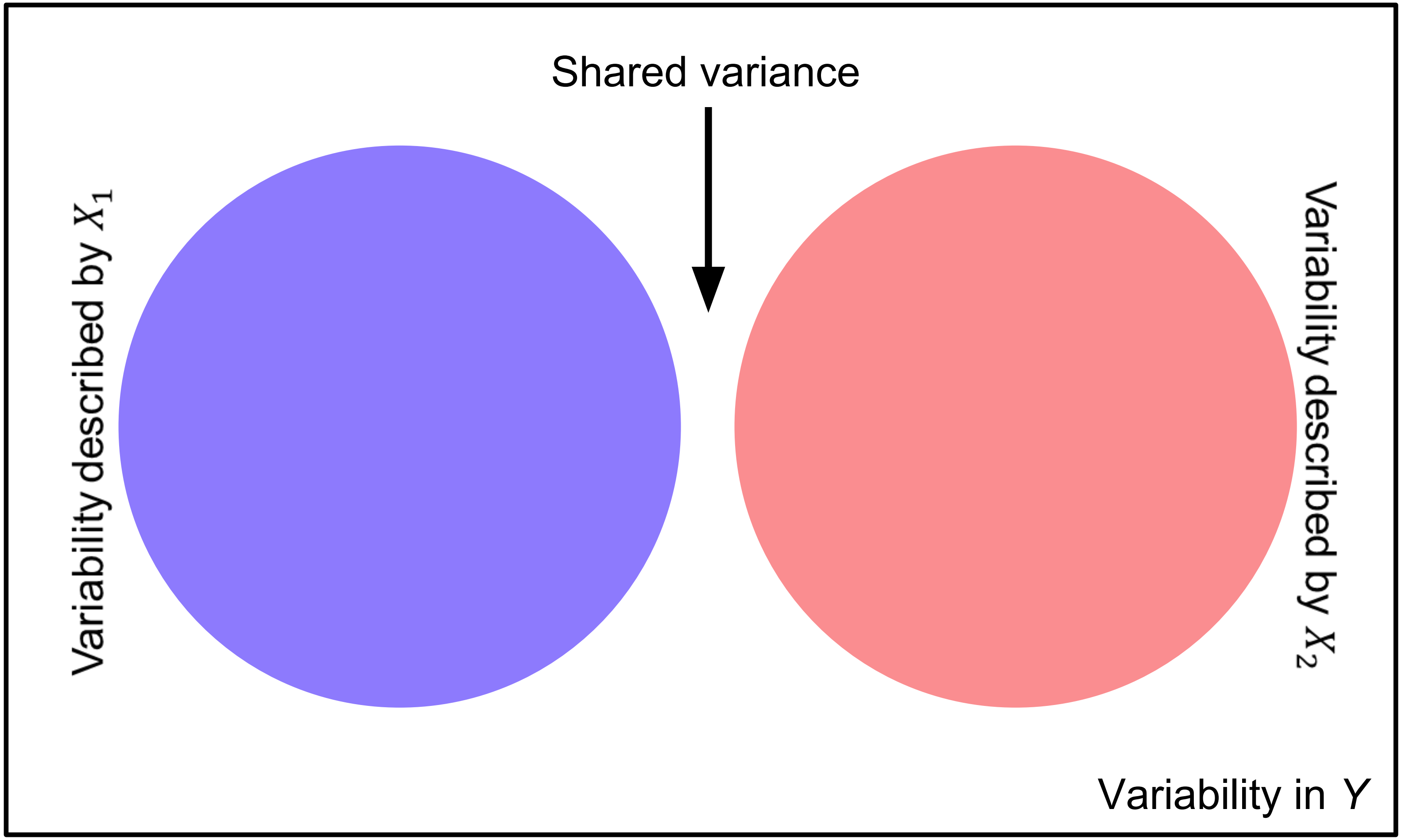
$$\sigma^2(X'X)^{-1} \; via \; (X'X)^{-1}$$

Derive variance inflation?
c.f. dependent predictors of interest

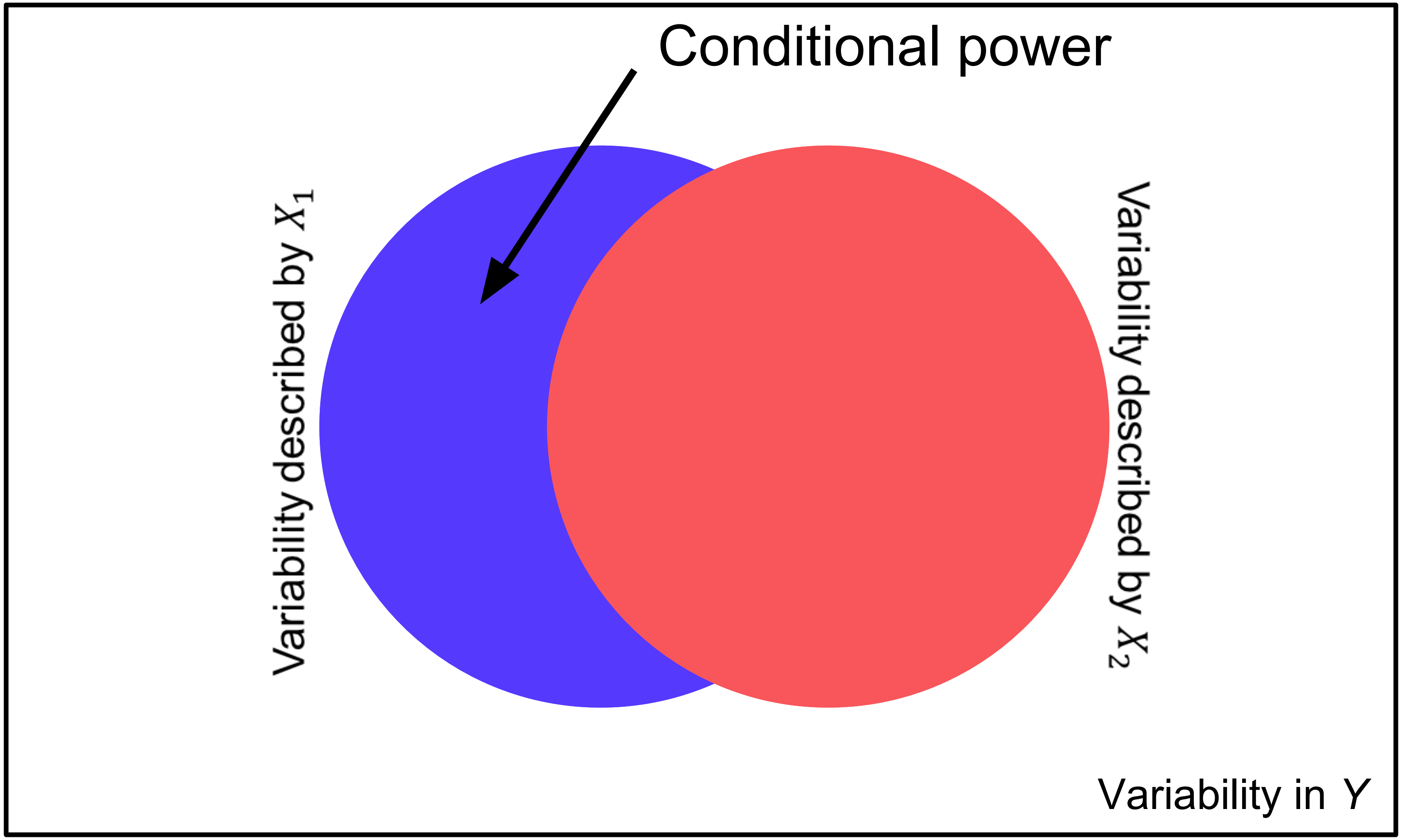
Orthogonal regressor cartoon



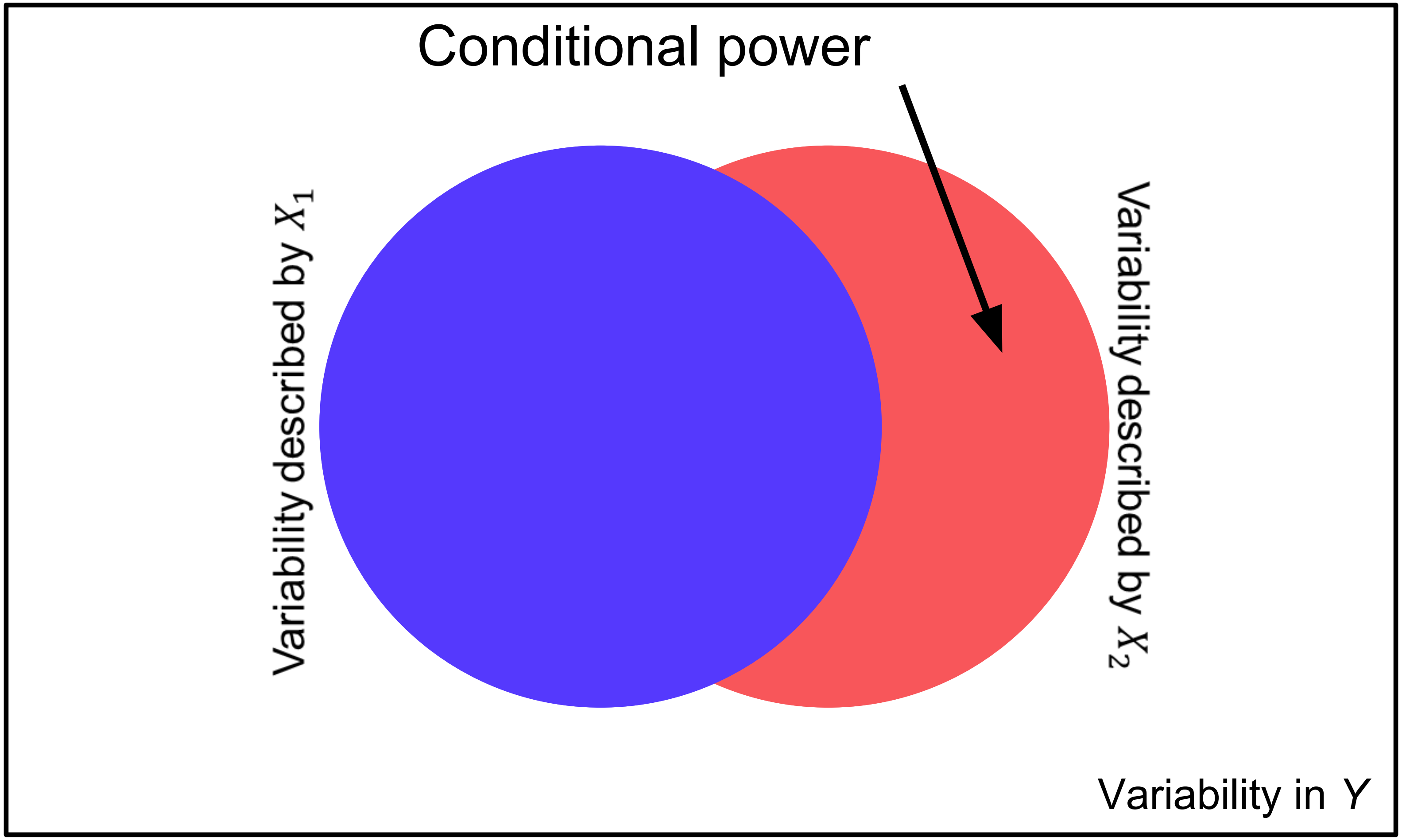
Correlated regressors



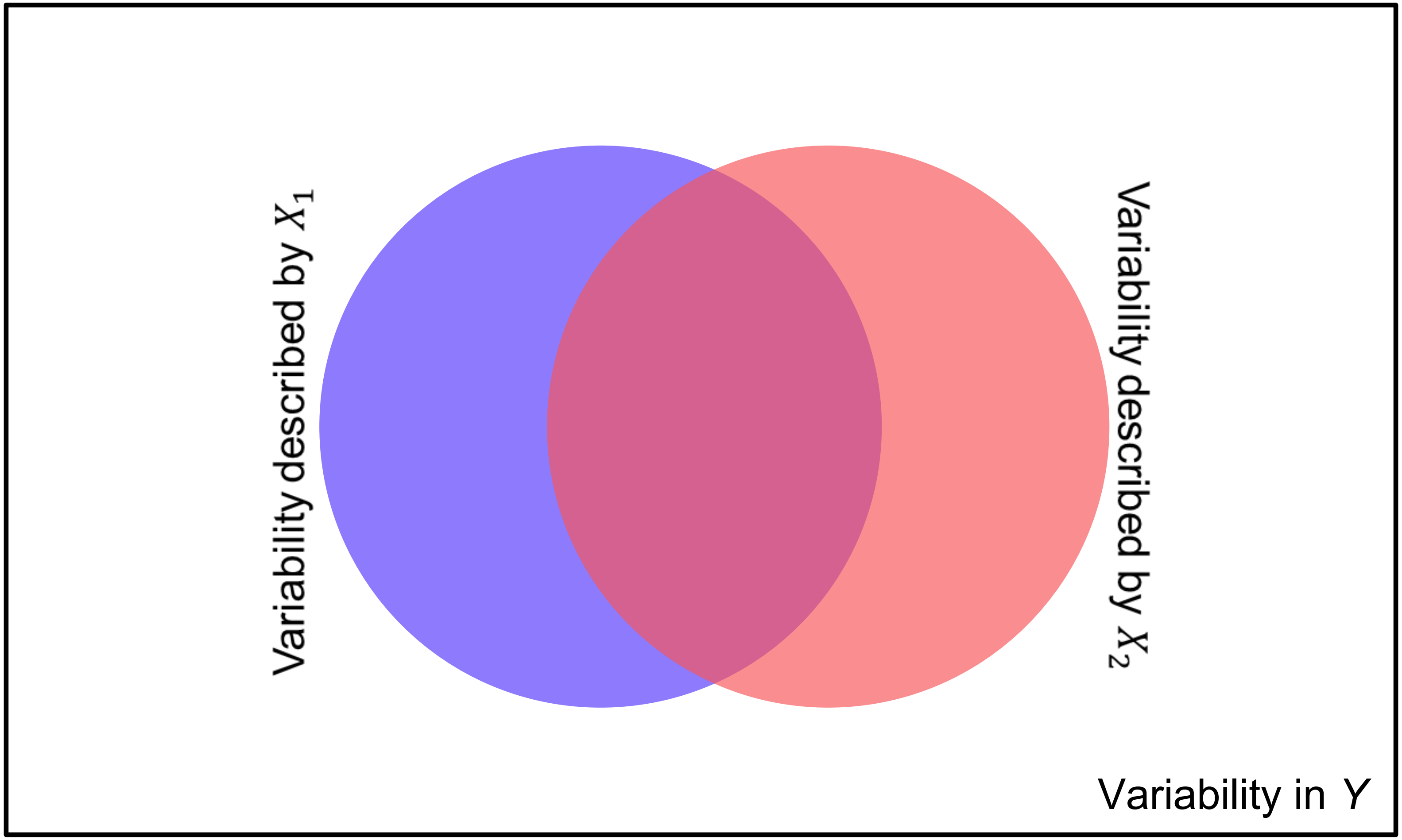
Correlated regressors



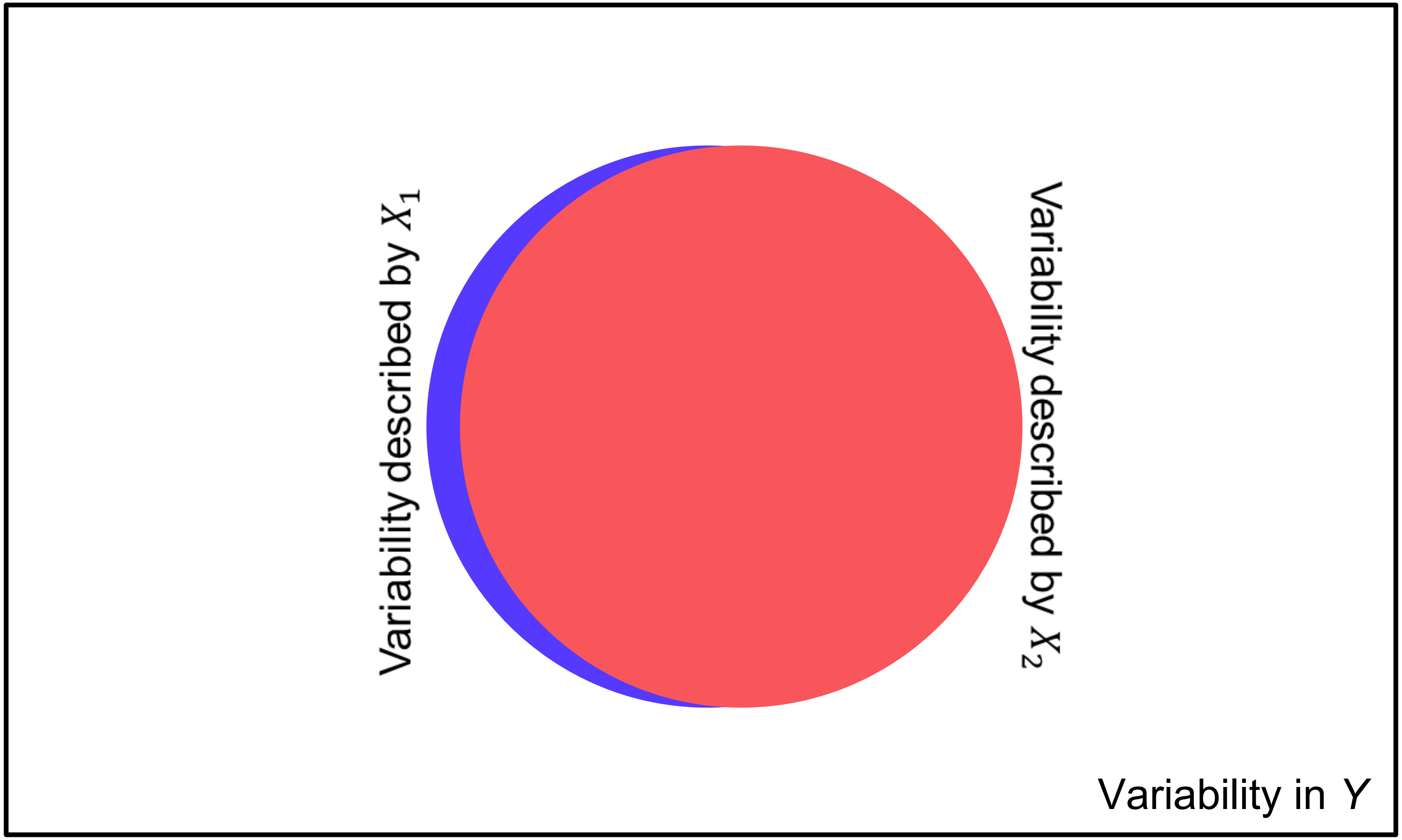
Correlated regressors



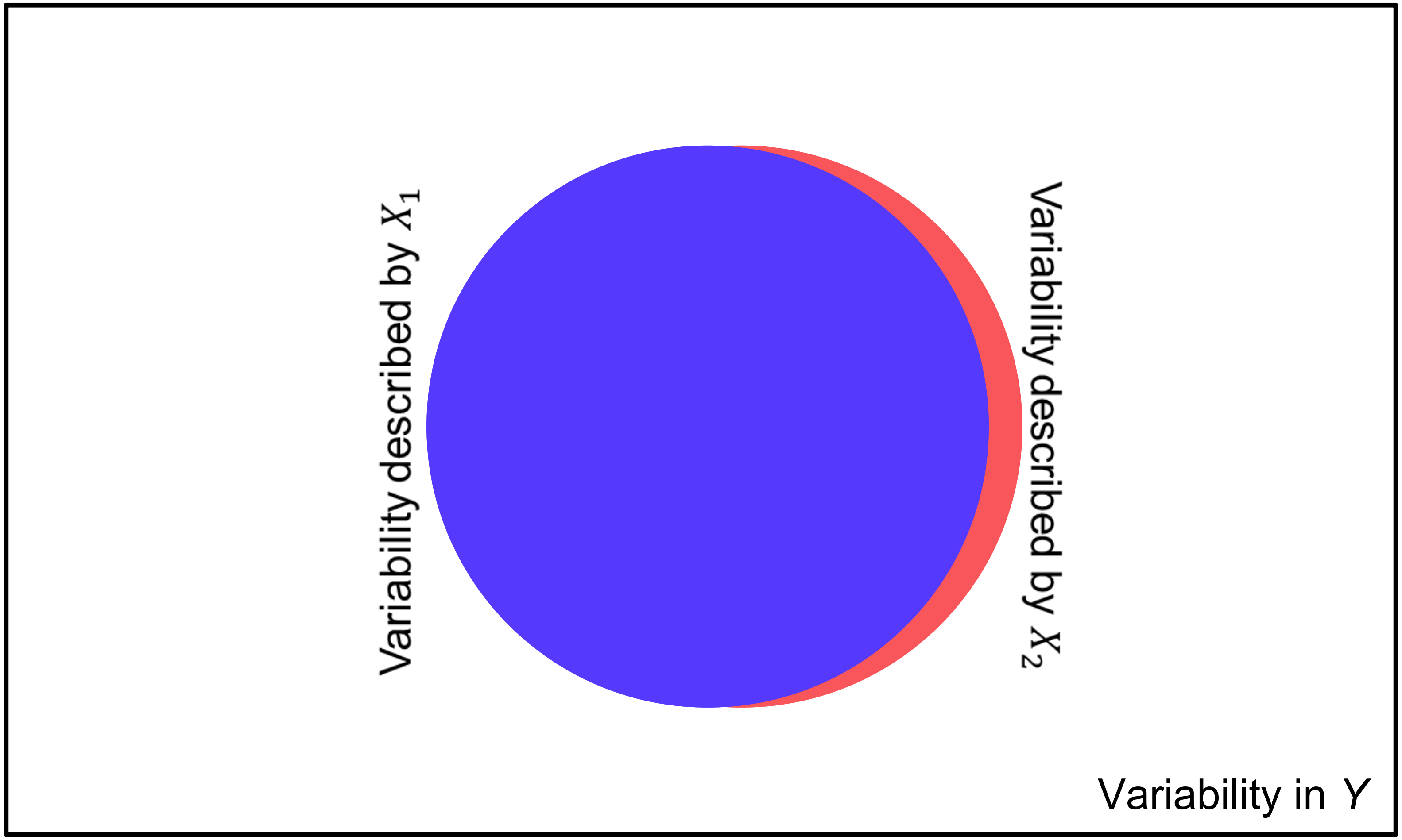
Correlated regressors



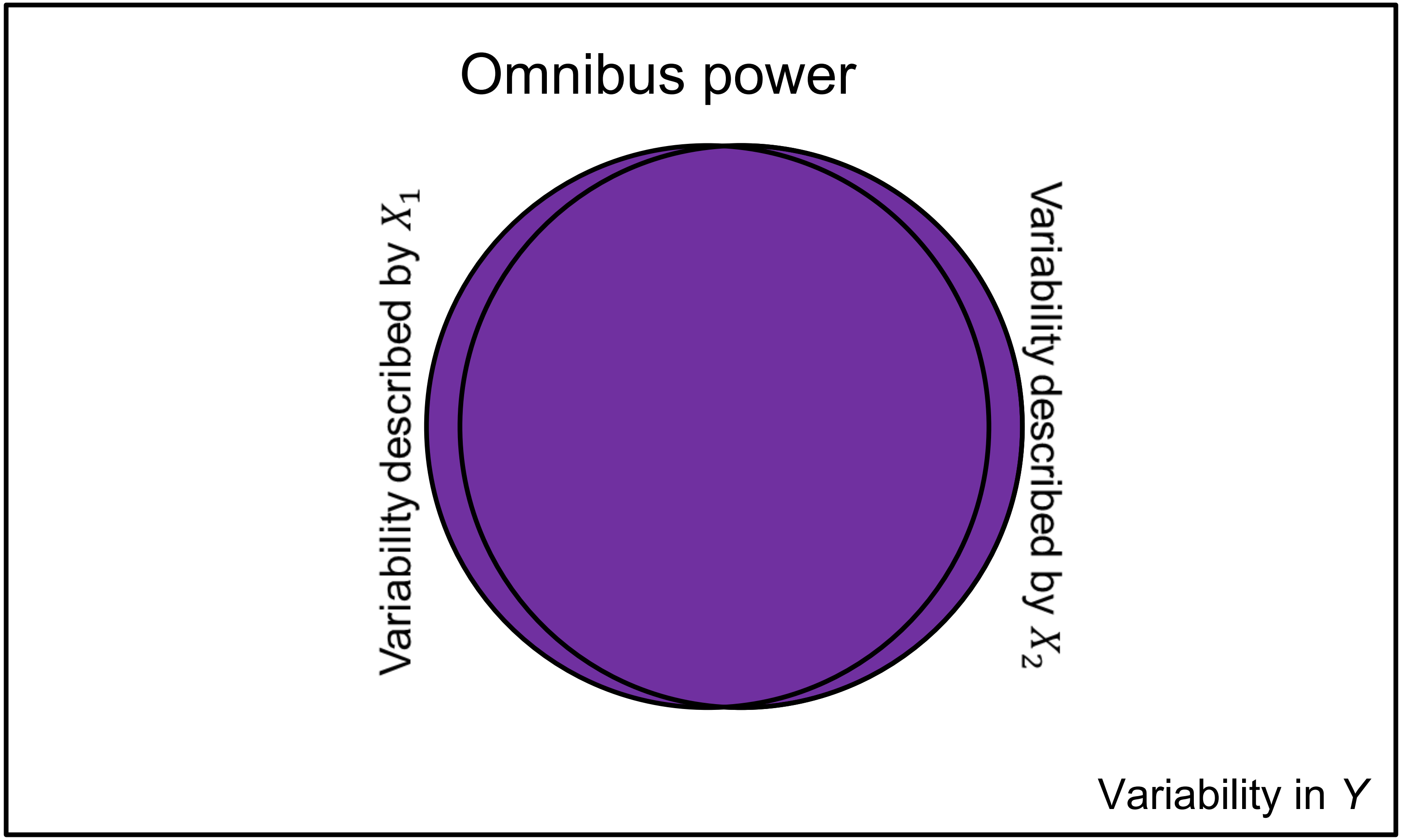
Correlated regressors



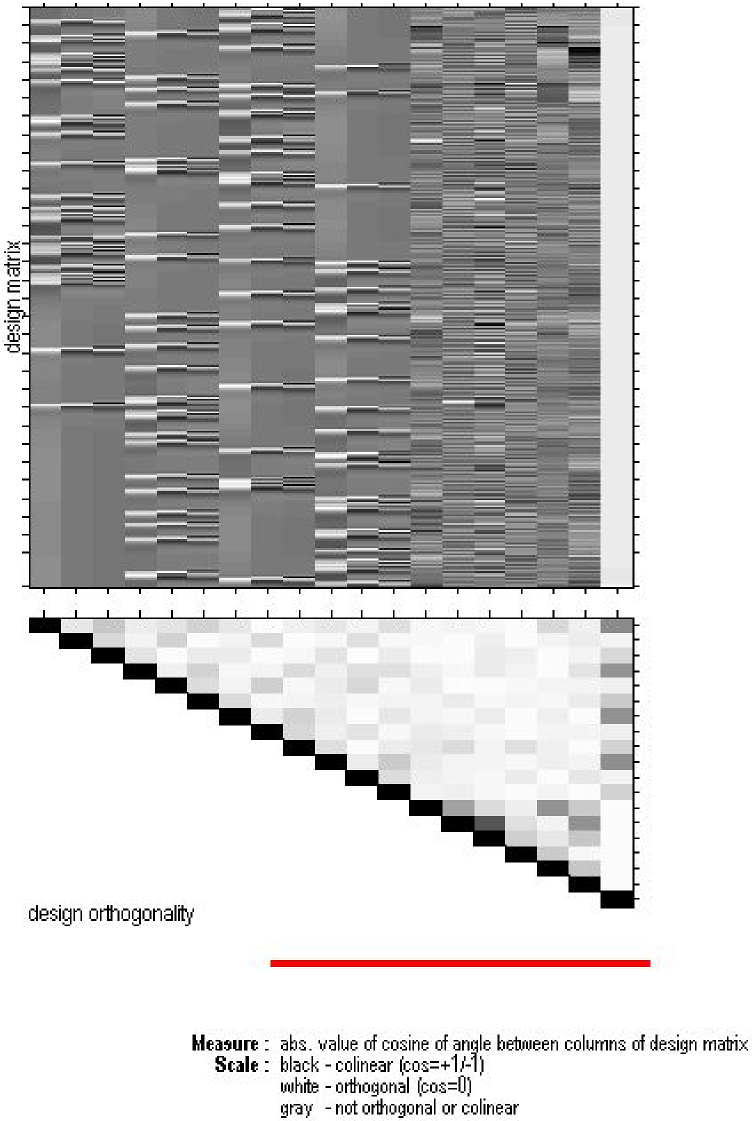
Correlated regressors



Correlated regressors



Design orthogonality

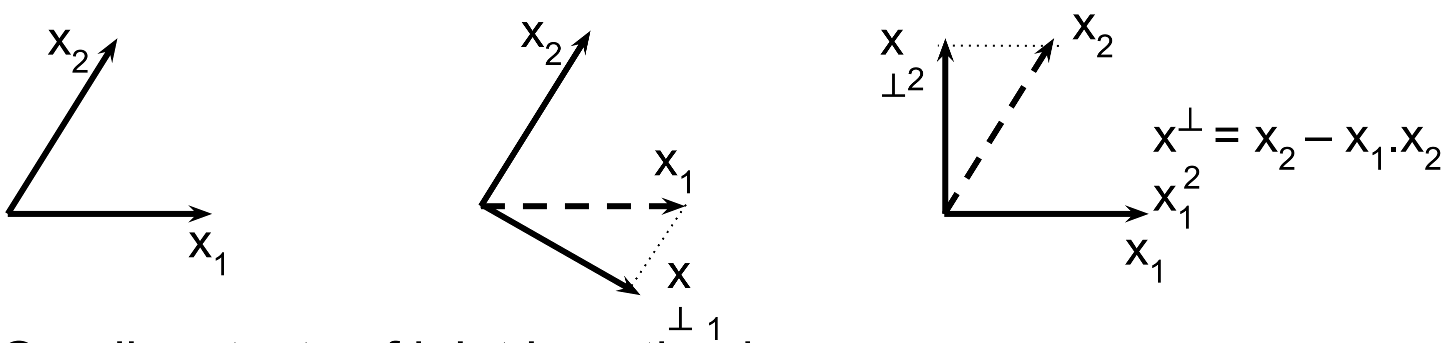


For each pair of columns of the design matrix, the orthogonality matrix depicts the magnitude of the **cosine of the angle** between them, with the range 0 to 1 mapped from white to black.

- 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/tests: summary

- Orthogonalize before (factorial designs) not after: strong assumptions about which regressor explains common variance (dangerous).
- Linear models implicitly "orthogonalize" individual regressors: When testing for the first regressor, we are effectively removing the part of the signal that can be accounted for by the second regressor:



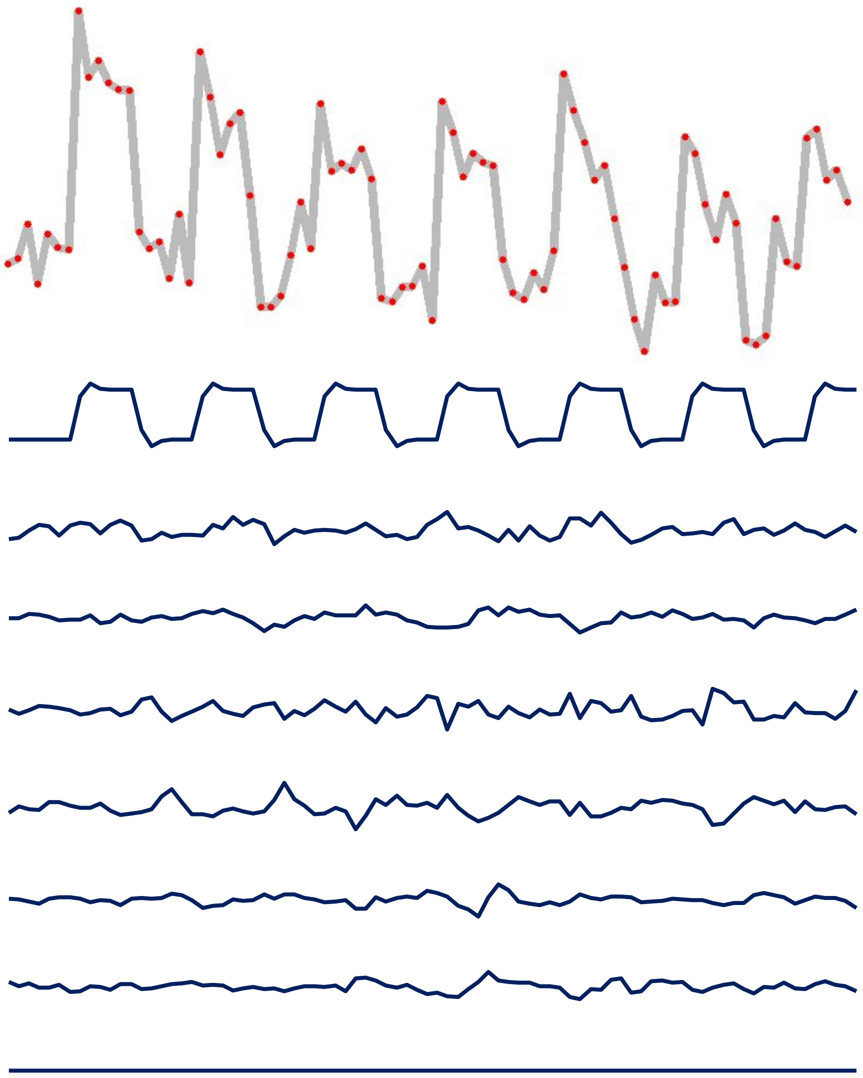
- Omnibus tests of joint hypothesis.
- Original regressors may not matter: interesting contrast should be orthogonal from the rest of the design matrix.

Why multiple predictors?
Problems they solve/create?

Multiplicity:
8 predictors
8 estimators
1 test statistic

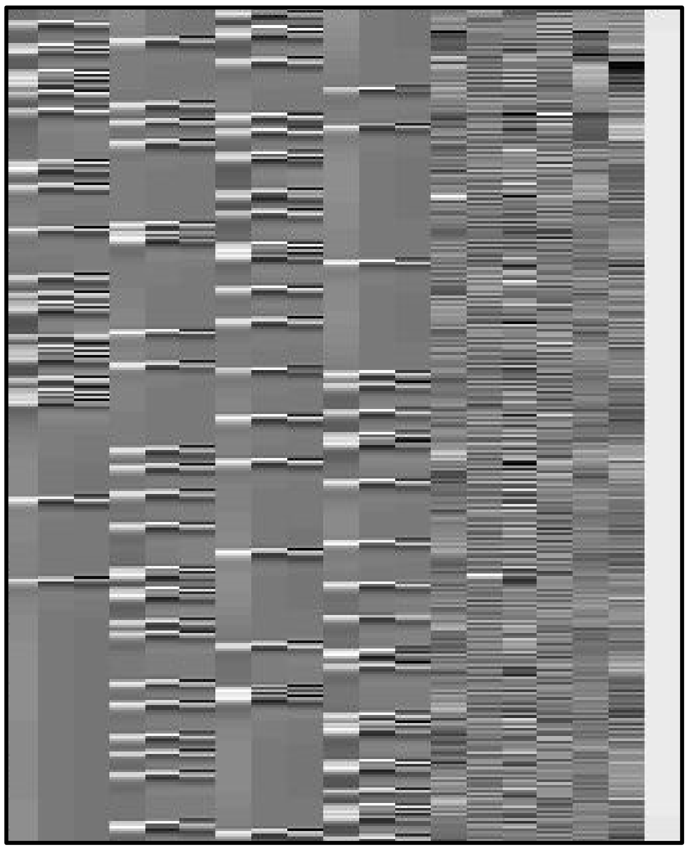
$$c^T\hat{\beta} \sim N(c^T\beta, \sigma^2 c^T (X^T X)^{-1} c)$$

estimators ≠ # test-stats (2*2 design, tests)
estimator dependence ≠ test-stat dependence



Univariate contrast estimation

[1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]



❑ 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.

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

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

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

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

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

Univariate hypothesis testing: summary

T-test *signal-to-noise* (estimate/ s.e. estimate).

- Alternative hypothesis:

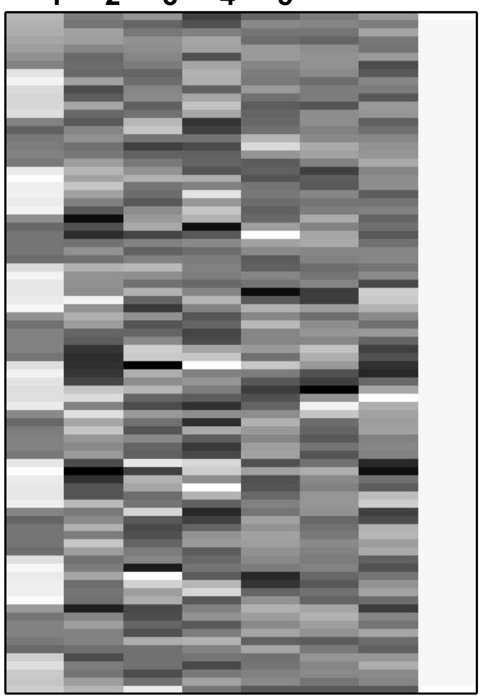
$$H_0:\quad c^T\beta = 0 \qquad \text{vs} \qquad H_A:\quad c^T\beta > 0$$

- *T*-contrasts linear combinations of $\widehat{\beta}$
- Functionally independent of regressor/contrast scale

Univariate hypothesis testing

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

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



Question:

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

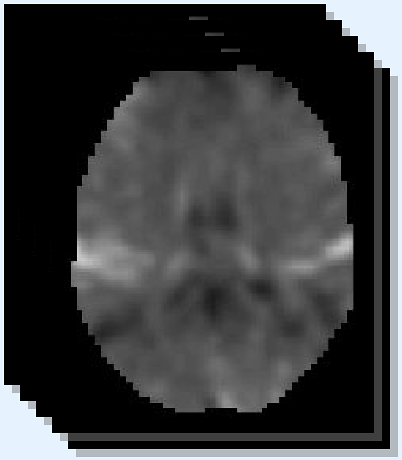
Null hypothesis: H_0: $c^T\beta=0$

Test statistic: $T = \dfrac{\text{Estimated effect}}{\text{Scale of error}}$ = **Effect * Precision**

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

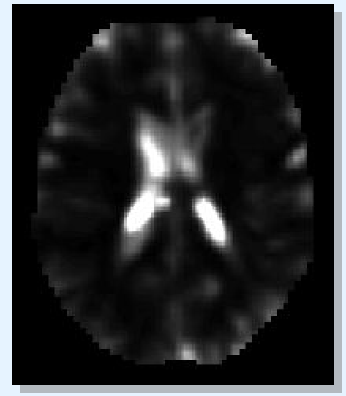
Univariate hypothesis testing

- For a given contrast *c*:



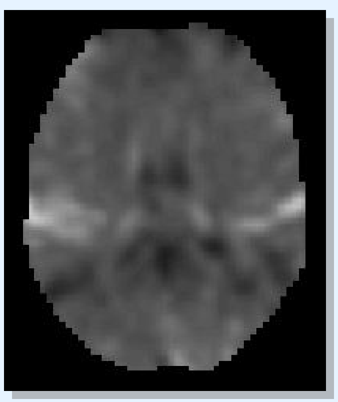
beta_???? images

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



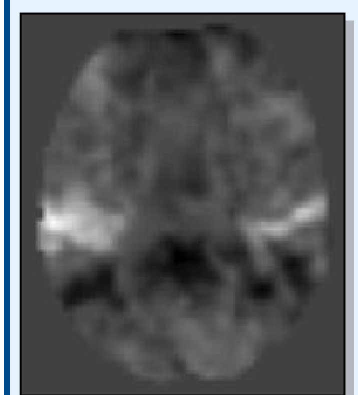
ResMS image

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



con_???? image

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



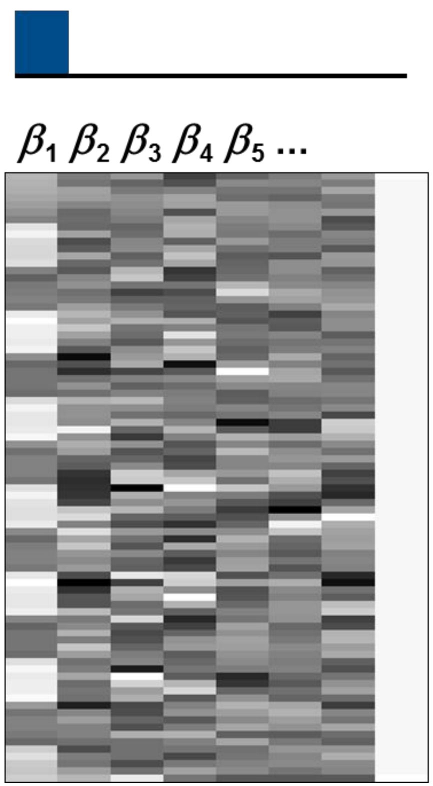
spmT_???? image

SPM{*t*}

Univariate hypothesis testing

- Passive word listening versus rest

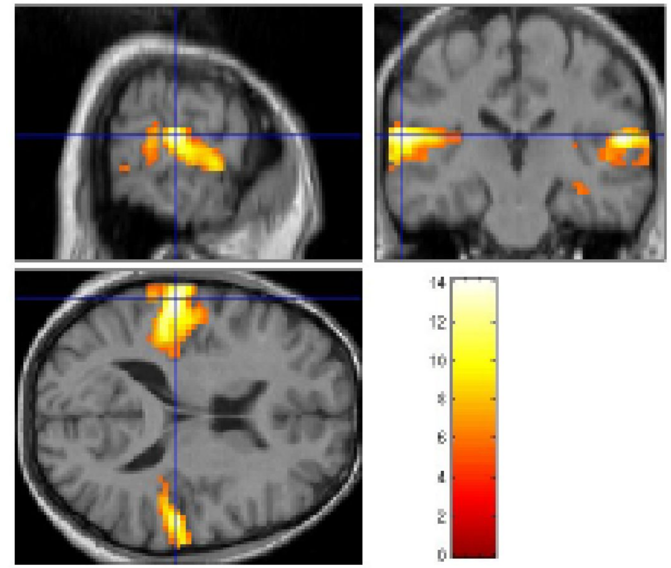
$c^T = 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0$



Q: activation during listening ?

Null hypothesis: $\beta_1 = 0$

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

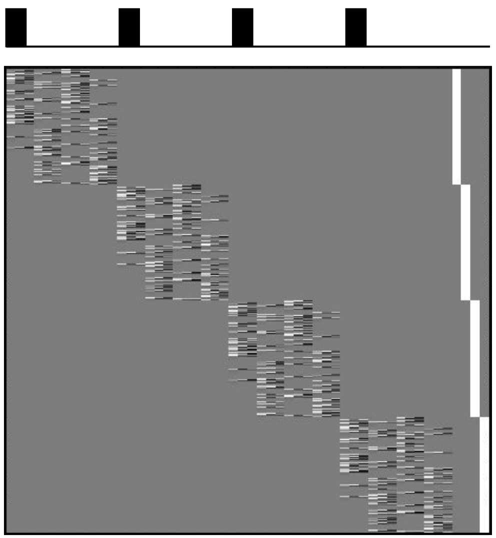


SPMresults: Threshold T = 3.2057 {p<0.001}
voxel-leve
l

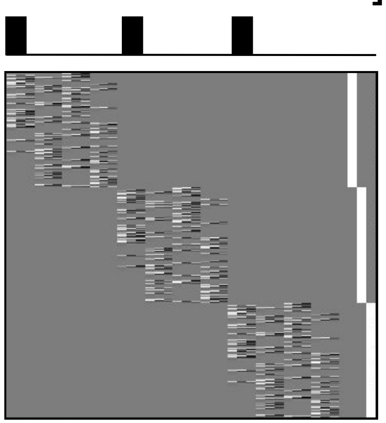
	(Z≡)	p uncorrected	Mm	mm	mm
13.94	**Inf**	**0.000**	**-63**	**-27**	**15**
12.04	Inf	0.000	-48	-33	12
11.82	Inf	0.000	-66	-21	6
13.72	**Inf**	**0.000**	**57**	**-21**	**12**
12.29	Inf	0.000	63	-12	-3
9.89	7.83	0.000	57	-39	6
7.39	**6.36**	**0.000**	**36**	**-30**	**-15**
6.84	**5.99**	**0.000**	**51**	**0**	**48**
6.36	**5.65**	**0.000**	**-63**	**-54**	**-3**

Notes on contrasts

[1 1 1 1] / 4



[1 1 1] / 3



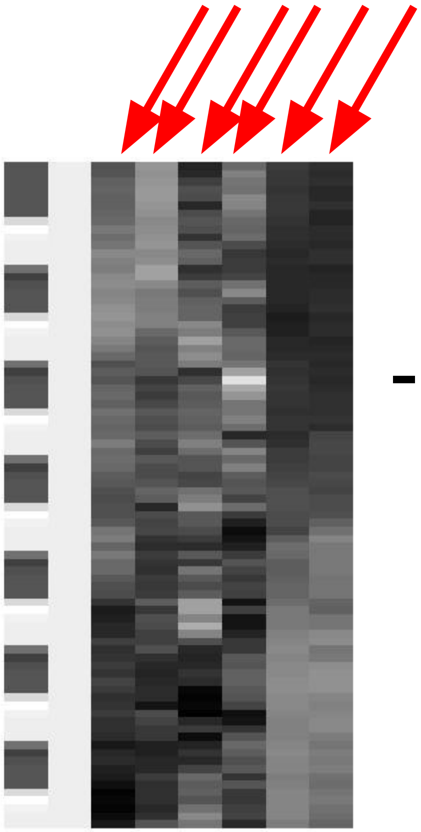
$$T = \frac{c^T \hat{\beta}}{\sqrt{\mathrm{var}(c^T \hat{\beta})}} = \frac{c^T \hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T \left(X^T X\right)^{-1} c}}$$

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.

  - Beware interpretation of the contrasts themselves (eg, $c^T \hat{\beta}$ for a second level analysis: sum ≠ average
  - Beware non-orthogonal contrasts. Are two linear combinations confounded?

Problems with multiplicity

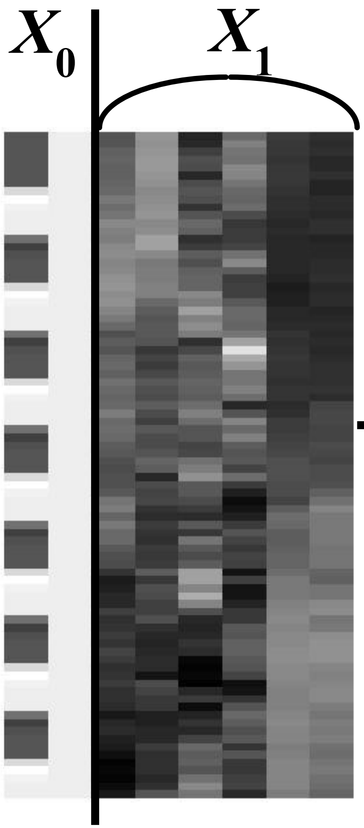


- Confounding bias
  - **Covariate adjustment** (costs power)
  - **Omnibus test** (costs interpretability)
- Similar for multiple overfitting (family-wise type-I error next lecture)

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

Model comparison:

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



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

$\rightarrow$ $\mathbf{RSS_0}$ $\sum \hat{\varepsilon}^2_{reduced}$

Test statistic: explained over unexplained variability (error)

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

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

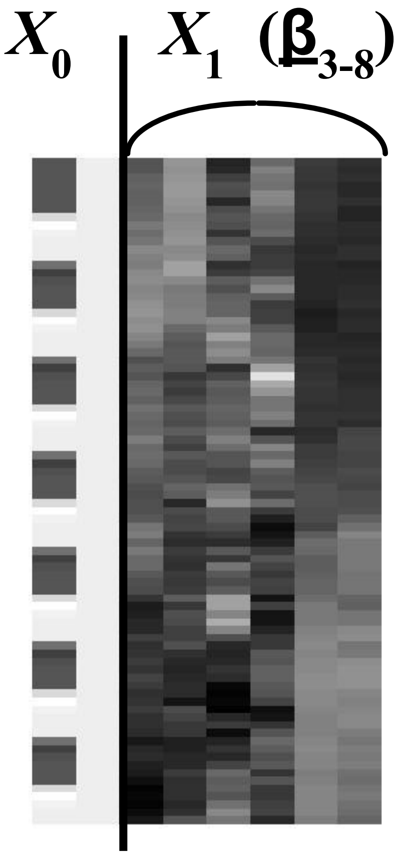
$v_1 = \mathrm{rank}(X) - \mathrm{rank}(X_0)$
$v_2 = N - \mathrm{rank}(X)$

Full model ? or Reduced model?

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

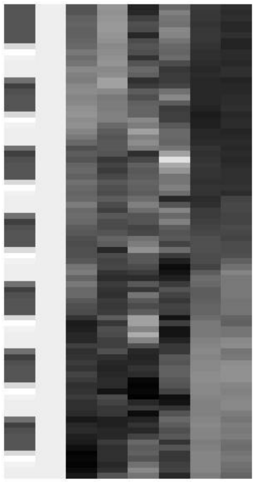
Joint linear hypothesis:

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

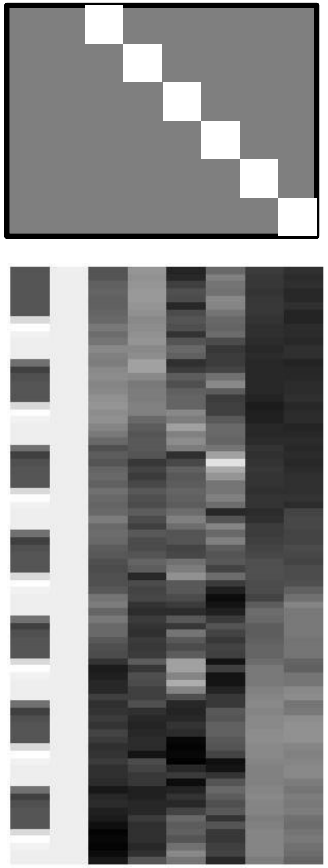


Full model ?

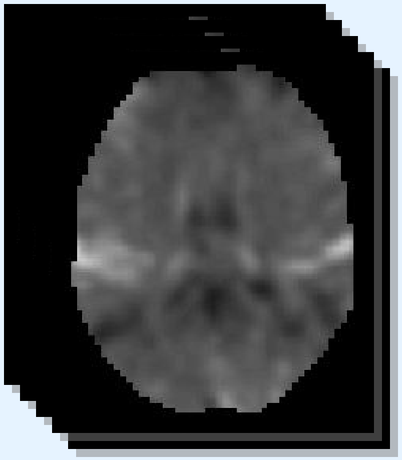
$$c^T = \begin{matrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{matrix}$$



Is any of $\beta_{3\text{-}8}$ non-zero?

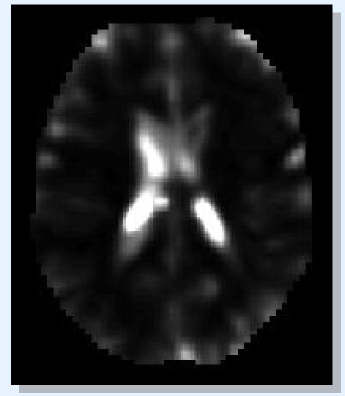


F-contrast in SPM



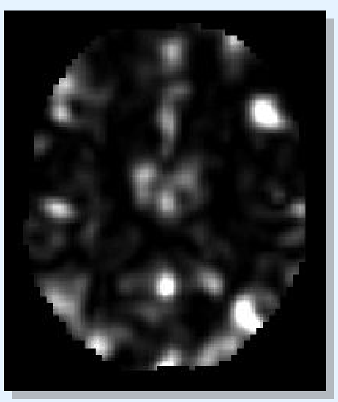
beta_???? images

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



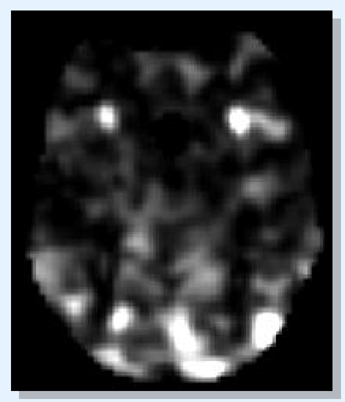
ResMS image

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



ess_???? images

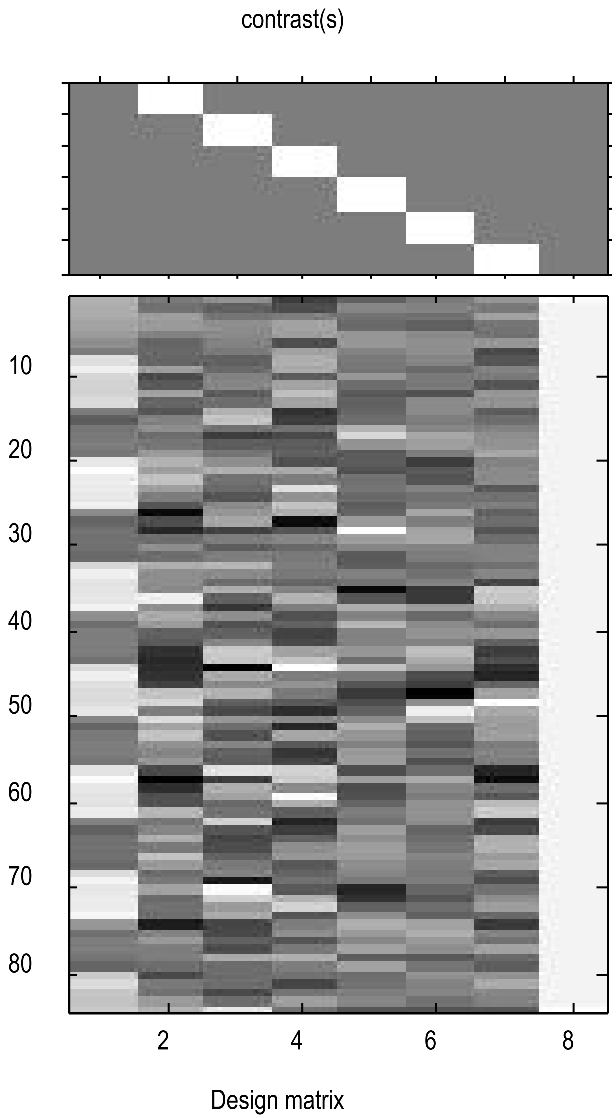
$$(RSS_0 - RSS)$$

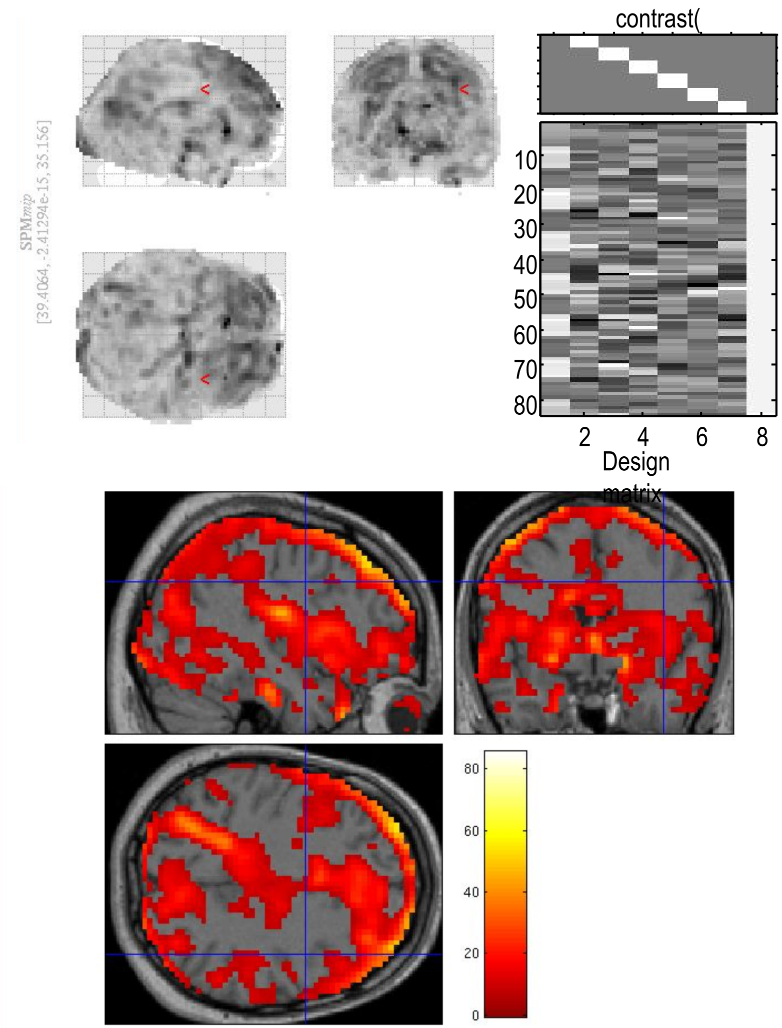


spmF_???? images

SPM{F}

F-test example: movement related effects





F-test: summary

F-test a nested submodel ➪ ***model comparison***.

- F test: weighted **sum-of-squares** of one or several combinations of the regression coefficients β.
- Needn't explicitly separate X = [X_0 X_1] thanks to **multidimensional contrasts**.
- Hypotheses:

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

- Univariate contrast: the square of the *t*-test, testing positive or negative effects.

Summary

Types of dependence...

- Noise (last week)
- Regressors (confound bias & variance)
- Contrasts (next lecture)
- Tests

Omnibus tests

- Combine dependent tests but weaken interpretation
- Threshold adjustment (next lecture)