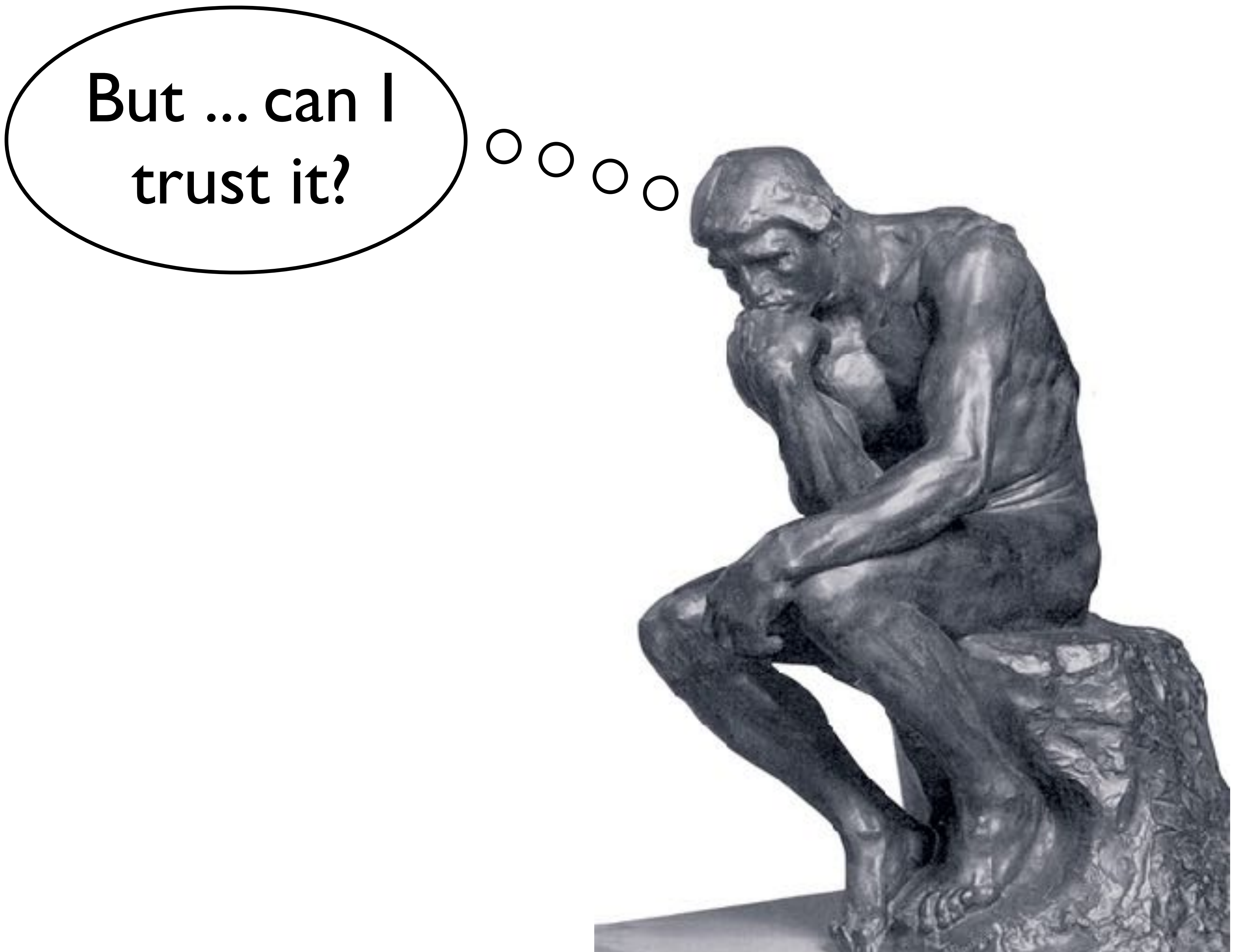
Inference

how surprising is your statistic? (thresholding)

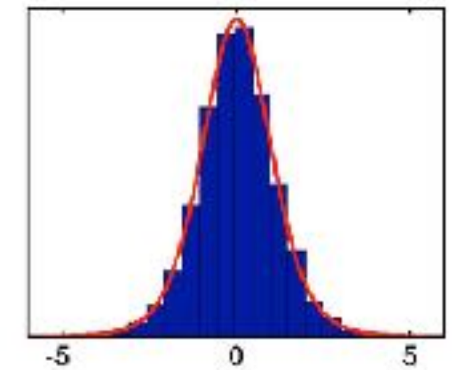


Outline

- Null-hypothesis and Null-distribution
- Multiple comparisons and Family-wise error
- Different ways of being surprised
    - Voxel-wise inference (Maximum z)
    - Cluster-wise inference (Maximum size)
- Parametric vs non-parametric tests
- Enhanced clusters
- FDR - False Discovery Rate

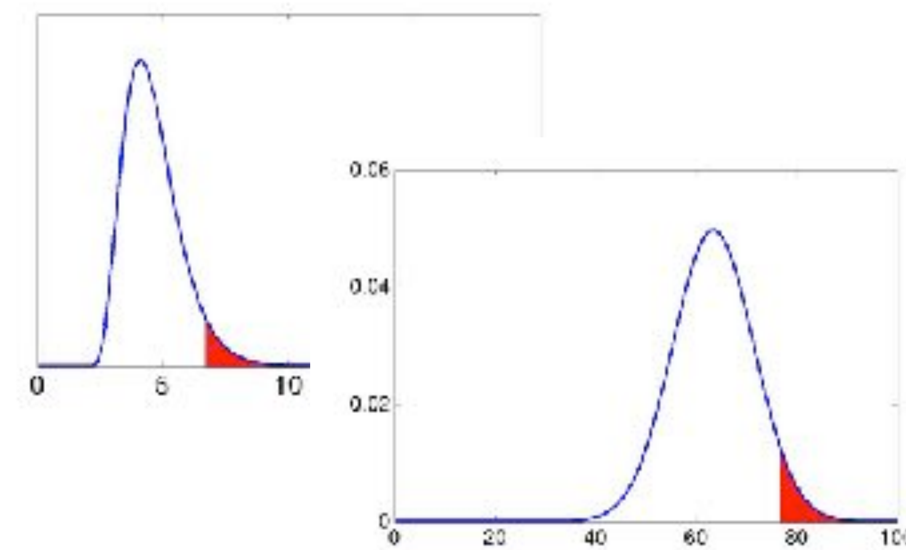
Parametric vs non-parametric

- As we described earlier, one of the great things about for example the t-test is that we know the null-distribution

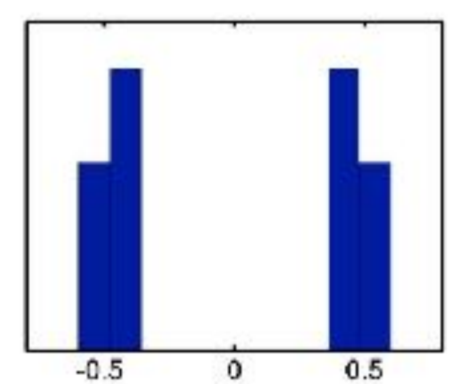


Provided that **e** ~ $N(0,\sigma^2)$

- But most distributions are not that simple



- And errors are not always normal-distributed

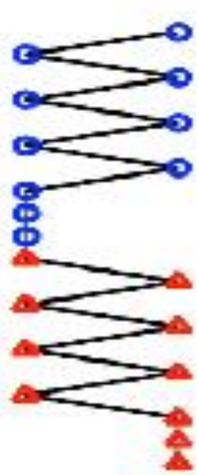


Example: VBM-style analysis

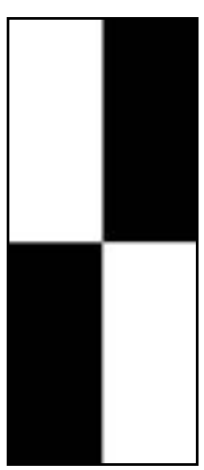
- Our data is segmented grey matter maps
- A voxel is either grey matter, or not.

Group #1
(Oxford students)

Group #2
(Train spotters)



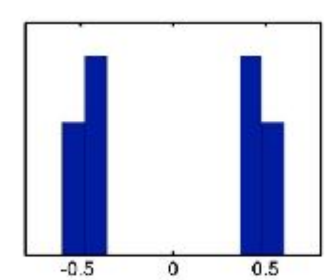
=



$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix}$$ Ok!

hist(**e**)

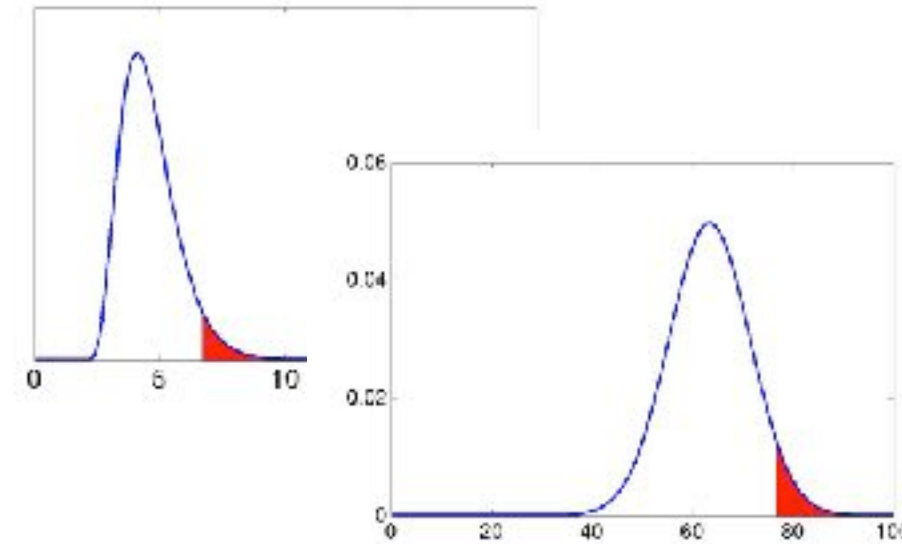


~ *N* ?

☹

Parametric vs non-parametric

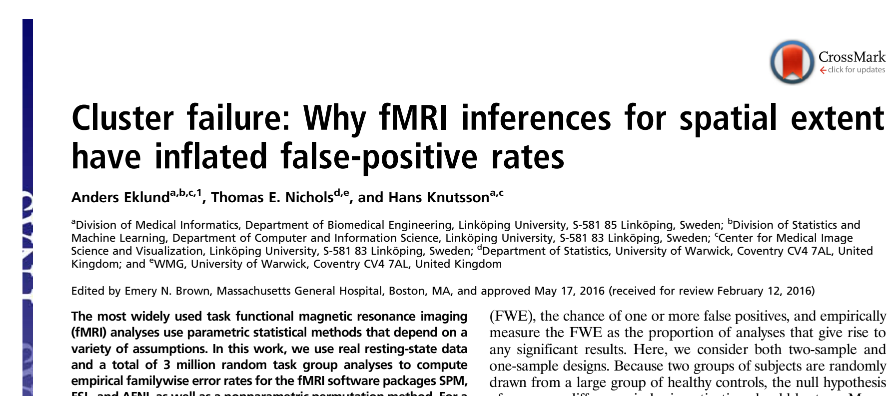
- There are <u>approximations</u> to the Max-z and Max-size statistics



- These are valid under certain sets of assumptions

  - Search area "large relative to boundary"
  - "High enough" cluster forming threshold
  - Normal distributed errors

- But can be a problem when applied outside of that set of assumptions



CrossMark
click for updates

Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates

Anders Eklund[a,b,c,1], Thomas E. Nichols[d,e], and Hans Knutsson[a,c]

[a]Division of Medical Informatics, Department of Biomedical Engineering, Linköping University, S-581 85 Linköping, Sweden; [b]Division of Statistics and Machine Learning, Department of Computer and Information Science, Linköping University, S-581 83 Linköping, Sweden; [c]Center for Medical Image Science and Visualization, Linköping University, S-581 83 Linköping, Sweden; [d]Department of Statistics, University of Warwick, Coventry CV4 7AL, United Kingdom; and [e]WMG, University of Warwick, Coventry CV4 7AL, United Kingdom

Edited by Emery N. Brown, Massachusetts General Hospital, Boston, MA, and approved May 17, 2016 (received for review February 12, 2016)

The most widely used task functional magnetic resonance imaging (fMRI) analyses use parametric statistical methods that depend on a variety of assumptions. In this work, we use real resting-state data and a total of 3 million random task group analyses to compute empirical familywise error rates for the fMRI software packages SPM,

(FWE), the chance of one or more false positives, and empirically measure the FWE as the proportion of analyses that give rise to any significant results. Here, we consider both two-sample and one-sample designs. Because two groups of subjects are randomly drawn from a large group of healthy controls, the null hypothesis

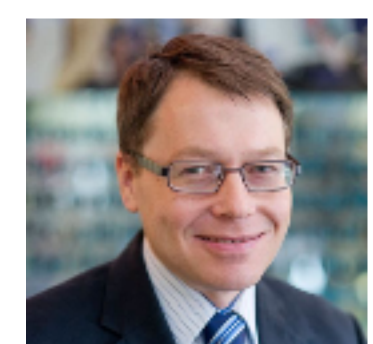
Parametric vs non-parametric

- Those approximations were based on Gaussian Random Field Theory, and was an impressive body of work

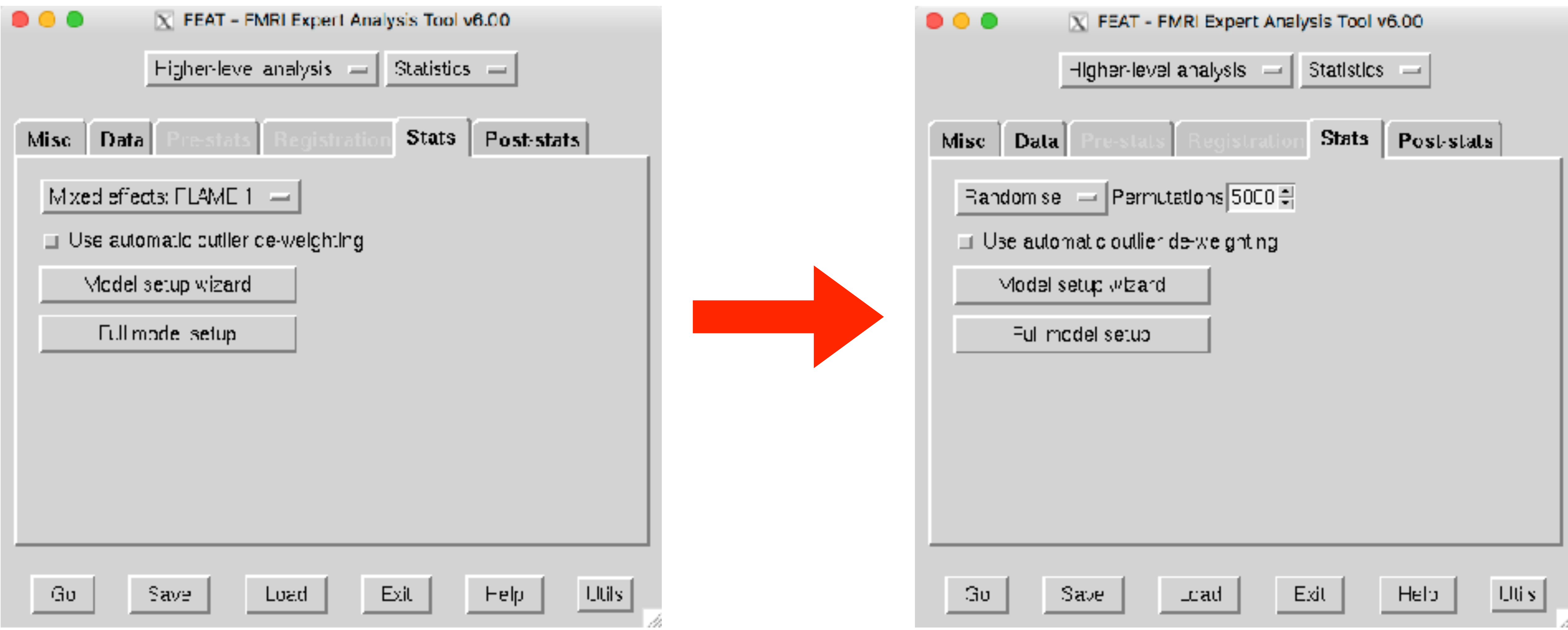

The Geometry of Random Images
Keith J. Worsley

- They served us fantastically well at a time when we had little choice

- But the future is non-parametric

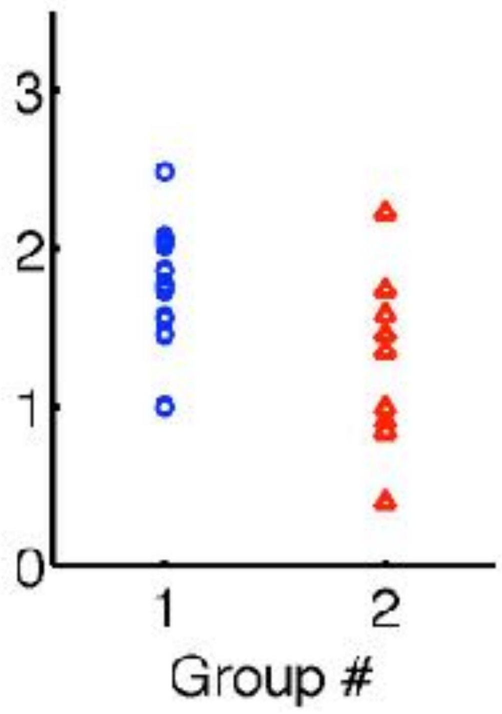
Parametric vs non-parametric



A simple permutation test

- We can permute the data itself to create a distribution that we can use to test our statistic.
  - + Makes very few assumptions about the data
  - + Works for any test statistic

We have performed an experiment



And calculated a statistic, e.g. a *t*-value

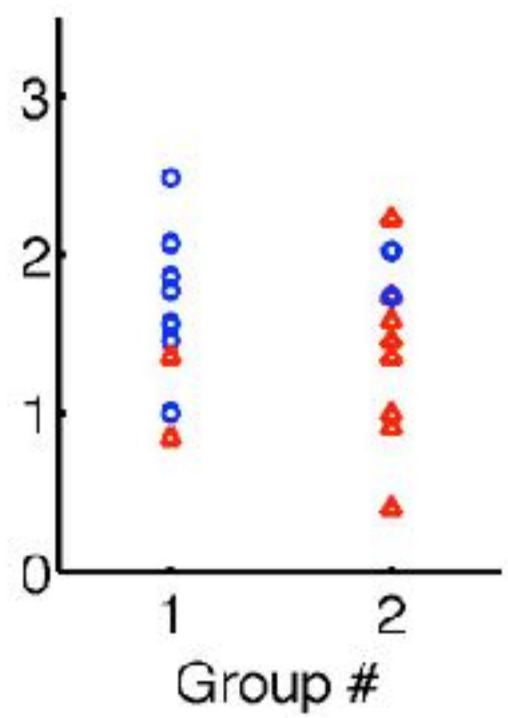
$$t = 2.27$$

If the null-hypothesis is true, there is no difference between the groups. That means we should be able to "re-label" the individual points without changing anything.

A simple permutation test

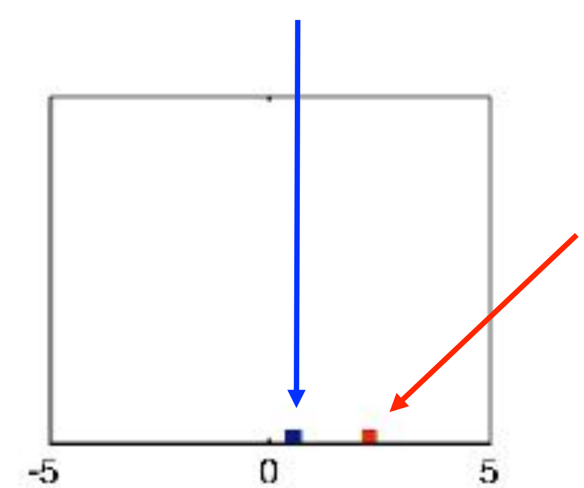
- We can permute the data itself to create a distribution that we can use to test our statistic.
  - \+ Makes very few assumptions about the data
  - \+ Works for any test statistic

One re-labelling



t-value after re-labelling

$$t = 0.67$$



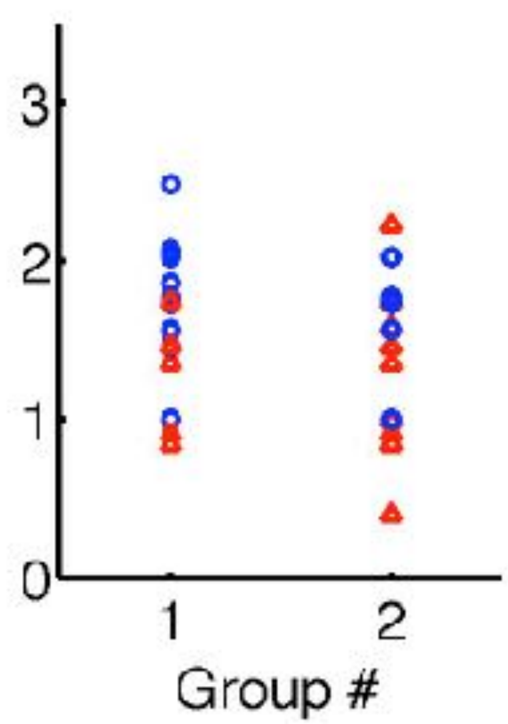
Original labelling

Let's start collecting them

A simple permutation test

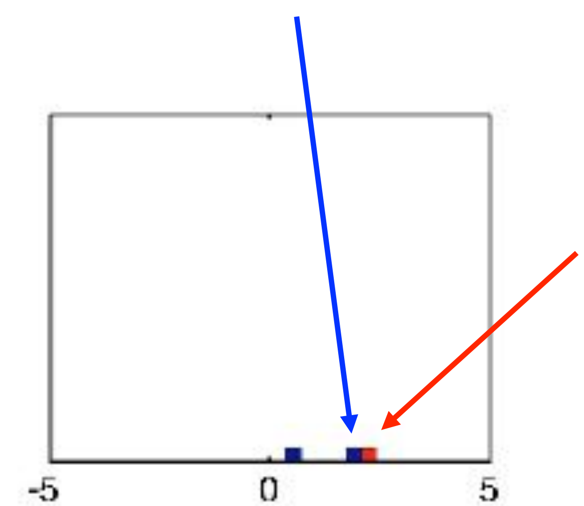
- We can permute the data itself to create a distribution that we can use to test our statistic.
  - + Makes very few assumptions about the data
  - + Works for any test statistic

Second re-labelling



t-value after re-labelling

$$t = 1.97$$



Original labelling

And another one

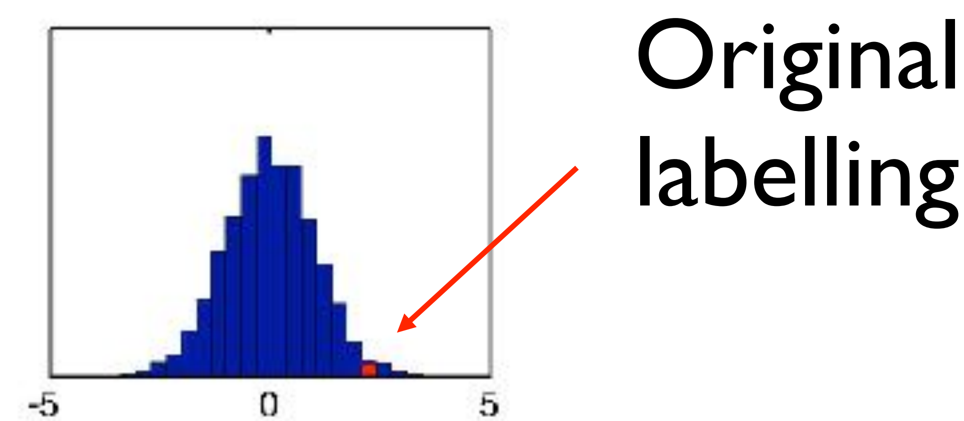
A simple permutation test

- We can permute the data itself to create a distribution that we can use to test our statistic.
  - \+ Makes very few assumptions about the data
  - \+ Works for any test statistic

Of the 5000 re-labellings, only 90 had a t-value > 2.27 (the original labelling).

I.e. there is only a ~1.8% (90/5000) chance of obtaining a value > 2.27 if there is no difference between the groups

C.f. $p(x \geq 2.27) = 1.79\%$ for t_{18}

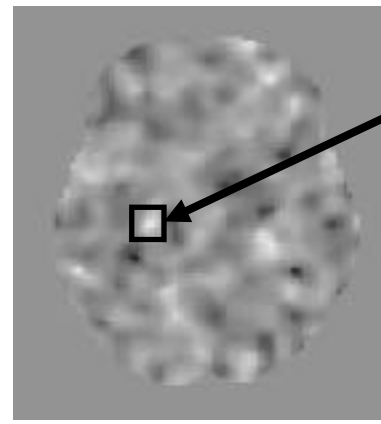


5000 re-labellings. Phew!

And we can use this for any statistic

We compared activation by painful stimuli in two groups of 5 subjects each.

This is what we got



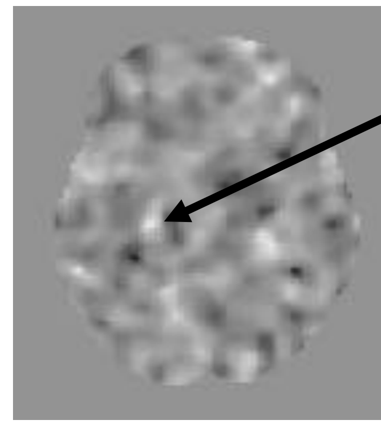
Very intriguing activation. $t_8 = 4.65$

But, can we trust it?

And we can use this for any statistic

This is what we got

We compared activation by painful stimuli in two groups of 5 subjects each.

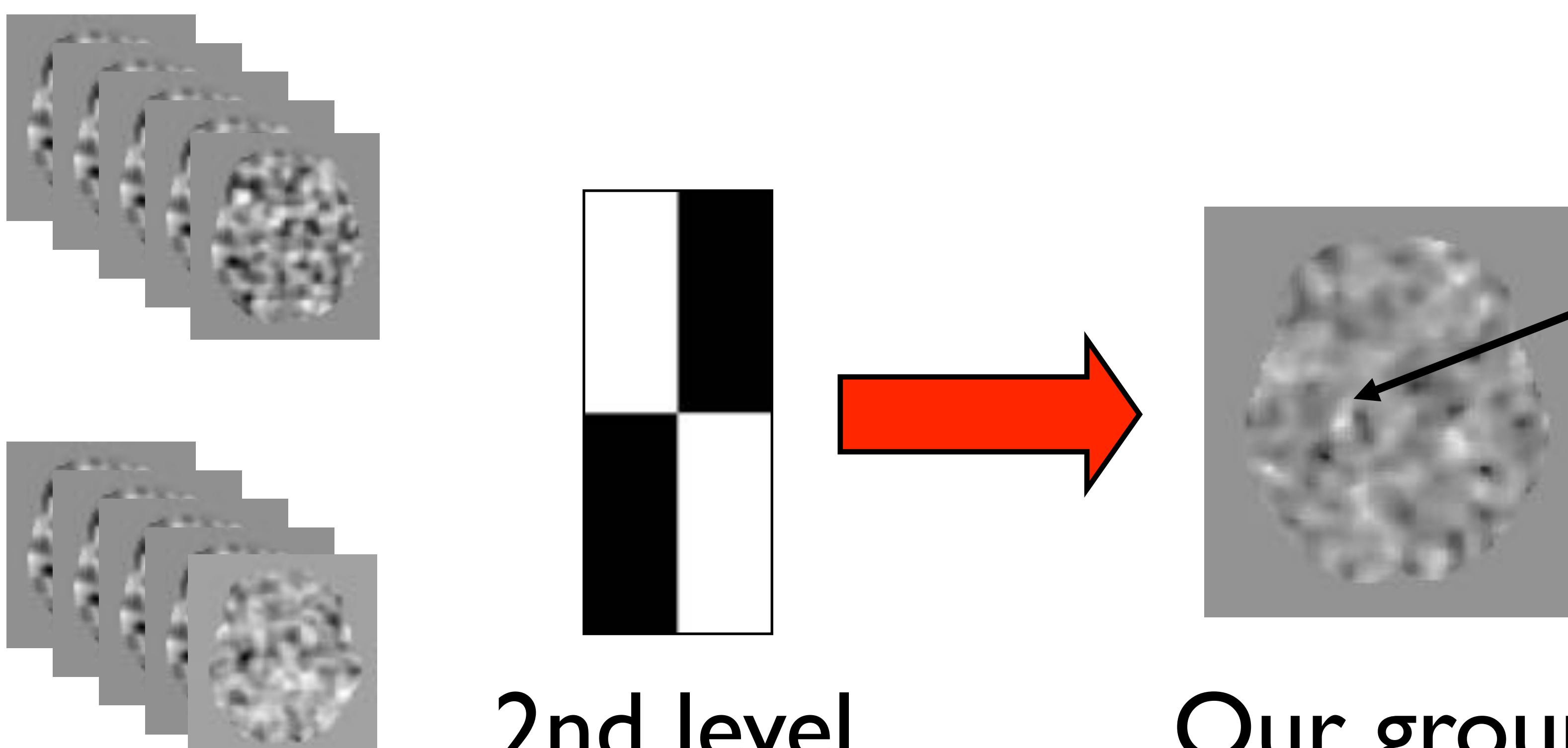


Very intriguing activation. $t_8 = 4.65$

But, can we trust it?

Group 1

Group 2



max(t)=4.65

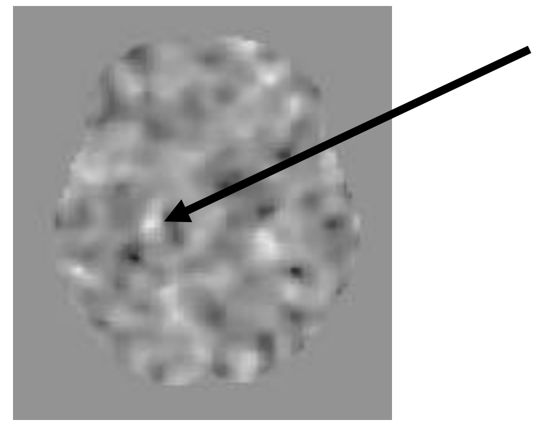
2nd level model

Our group difference map

And we can use this for any statistic

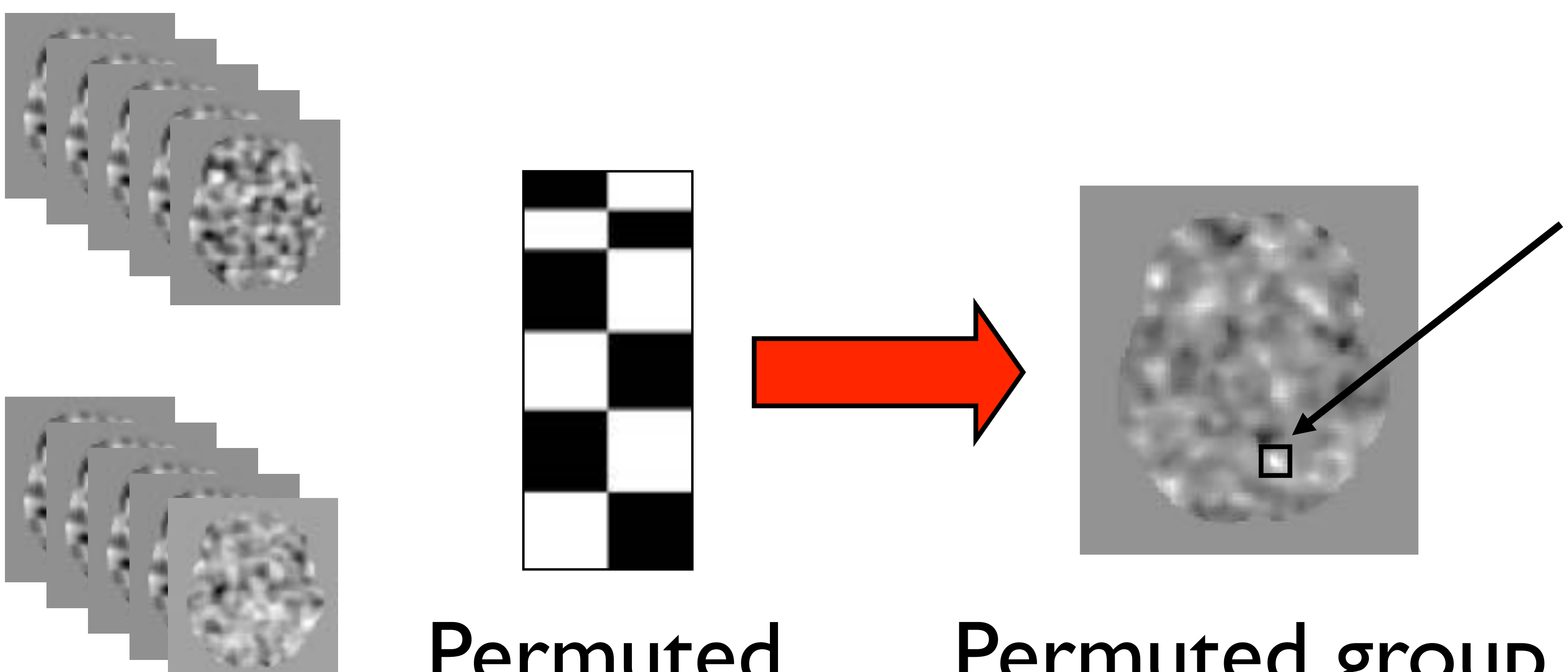
We compared activation by painful stimuli in two groups of 5 subjects each.

This is what we got



Very intriguing activation. $t_8 = 4.65$

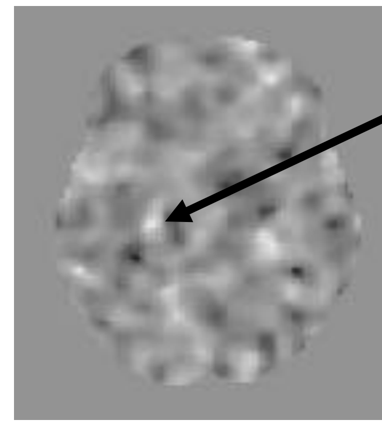
But, can we trust it?



And we can use this for any statistic

This is what we got

We compared activation by painful stimuli in two groups of 5 subjects each.

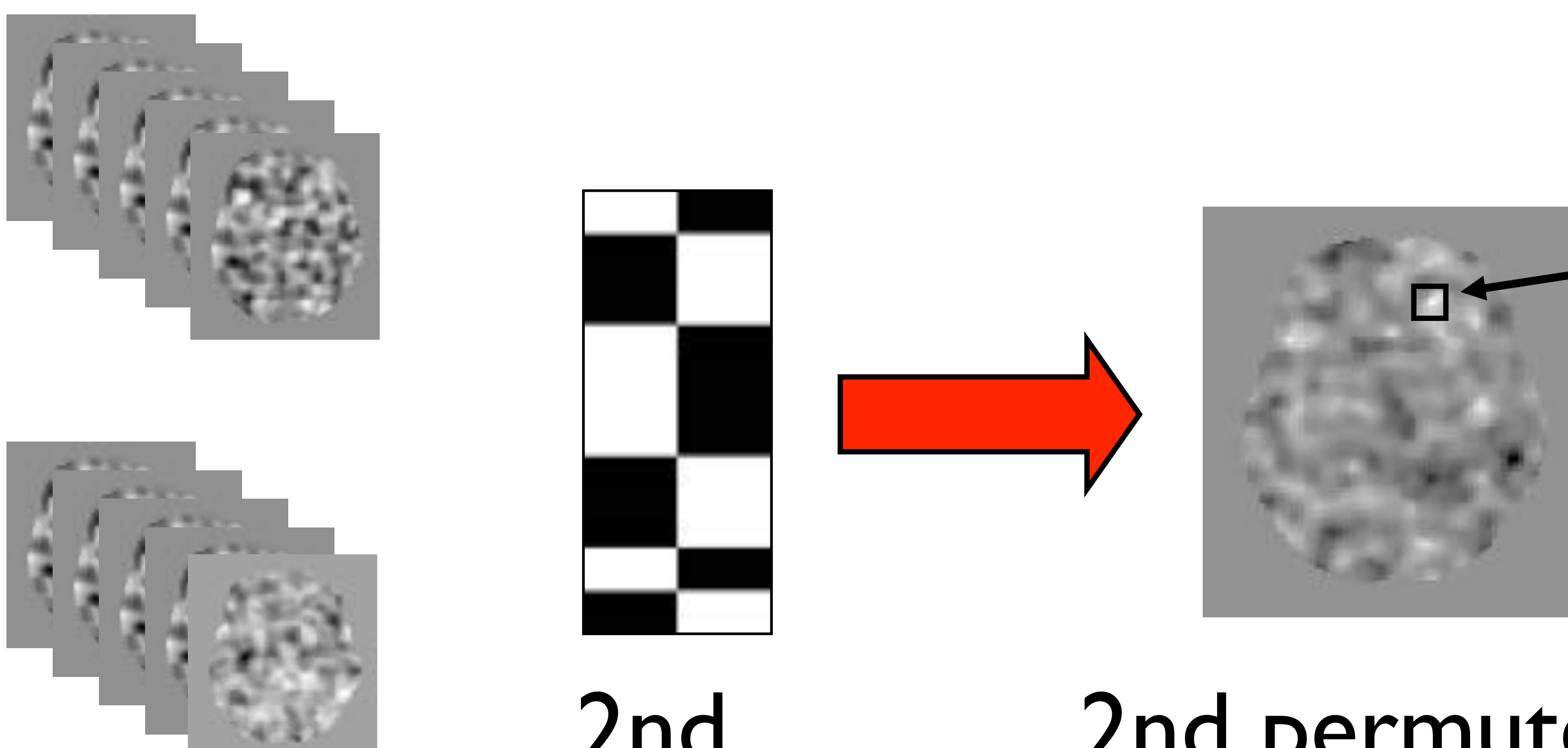


Very intriguing activation. $t_8 = 4.65$

But, can we trust it?

Group 1

Group 2



max(t)=5.43

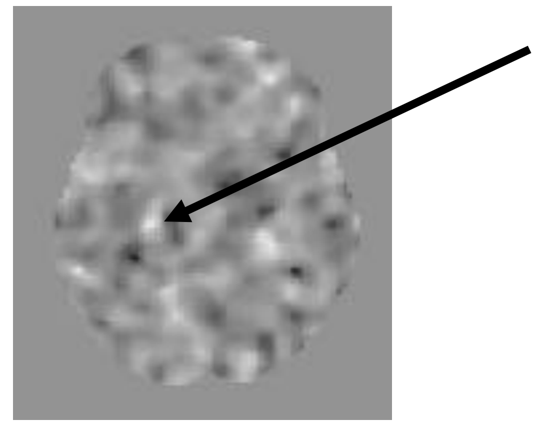
2nd Permutation

2nd permuted map

And we can use this for any statistic

This is what we got

We compared activation by painful stimuli in two groups of 5 subjects each.

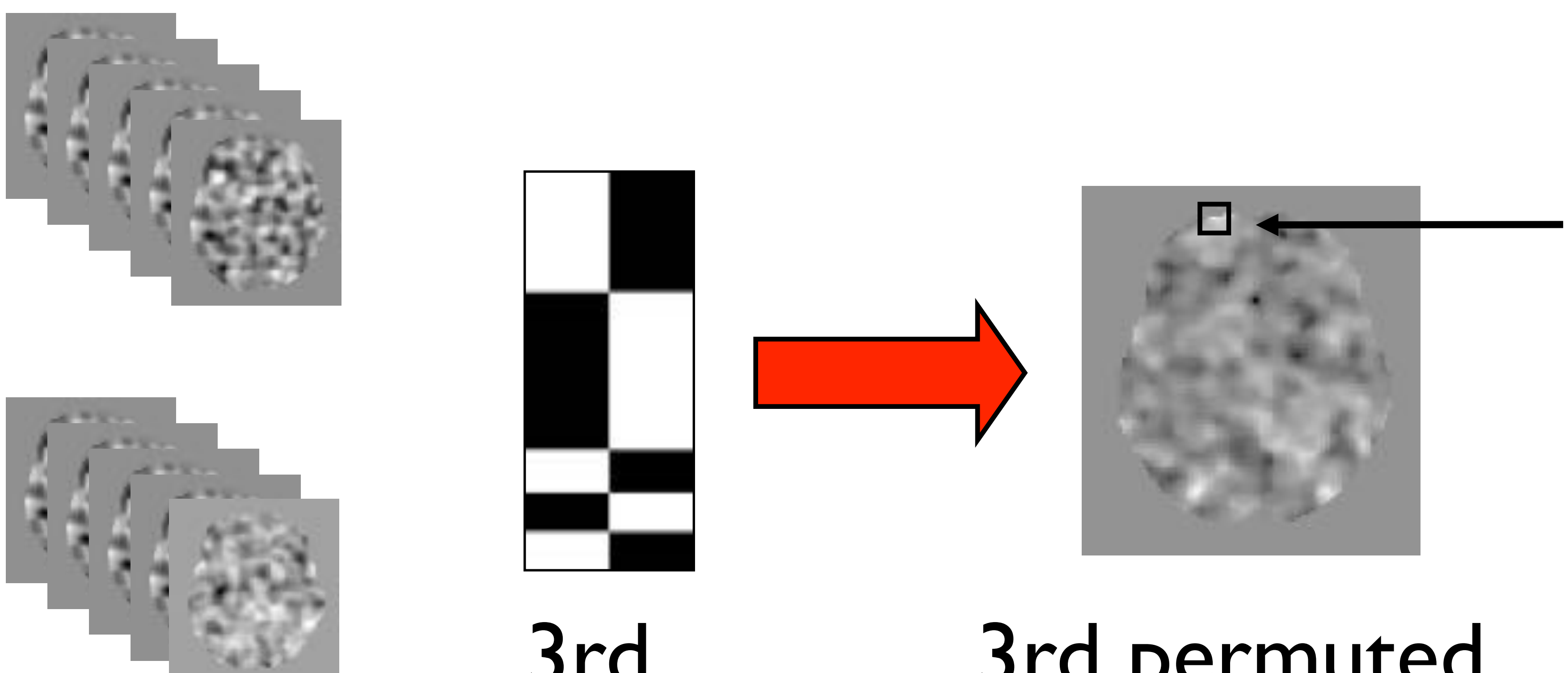


Very intriguing activation. $t_8 = 4.65$

But, can we trust it?

Group 1

Group 2



max(t)=5.84

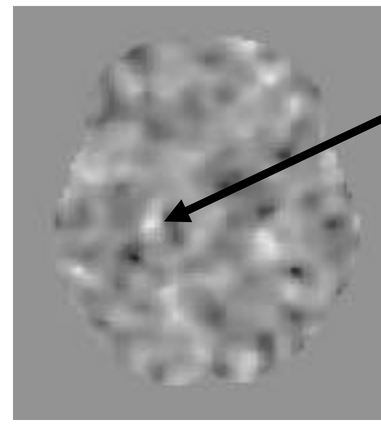
3rd Permutation

3rd permuted map

And we can use this for any statistic

This is what we got

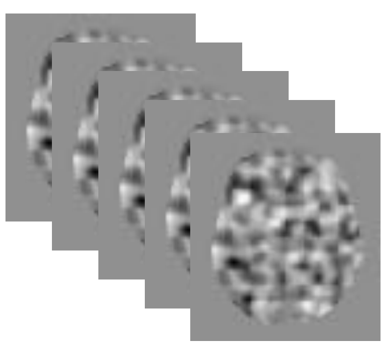
We compared activation by painful stimuli in two groups of 5 subjects each.



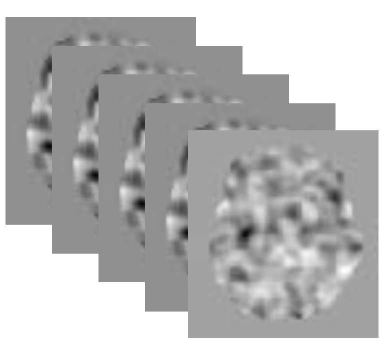
Very intriguing activation. $t_8 = 4.65$

But, can we trust it?

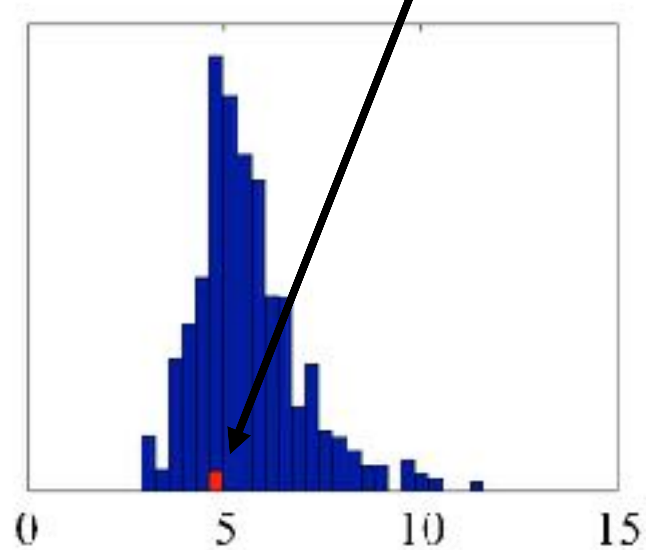
Group 1



Group 2



Original labelling

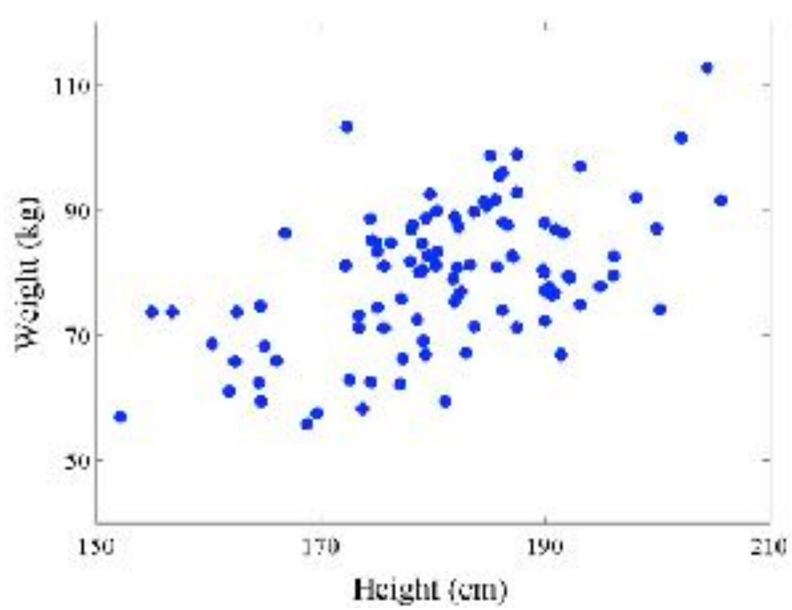


5000 permutations

3925 permutations yielded higher max(t)-value than original labelling. We cannot reject the null-hypothesis.

But beware the "exchangeability"

- When we swap the labels of two data-points we need to make sure that they are "exchangeable"
- I will start to explain "exchangeability" through a case that is not
- But first we need to learn about covariance matrices

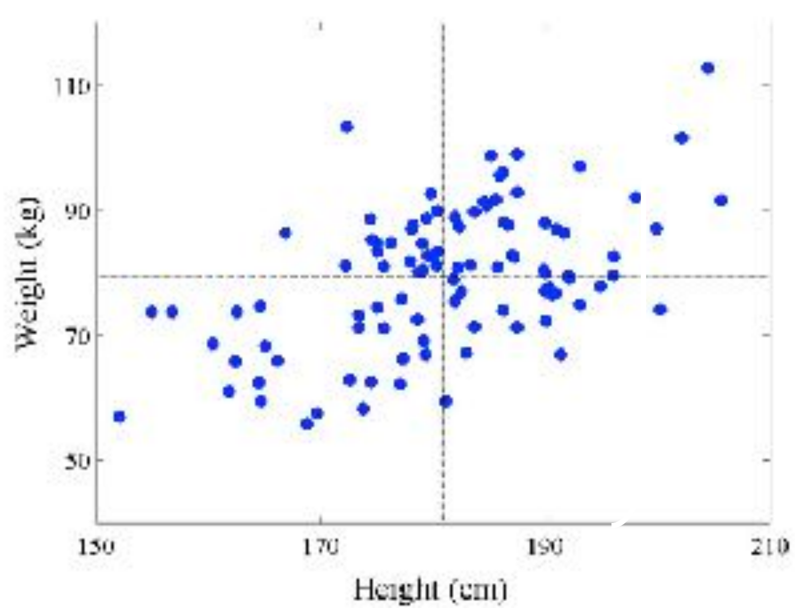


Height and weight of a random sample of Swedish men

Covariance matrices

- When we swap the labels of two data-points we need to make sure that they are "exchangeable"
- I will start to explain "exchangeability" through a case that is not
- But first we need to learn about covariance matrices

Mean height ≈181 cm

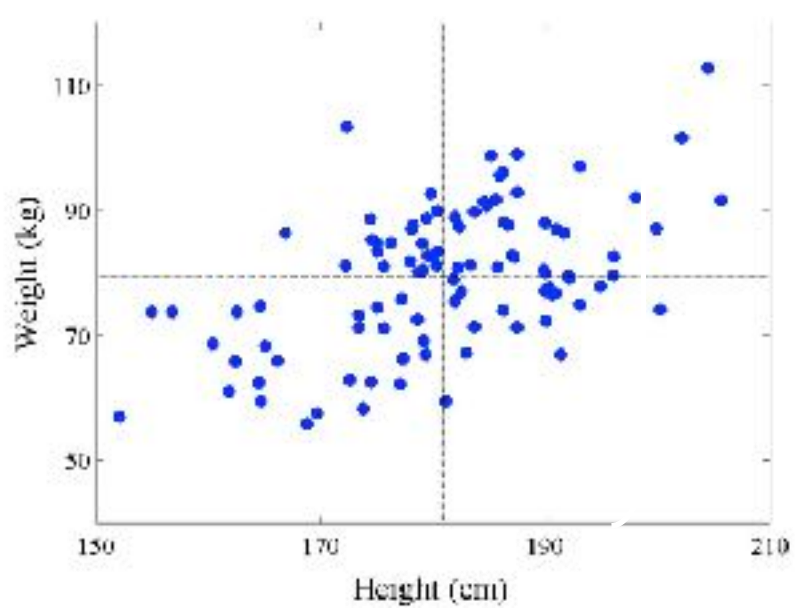


Mean weight ≈79.4 kg

Characterised by two means

Covariance matrices

- When we swap the labels of two data-points we need to make sure that they are "exchangeable"
- I will start to explain "exchangeability" through a case that is not
- But first we need to learn about covariance matrices

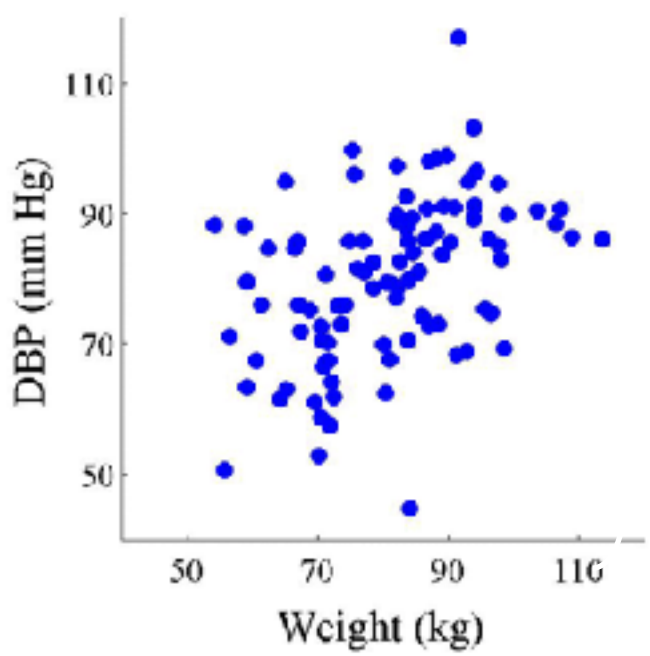


$$\Sigma = \begin{bmatrix} 130 & 52 \\ 52 & 165 \end{bmatrix}$$

And a covariance - matrix

Covariance matrices

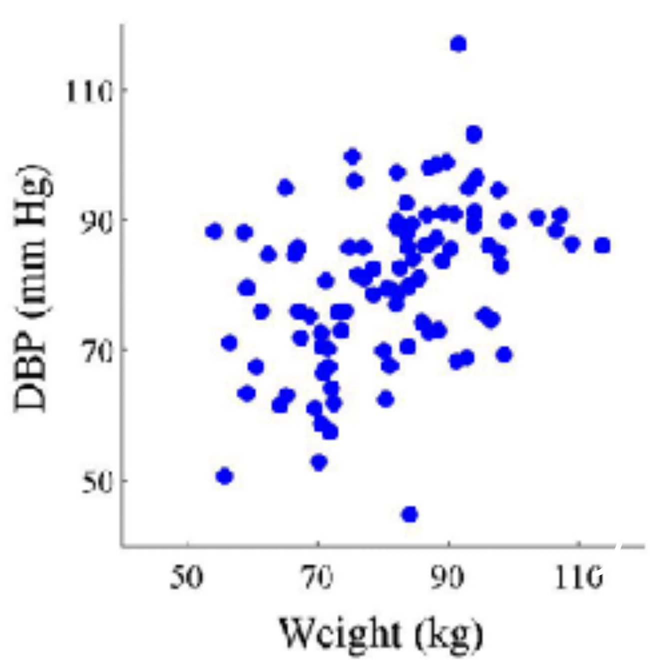
- When we swap the labels of two data-points we need to make sure that they are "exchangeable"
- I will start to explain "exchangeability" through a case that is not
- But first we need to learn about covariance matrices



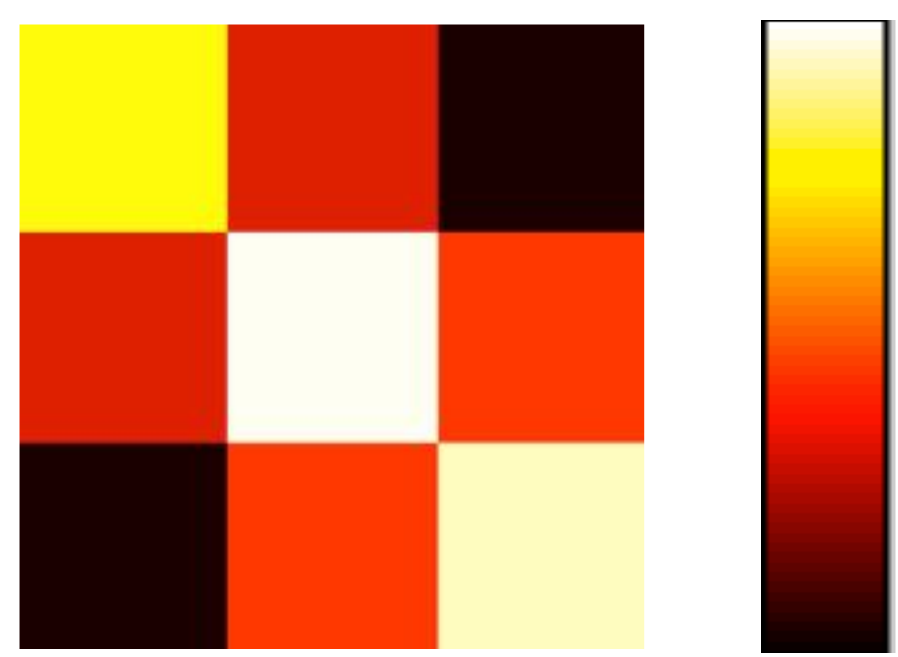
$$\Sigma = \begin{bmatrix} 130 & 52 & 4.8 \\ 52 & 165 & 69 \\ 4.8 & 69 & 156 \end{bmatrix}$$

Covariance matrices

- When we swap the labels of two data-points we need to make sure that they are "exchangeable"
- I will start to explain "exchangeability" through a case that is not
- But first we need to learn about covariance matrices

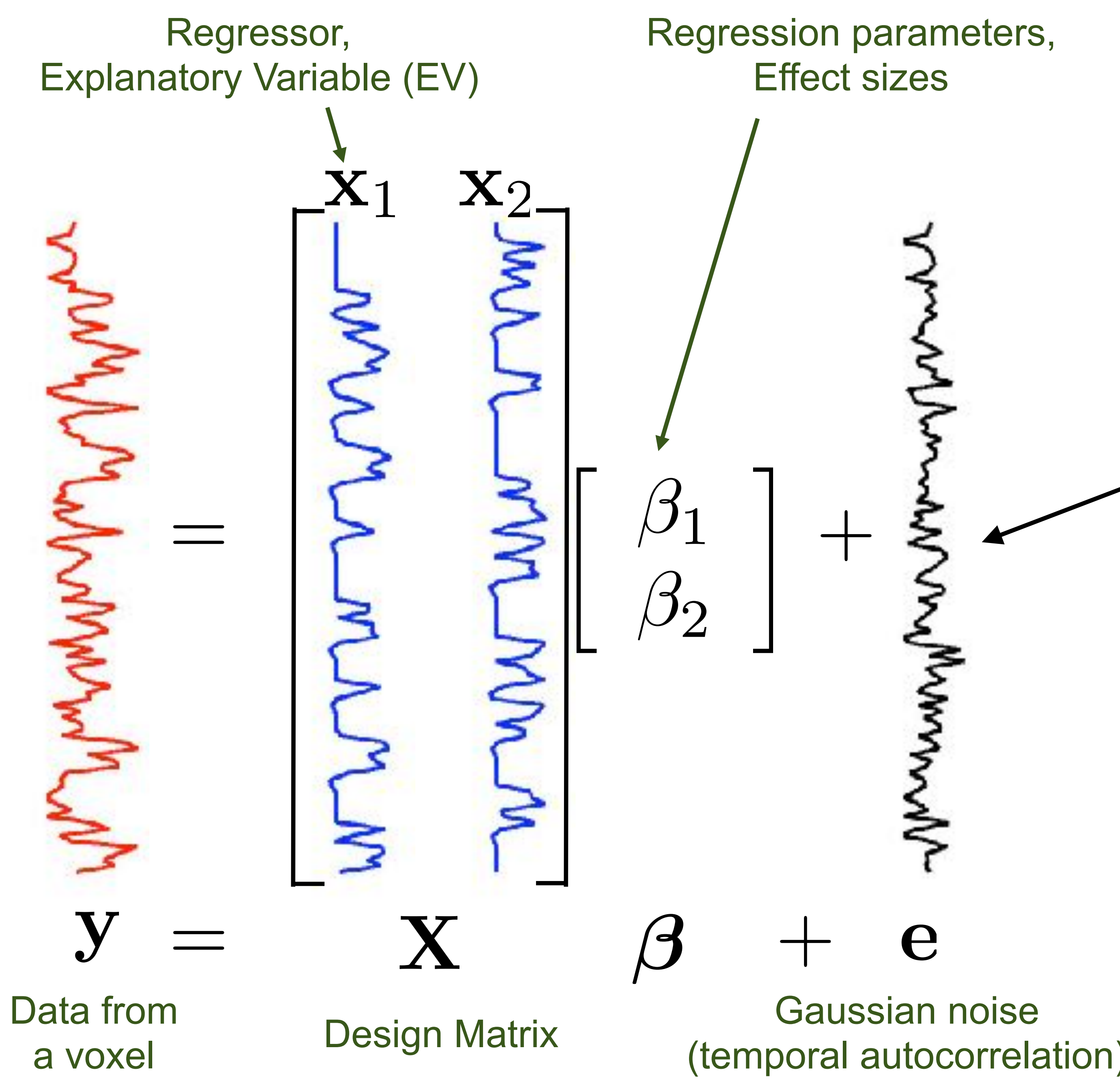


$\Sigma =$



1st level fMRI data is not exchangeable

- You may, or may not, have seen this slide in the 1st level GLM talk.



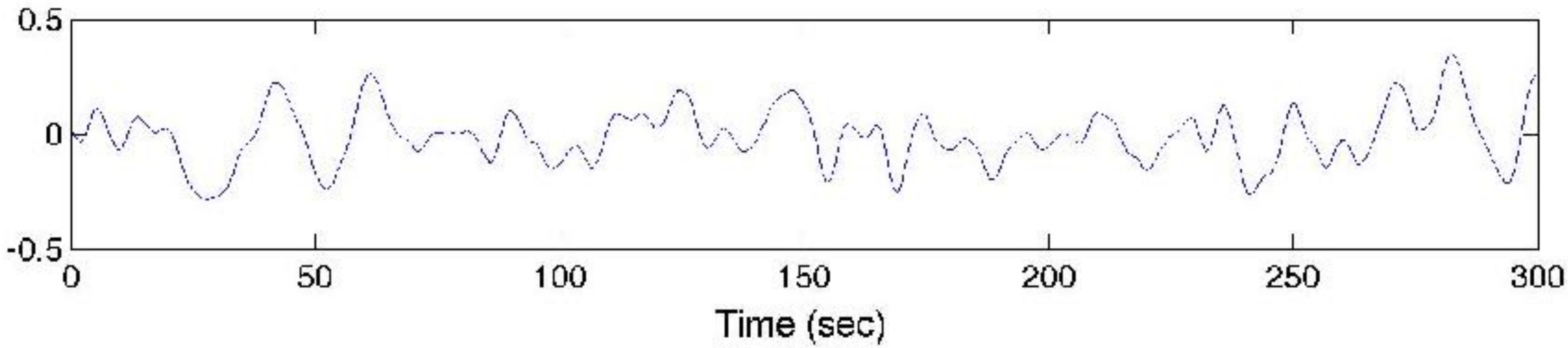
This time we will look more closely at this part

$$\mathbf{e} \sim N(\mathbf{0}, \Sigma)$$

Our old friend "the covariance matrix"

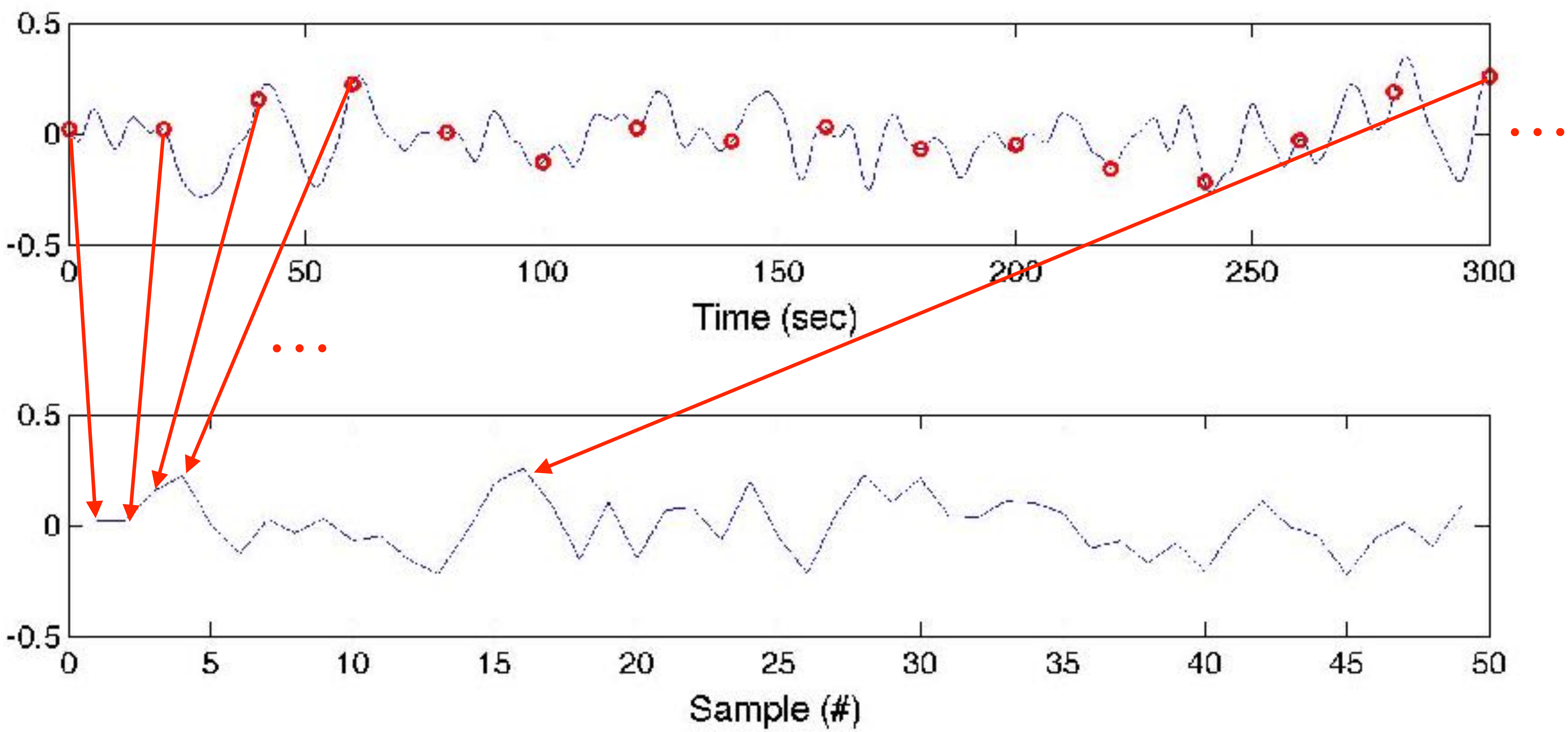
1st level fMRI data is not exchangeable

- One important component of noise in fMRI consists of physiological/neuronal events convolved by the HRF



1st level fMRI data is not exchangeable

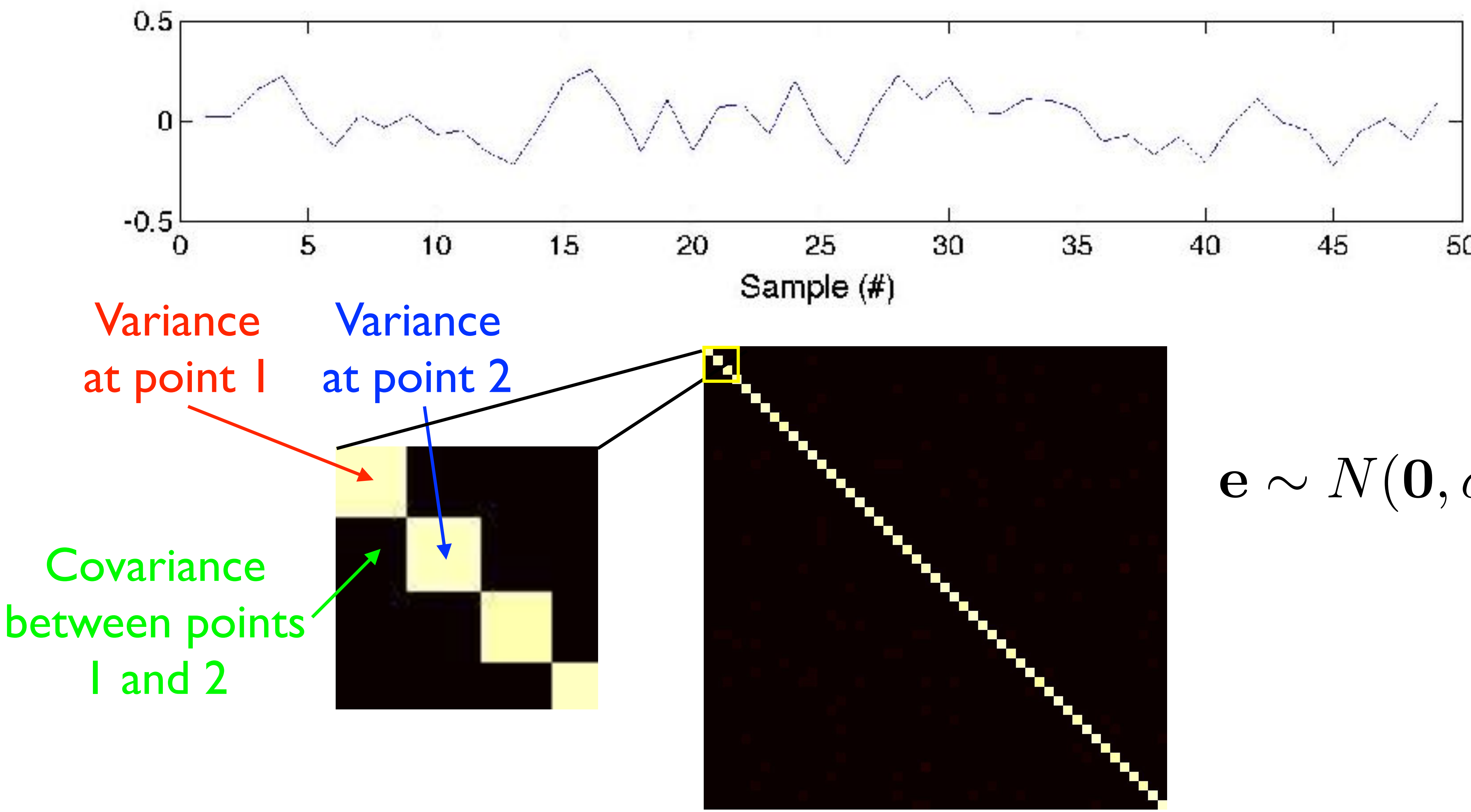
- One important component of noise in fMRI consists of physiological/neuronal events convolved by the HRF



If we sample this every 20 seconds it no longer looks “smooth”

1st level fMRI data is not exchangeable

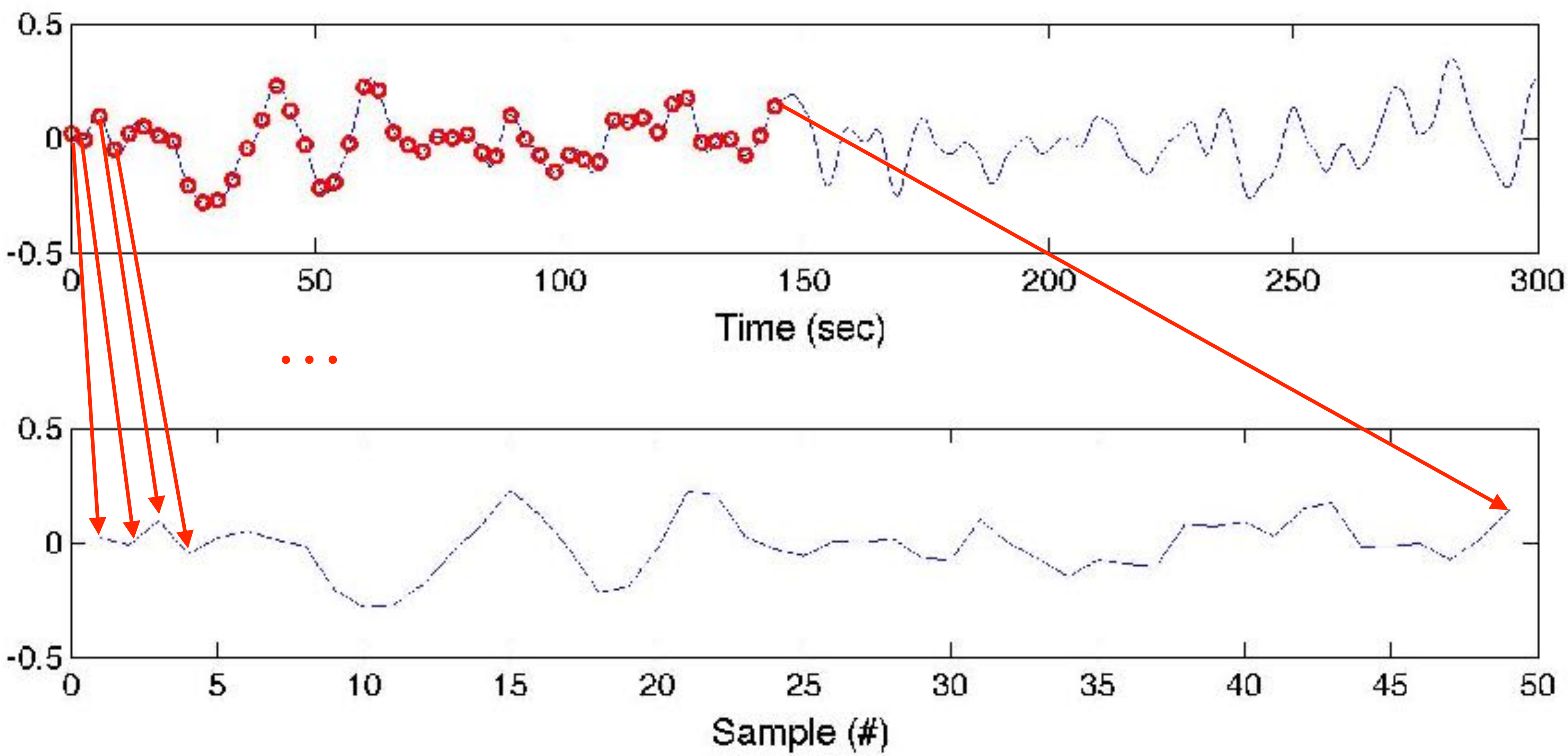
- One important component of noise in fMRI consists of physiological/neuronal events convolved by the HRF



$$\mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

1st level fMRI data is not exchangeable

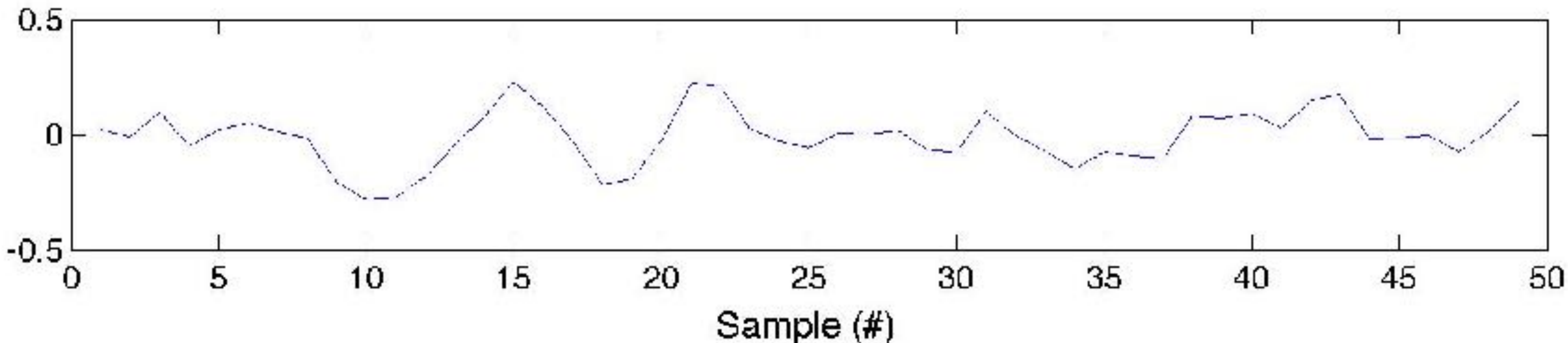
- One important component of noise in fMRI consists of physiological/neuronal events convolved by the HRF

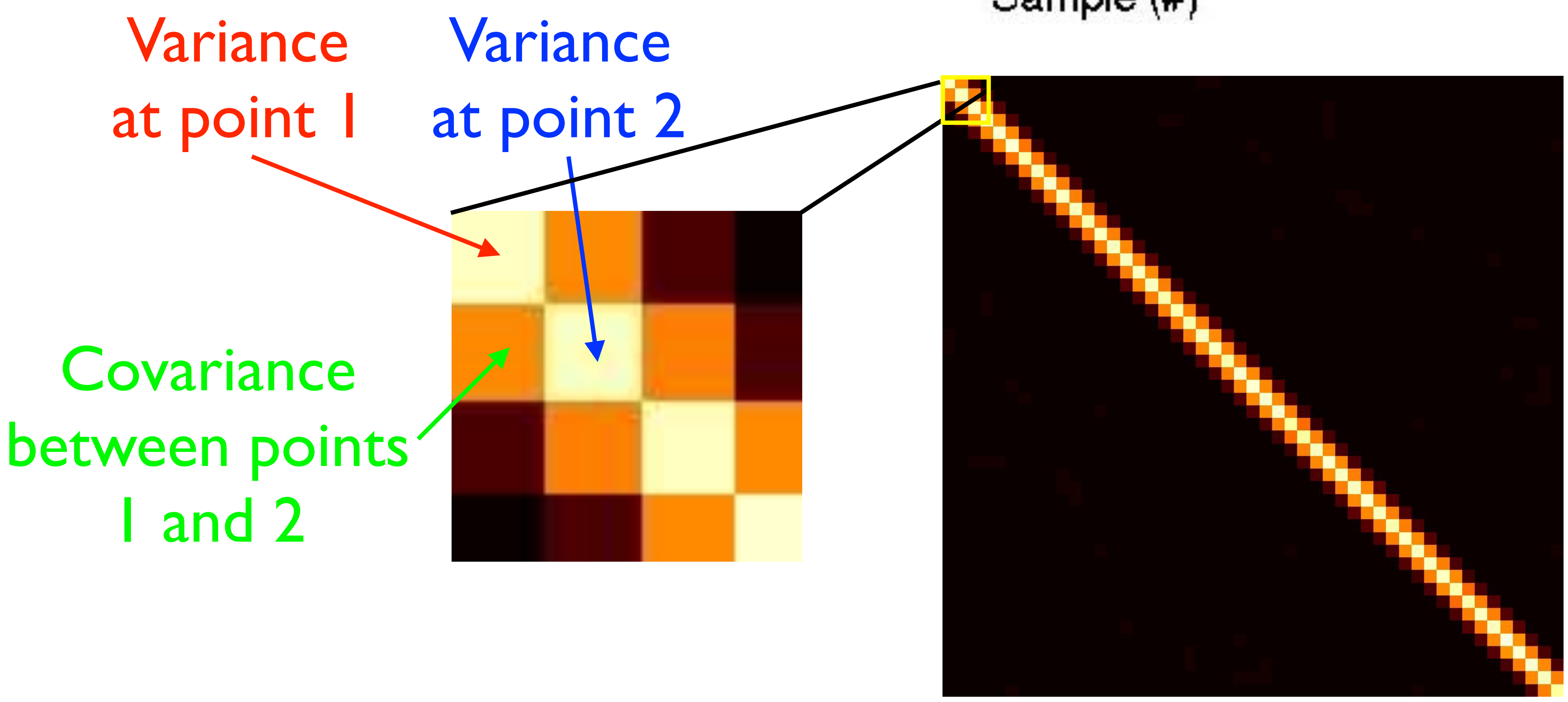


But that is not a realistic TR. What about every 3 seconds?

1st level fMRI data is not exchangeable

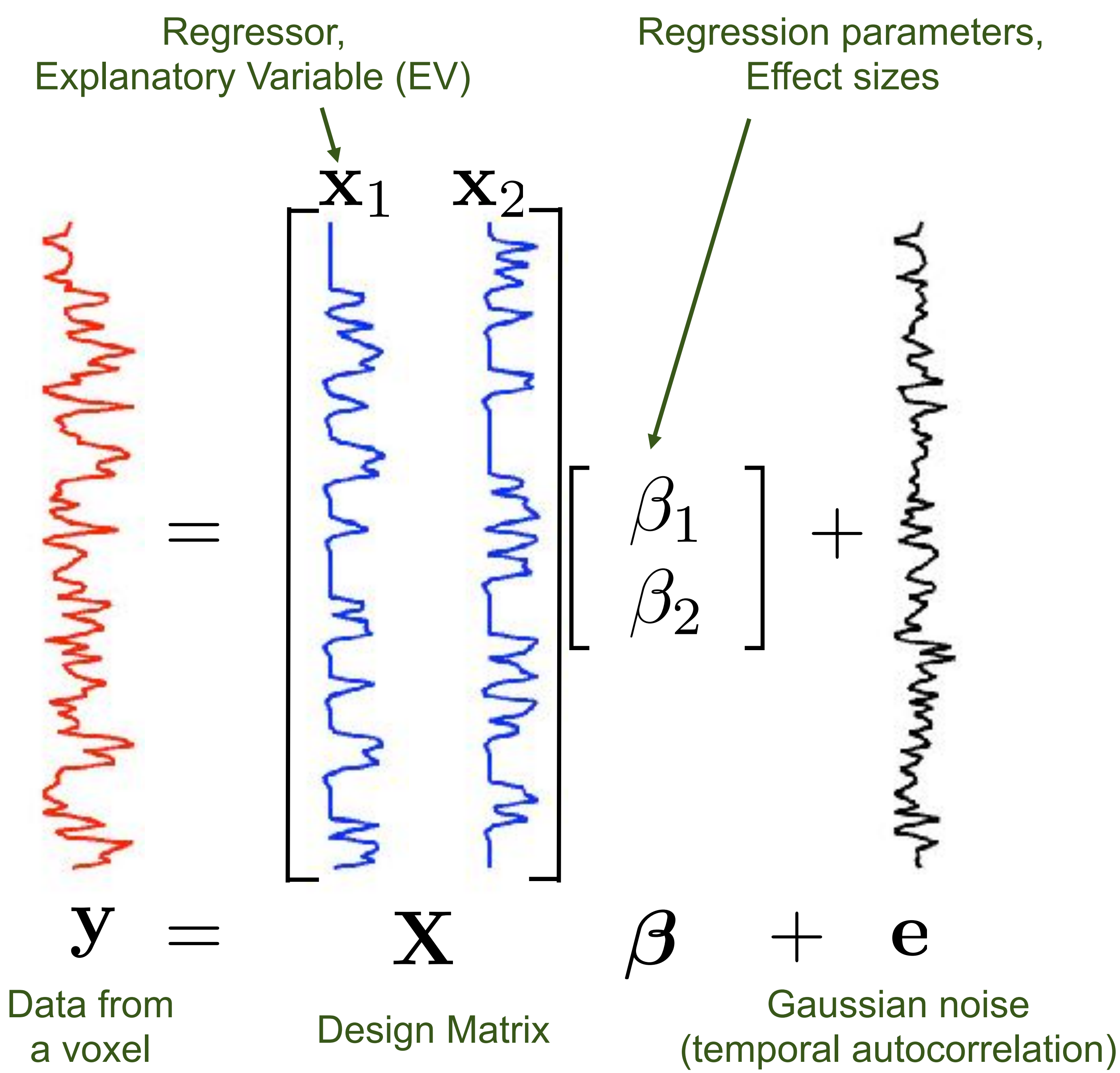
- One important component of noise in fMRI consists of physiological/neuronal events convolved by the HRF





1st level fMRI data is not exchangeable

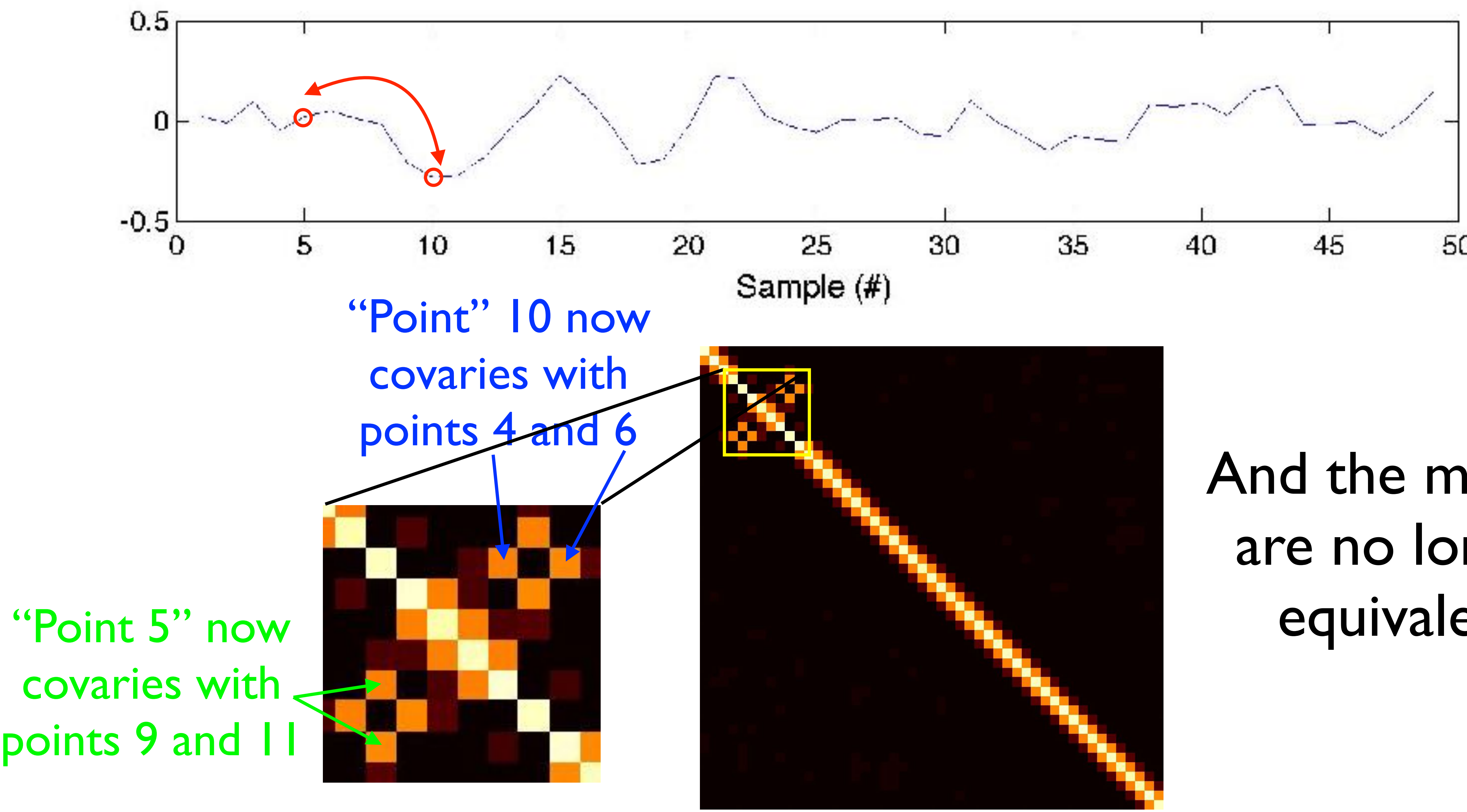
- Let us now return to our model again



- The model consists of our regressors $\mathbf{X}$ and the noise model $\mathbf{e} \sim N(\mathbf{0}, \Sigma)$
- All permutations must result in “equivalent models”
- Let us now see what happens if we swap two data-points (points 5 and 10)

1st level fMRI data is not exchangeable

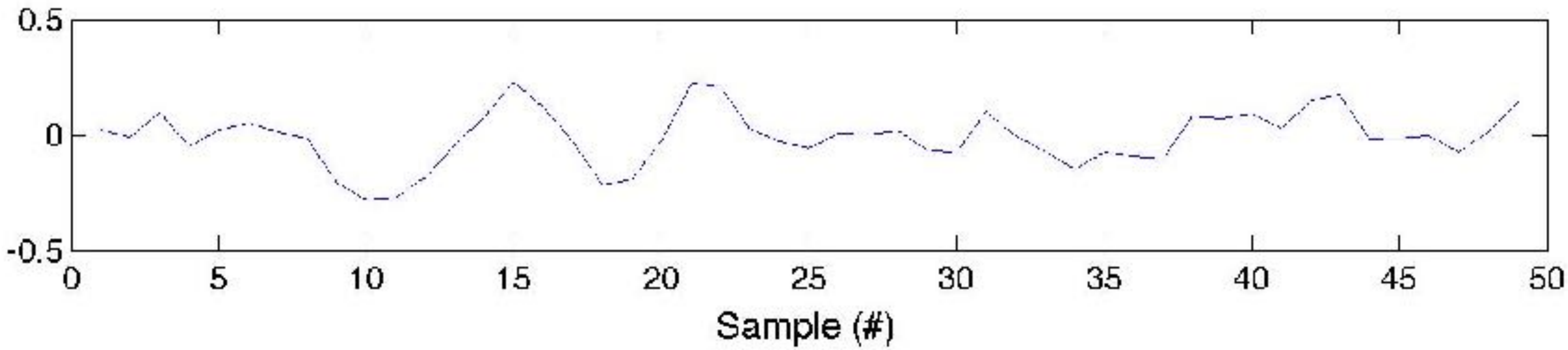
- One important component of noise in fMRI consists of physiological/neuronal events convolved by the HRF



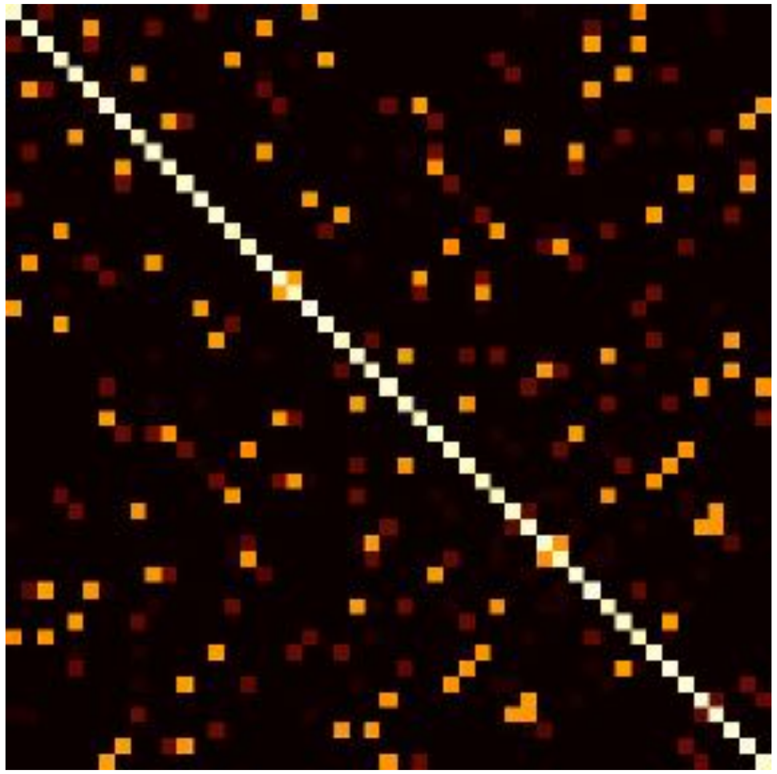
And the models are no longer equivalent

1st level fMRI data is not exchangeable

- One important component of noise in fMRI consists of physiological/neuronal events convolved by the HRF



And for a random permutation …

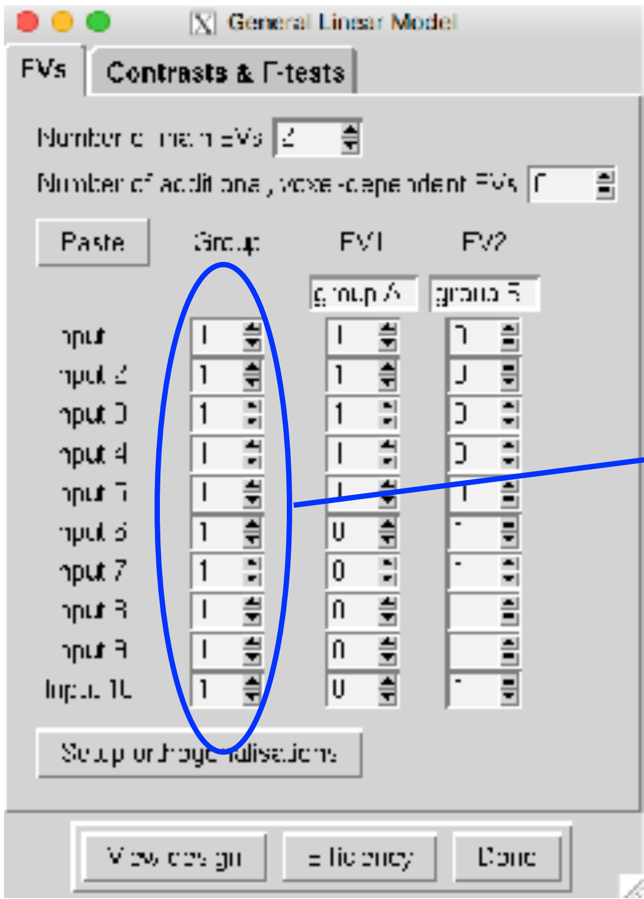


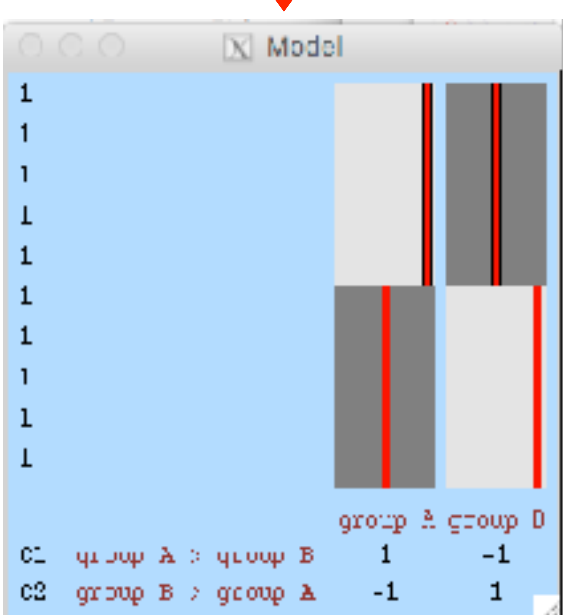
And the models are no longer equivalent

Back to exchangeability

- Data-points are not "exchangeable" if swapping them means that the noise covariance-matrix ends up looking different.
- Formally we say that "The joint distribution of the data must be unchanged by the permutations under the null-hypothesis".
- If the noise covariance-matrix has non-zero off-diagonal elements (covariances) you need to beware.
- You typically never estimate or see the covariance-matrix. You need to "imagine it" and determine from that if there is a problem.

Examples of exchangeability: Two groups unpaired

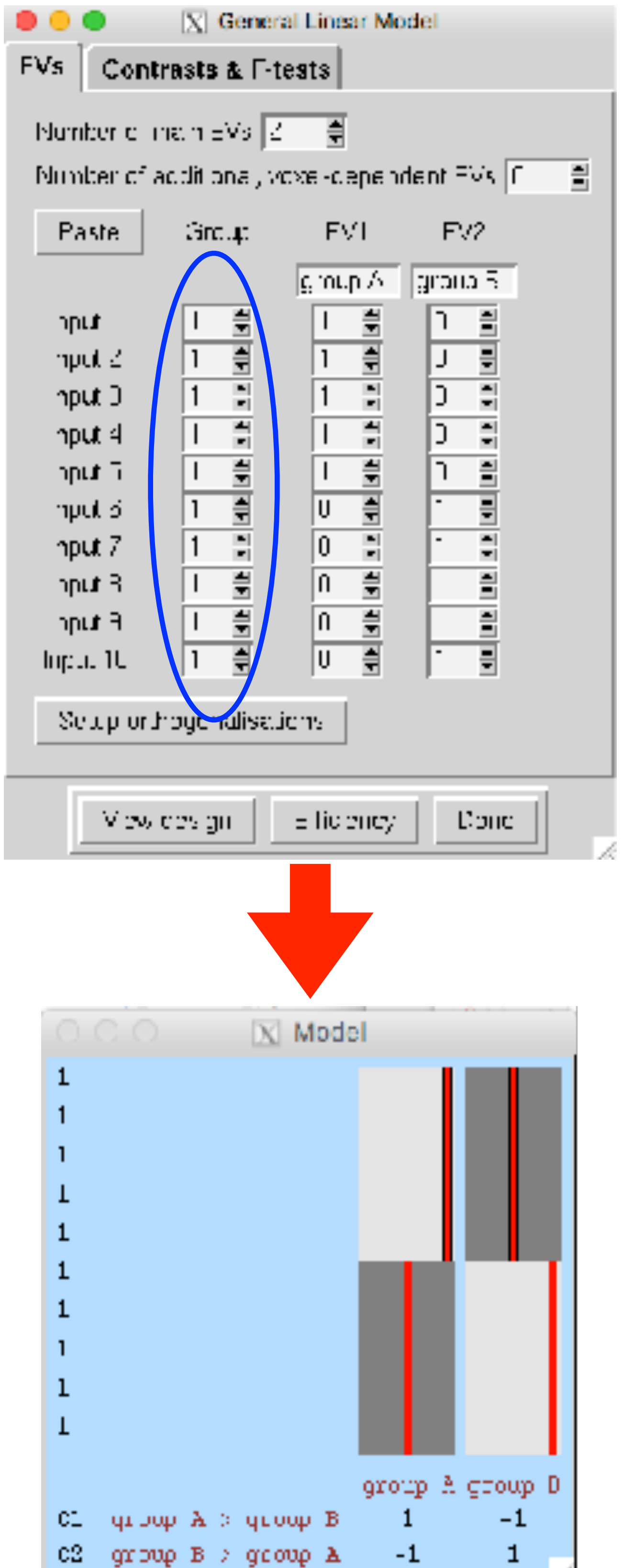




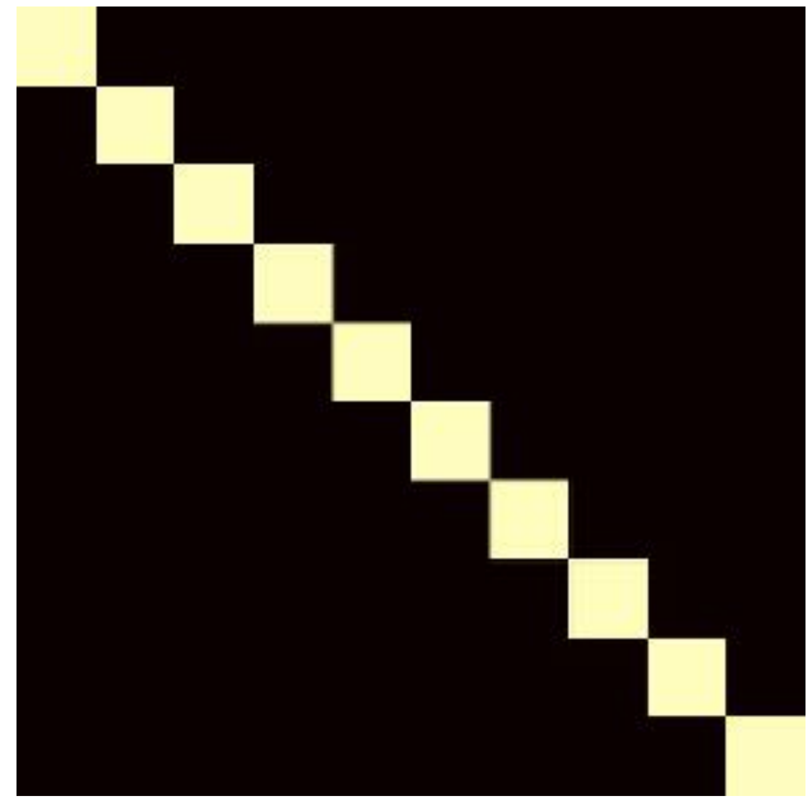
This is the "exchangeability group". Here all scans are in the same group, which means any scan can be exchanged for any other.

N.B. The "group" labelling is used for completely different purposes when using FLAME/GRFT

Examples of exchangeability: Two groups unpaired

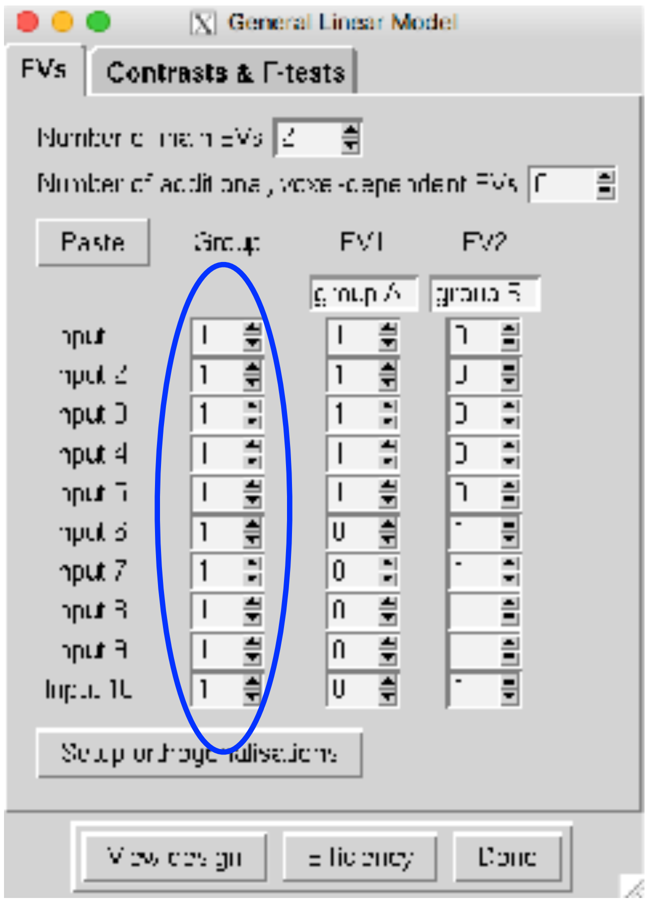


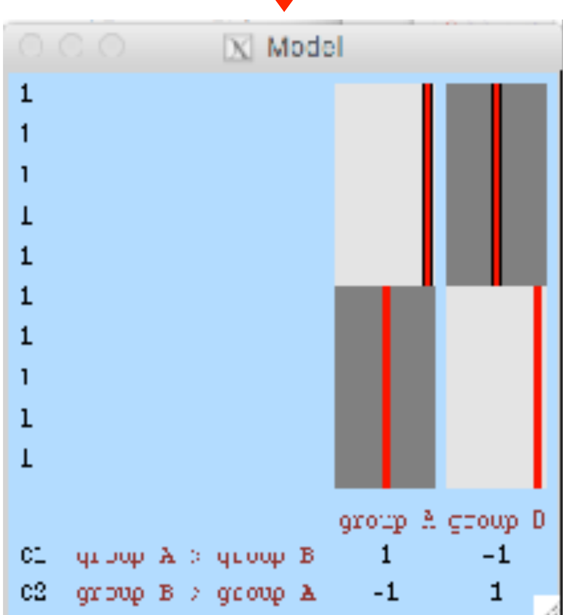
Assumed covariance matrix



The implicit assumption here is that data from all subjects have the same uncertainty and are all independent

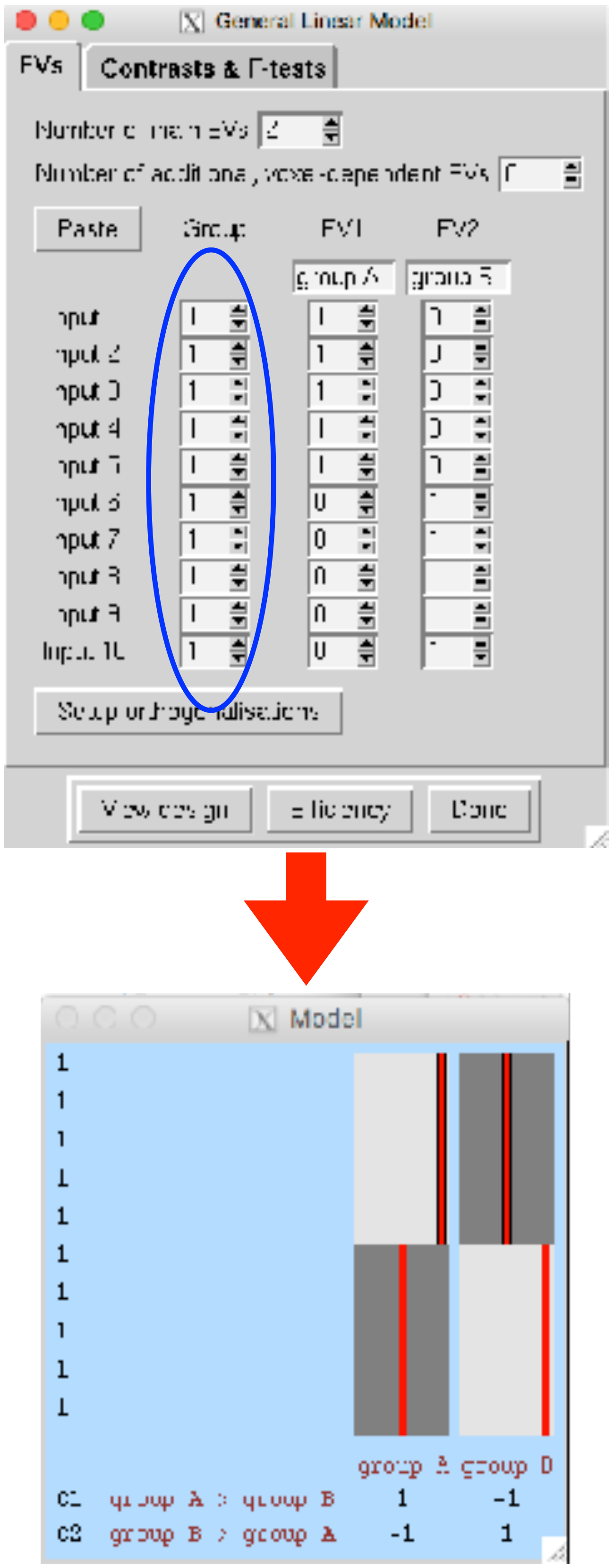
Examples of exchangeability: Two groups unpaired





Original	Perm 1	Perm 2	…
1	6	6	
2	3	1	
3	7	7	
4	8	4	
5	5	9	
6	1	5	
7	2	8	
8	4	3	
9	9	10	
10	10	2	

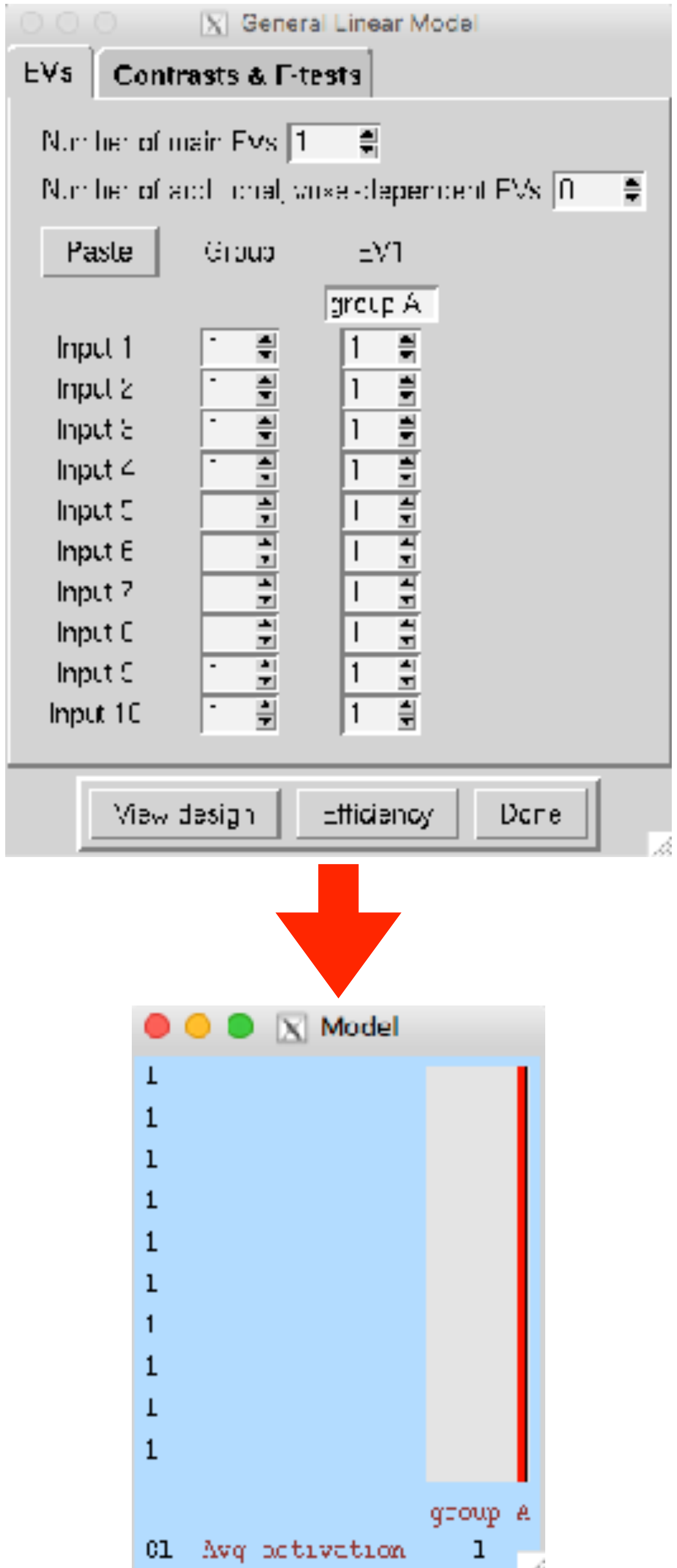
Examples of exchangeability: Two groups unpaired



Original	Perm 1	Perm 2	~~Perm 3~~
1	6	6	1
2	3	1	7
3	7	7	6
4	8	4	9
5	5	9	4
6	1	5	2
7	2	8	8
8	4	3	5
9	9	10	3
10	10	2	10

N.B. Equivalent

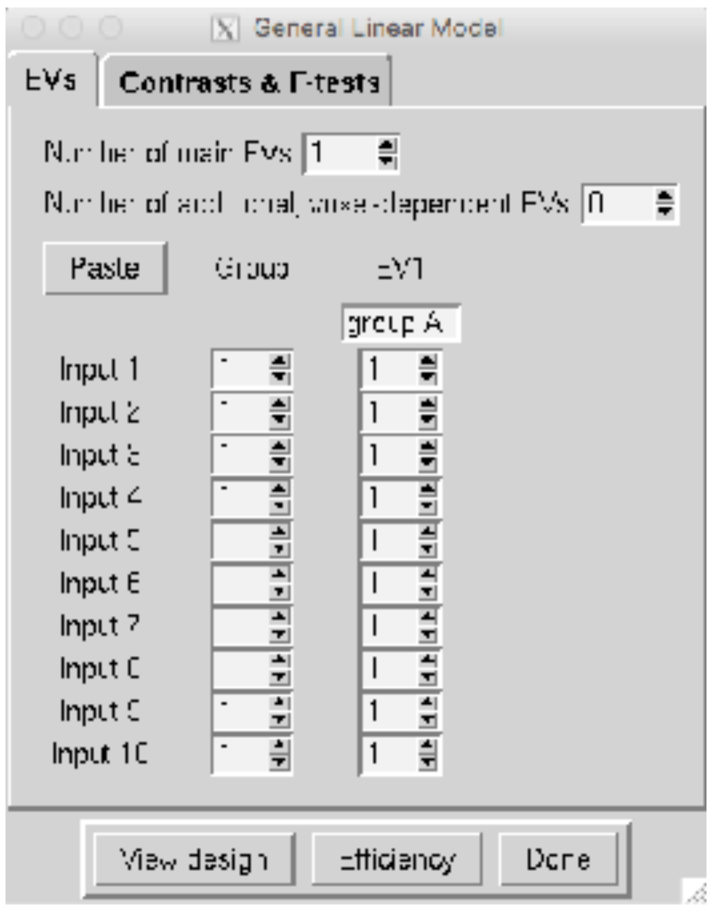
Examples of exchangeability: Single group average

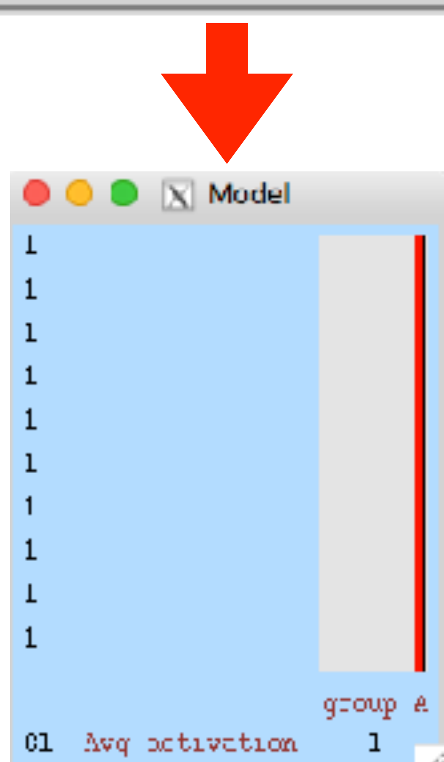


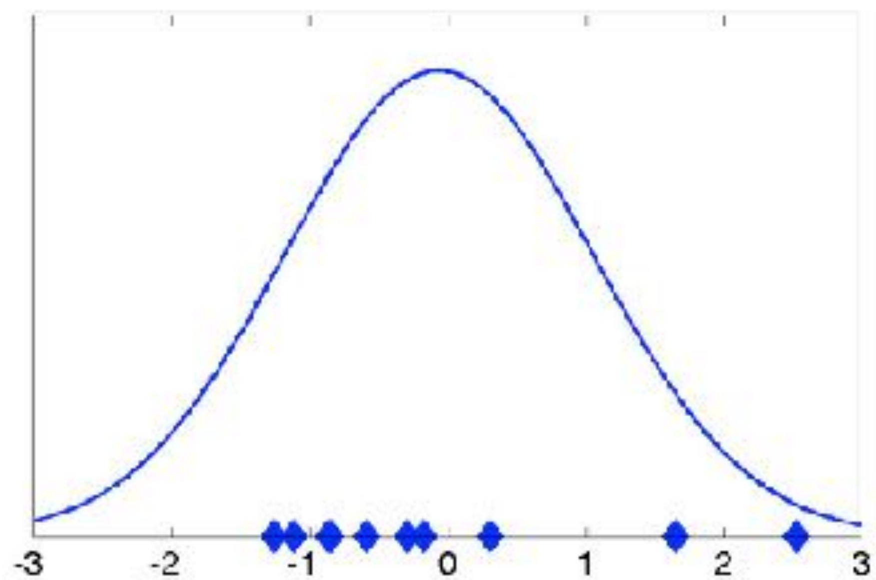
Here we model a single mean and want to know if that is different from zero

But there isn't really anything to permute, or is there?

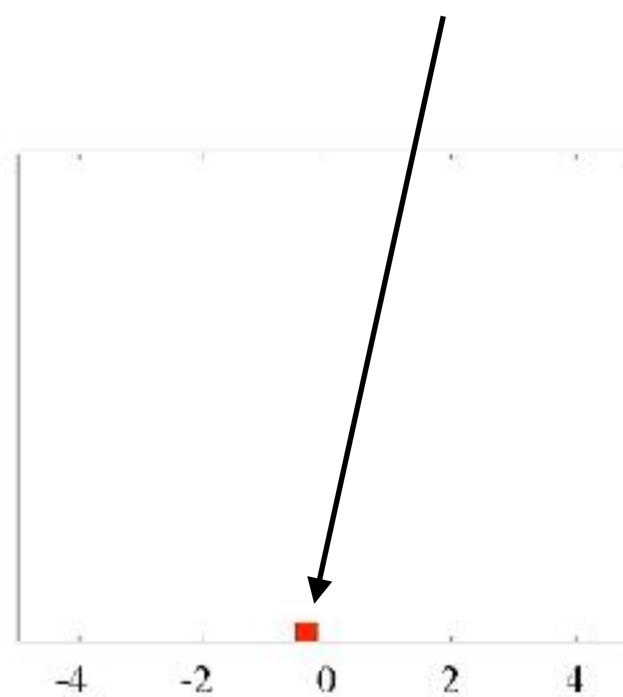
Examples of exchangeability: Single group average



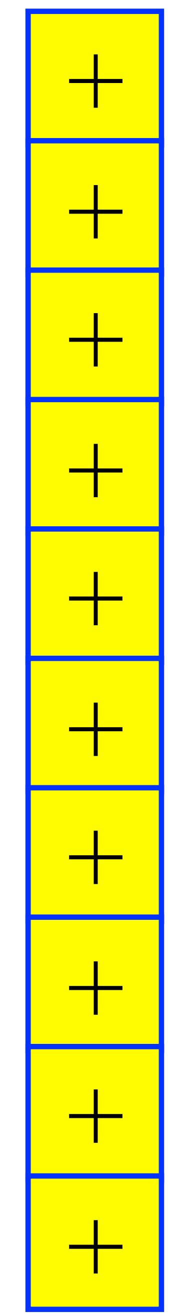




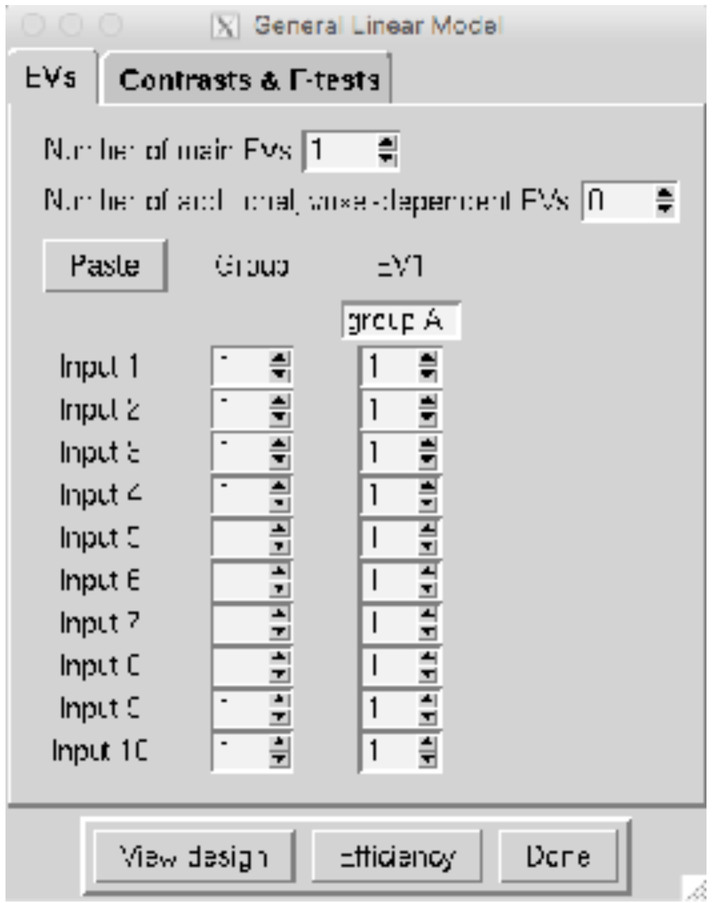
$$t = -0.17$$

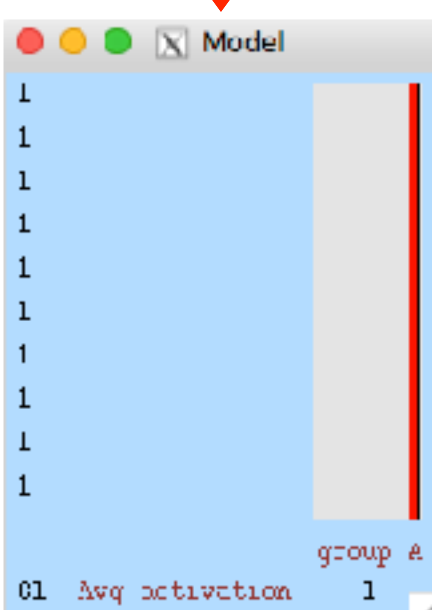


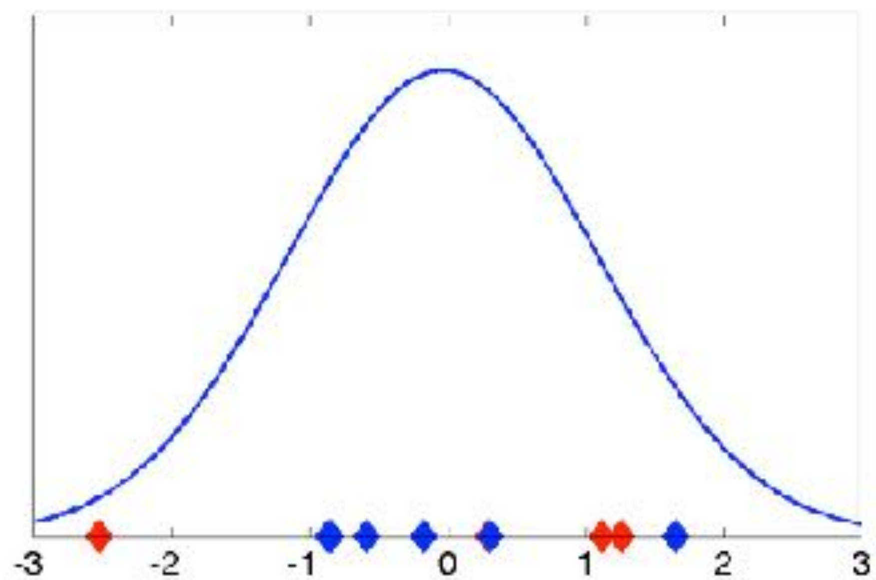
Original



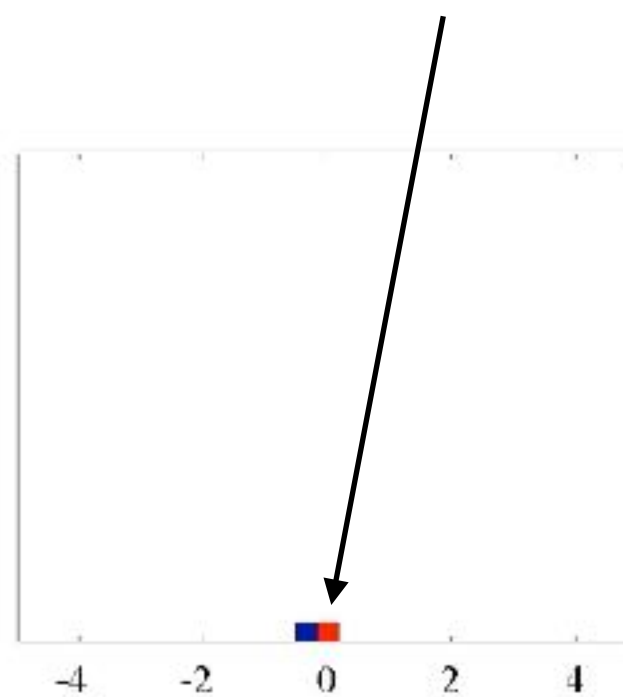
Examples of exchangeability: Single group average



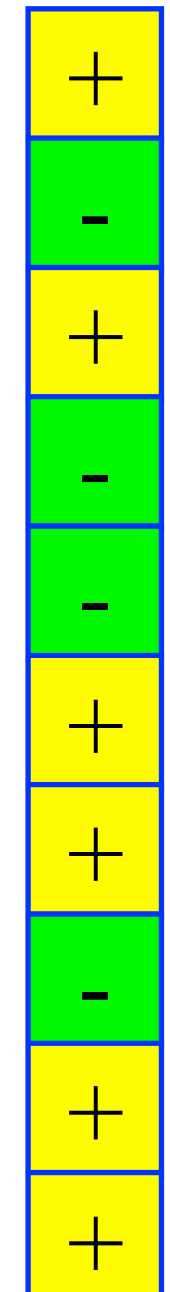




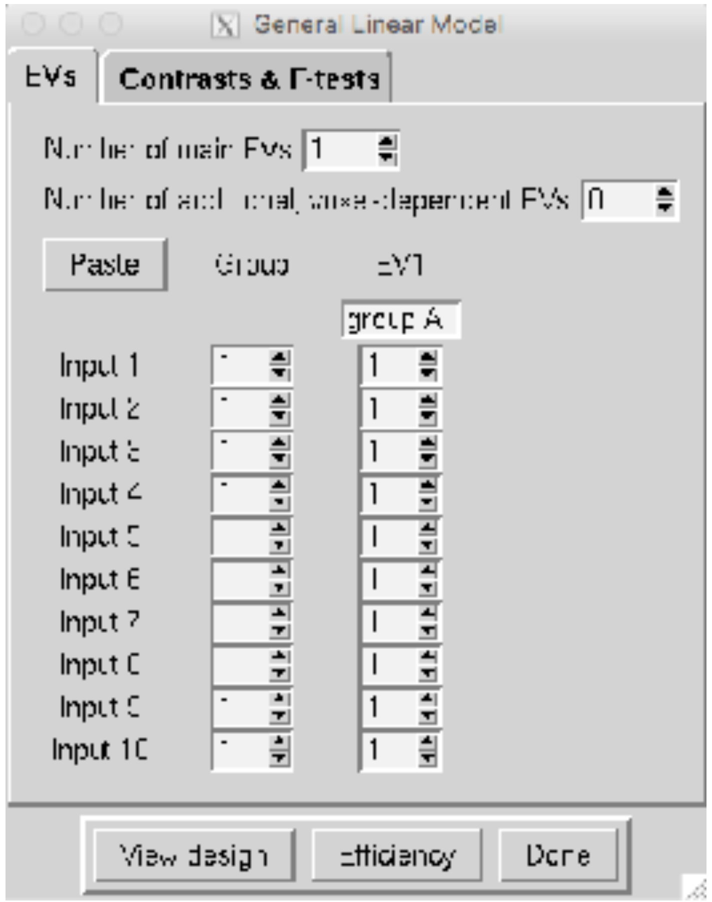
$$t = -0.09$$

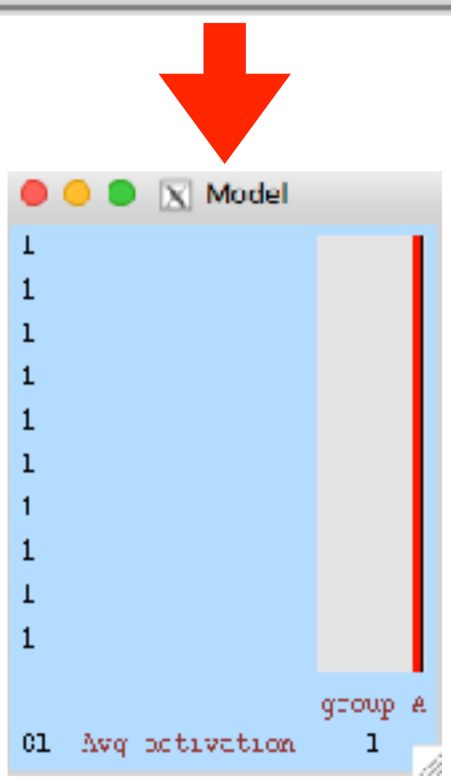


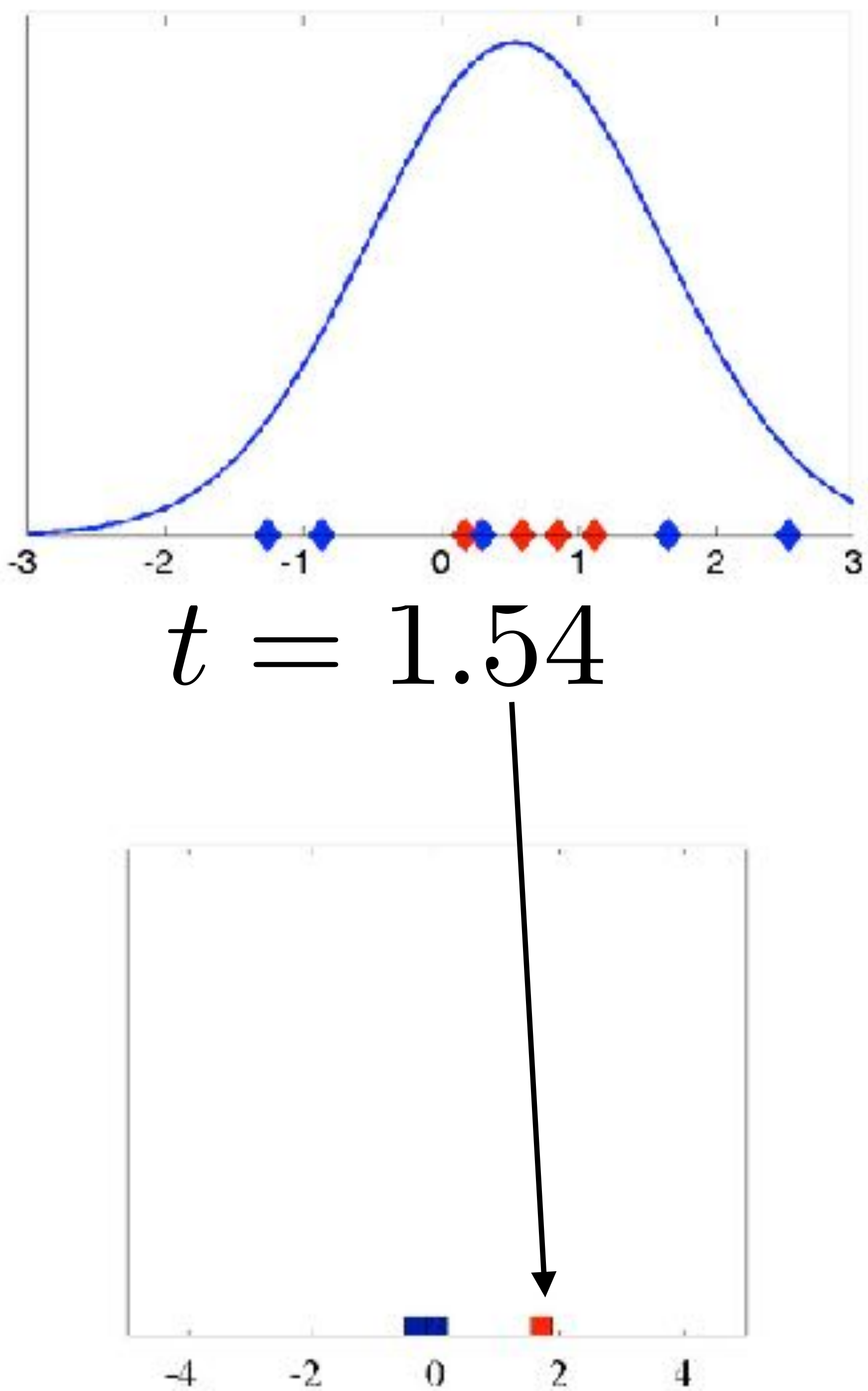
First flip



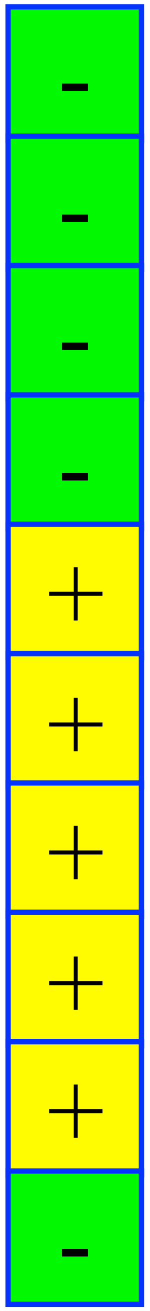
Examples of exchangeability: Single group average



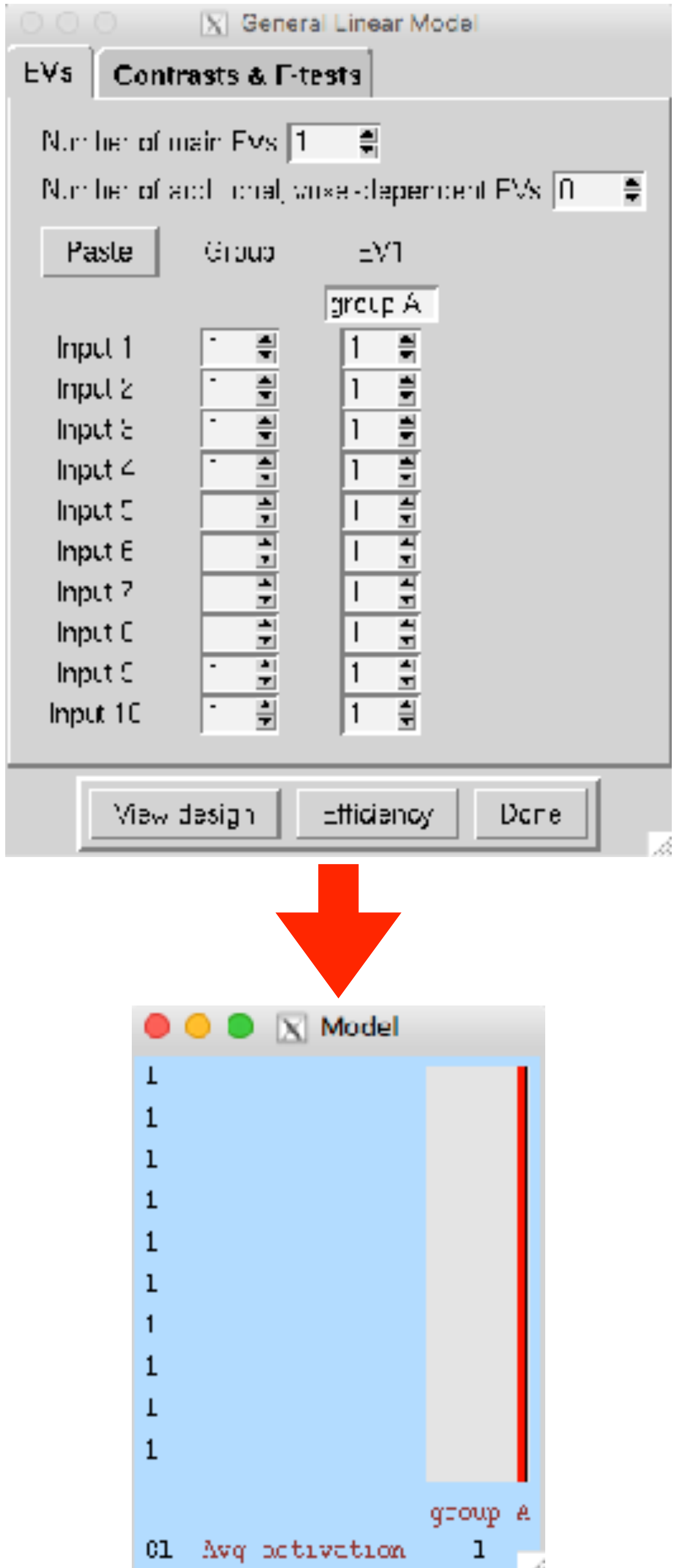




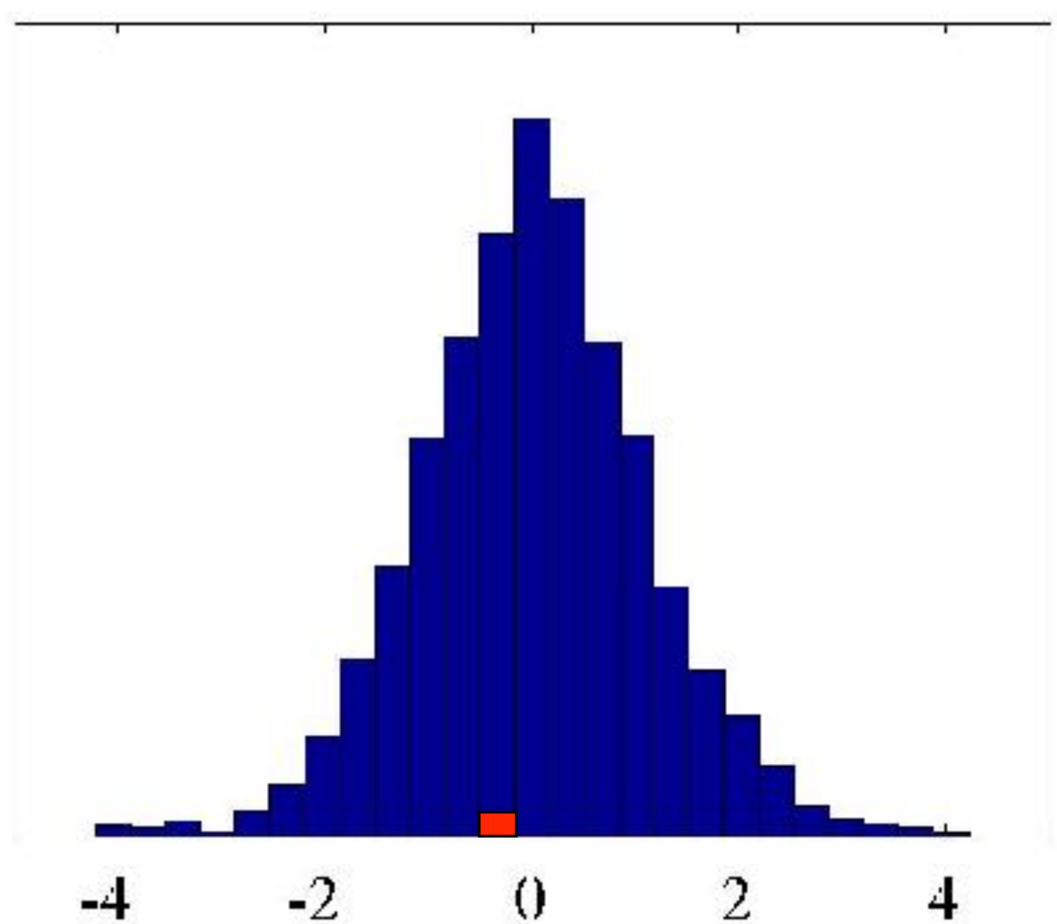
Second flip

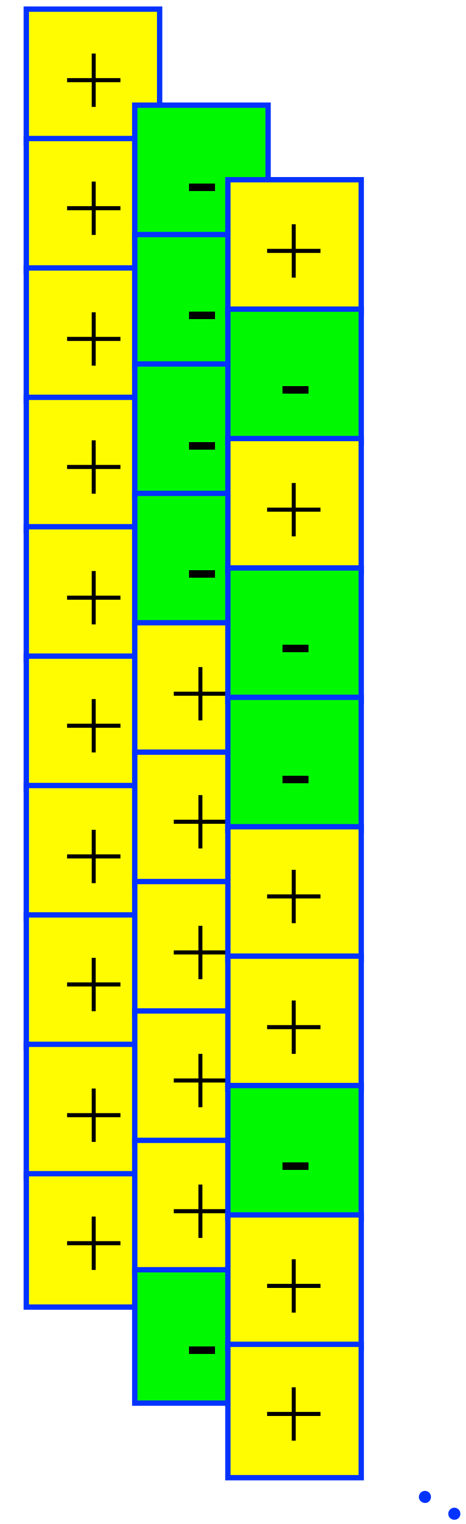


Examples of exchangeability: Single group average

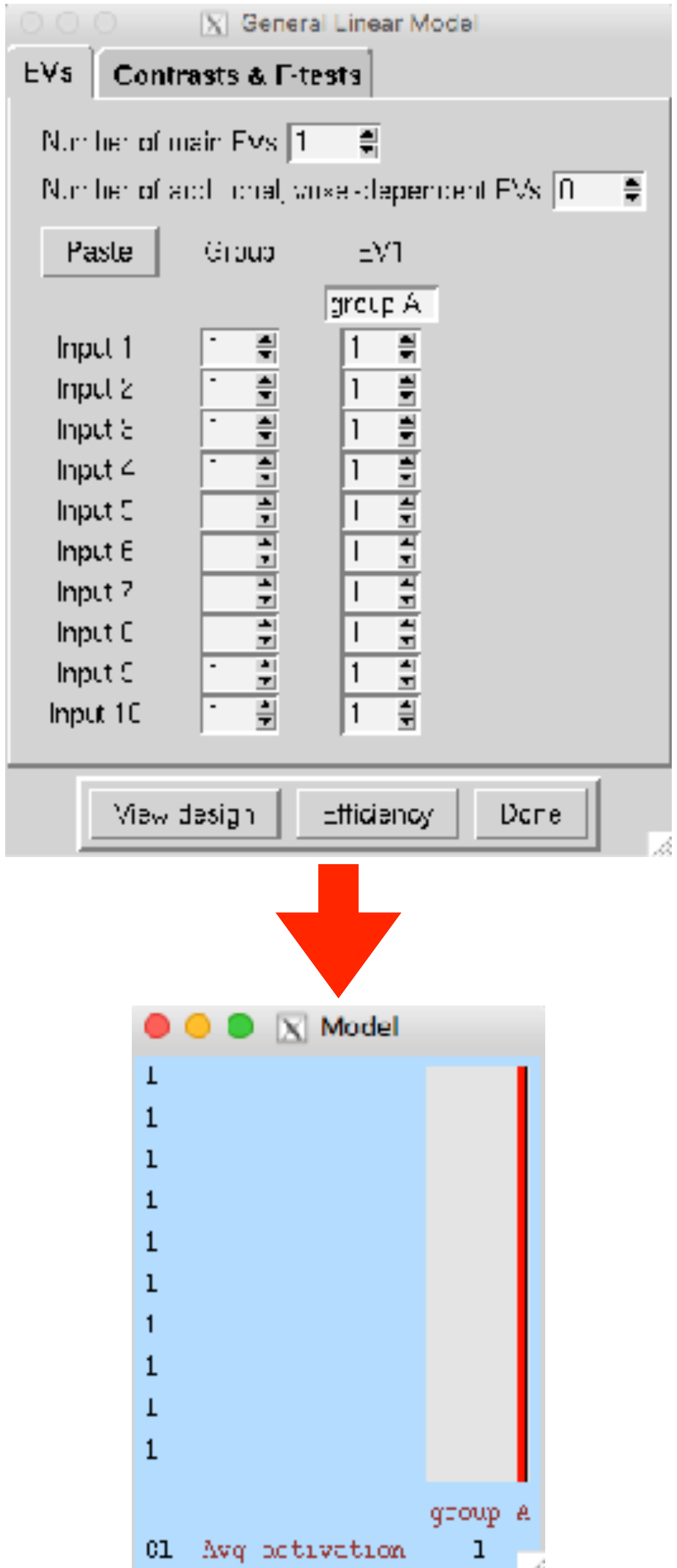


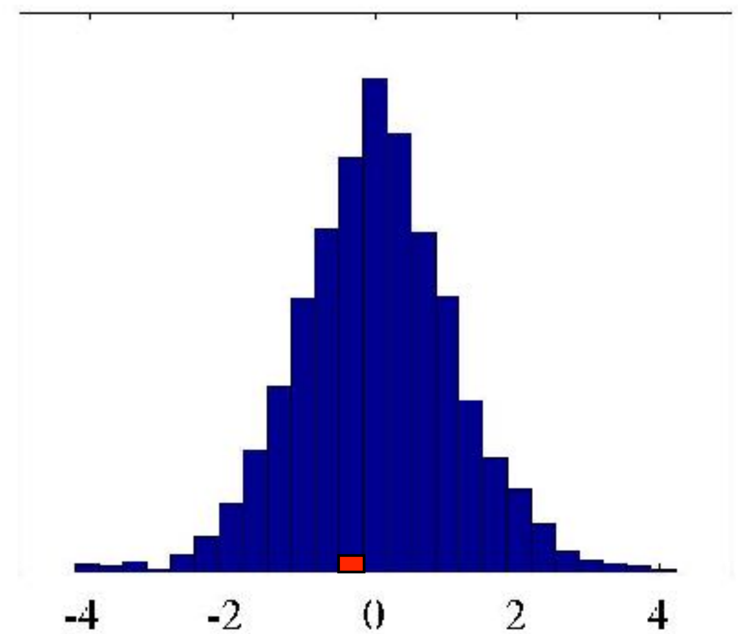
Etc …





Examples of exchangeability: Single group average

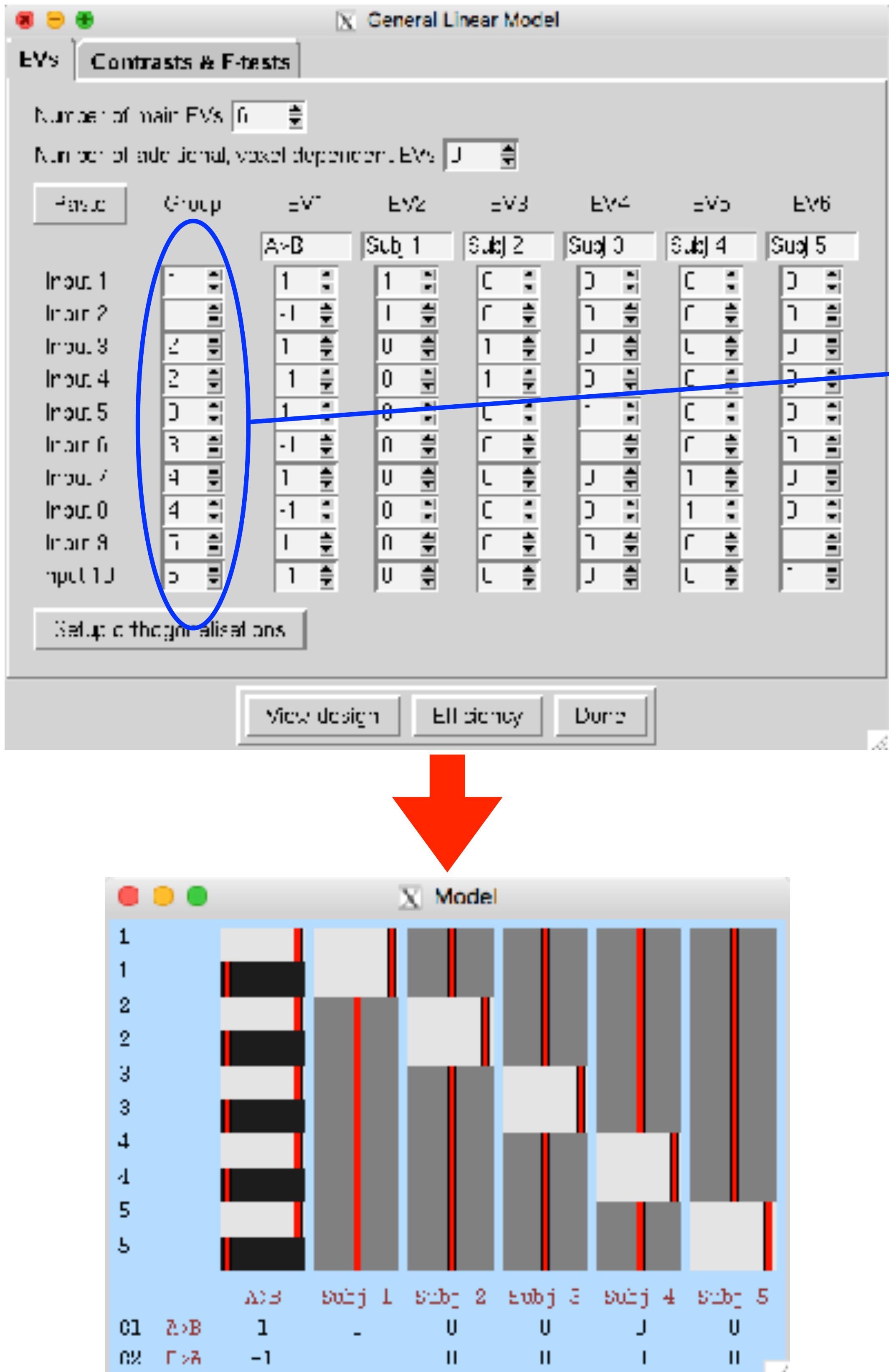




And the assumptions are:

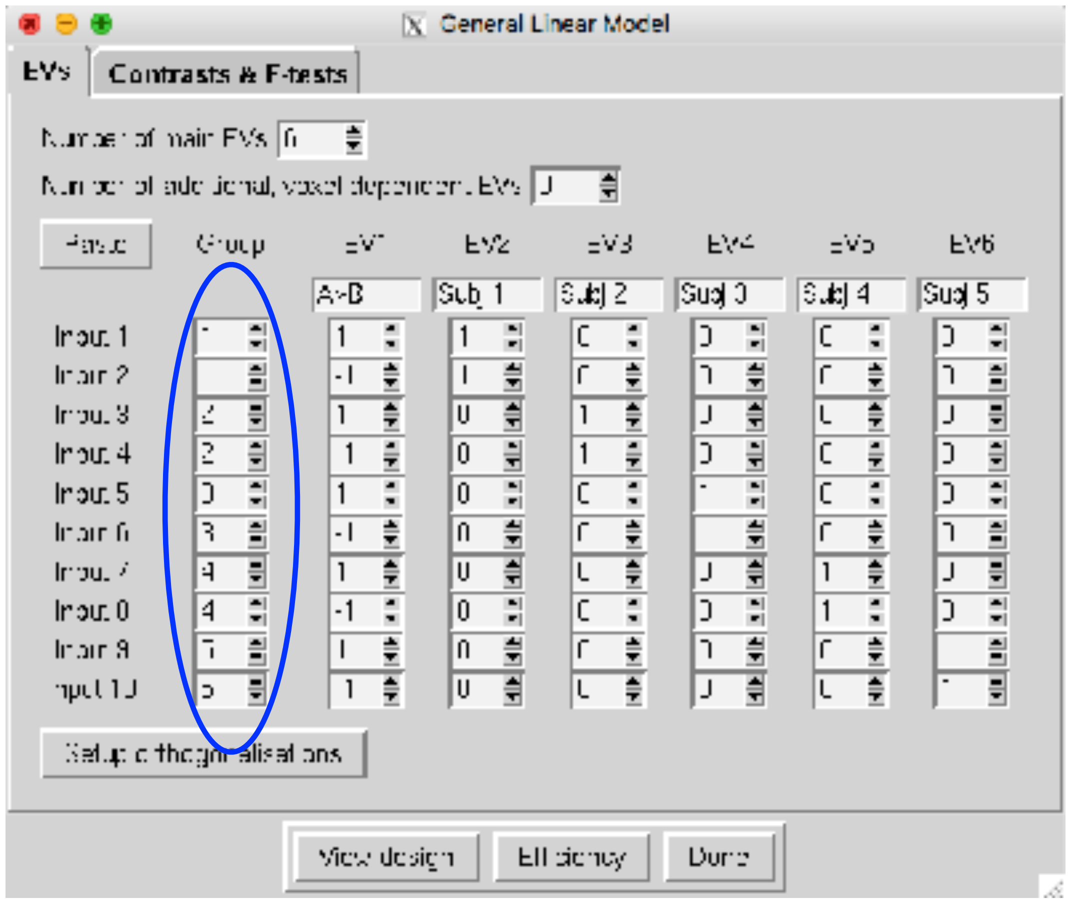
- Symmetric errors
- Errors independent
- Subjects drawn from a single population

Examples of exchangeability: Two groups paired

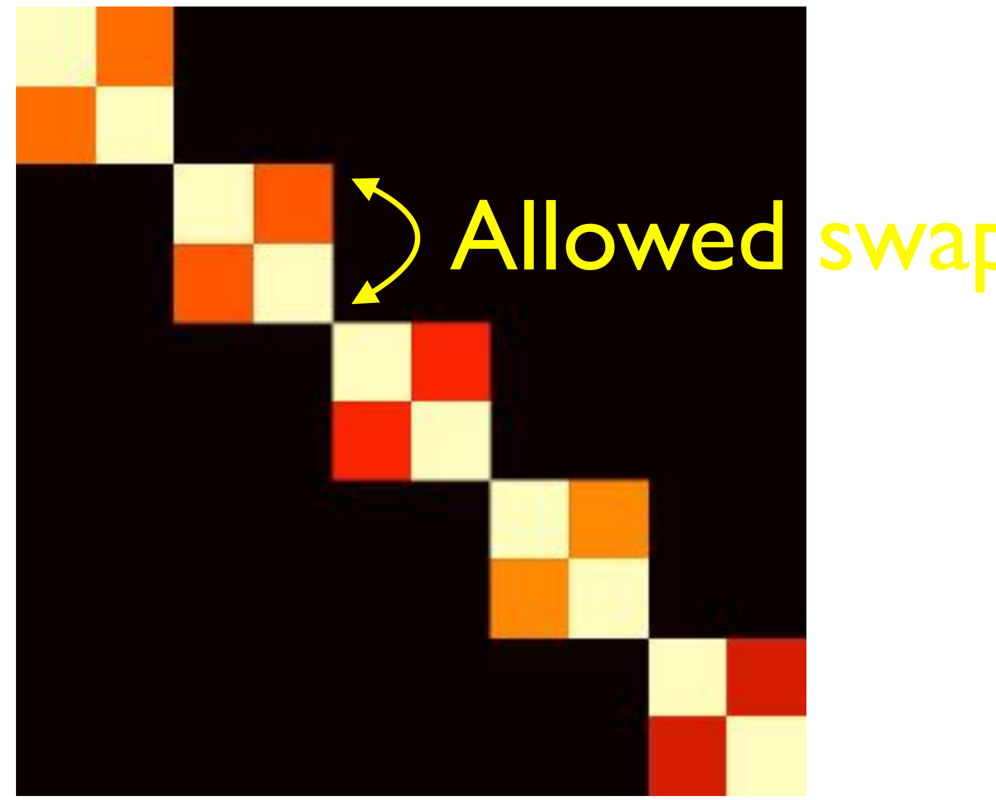


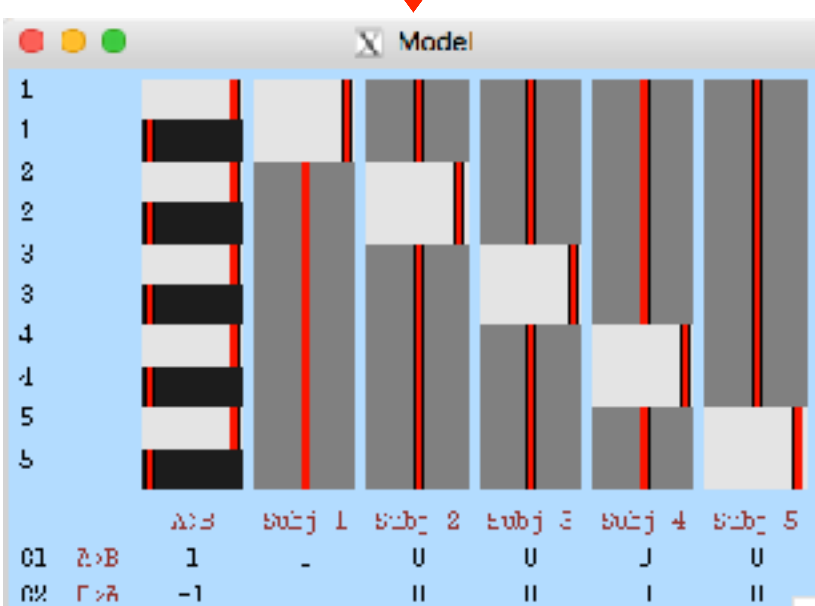
Here we can only exchange scans within each subject. I.e. Input 1 for Input 2, Input 3 for Input 4 etc

Examples of exchangeability: Two groups paired



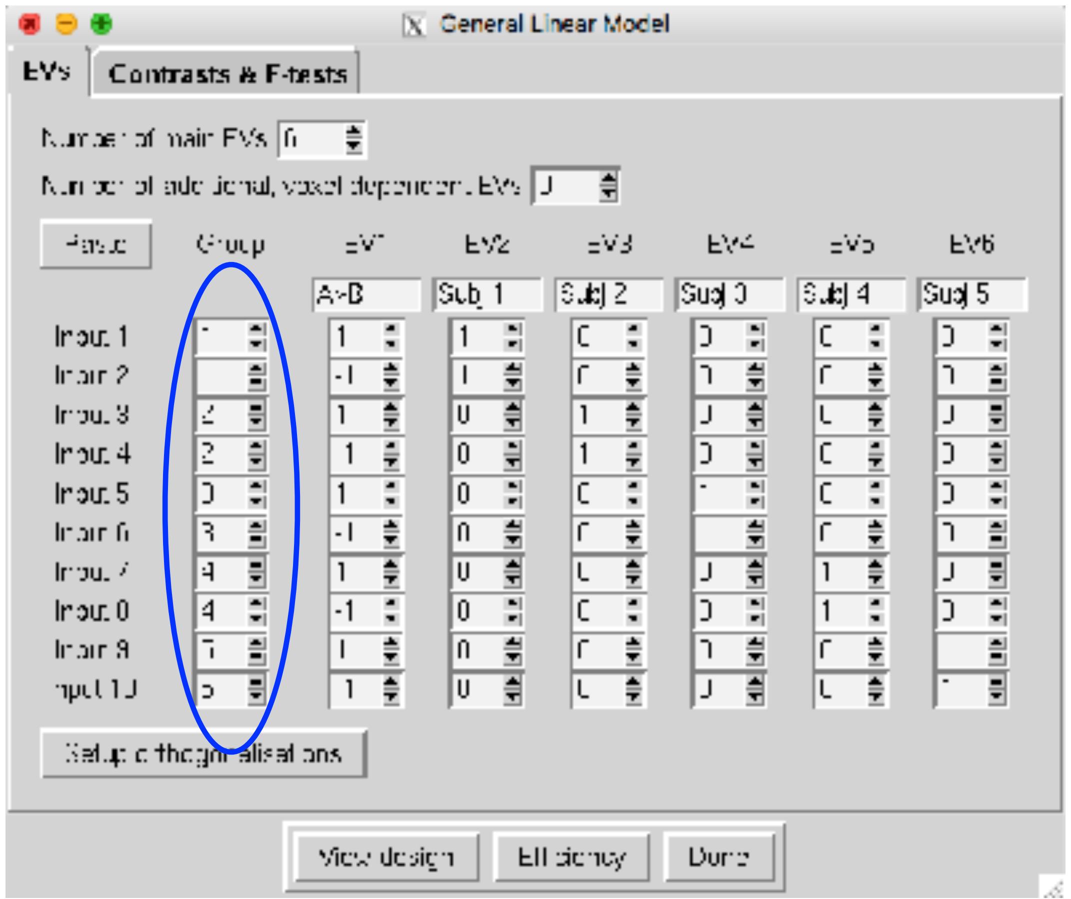
Assumed covariance matrix

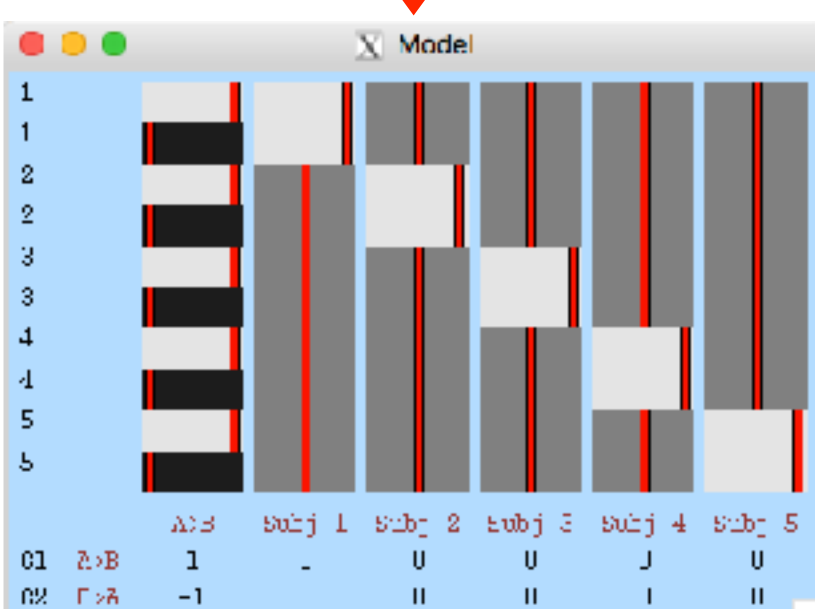




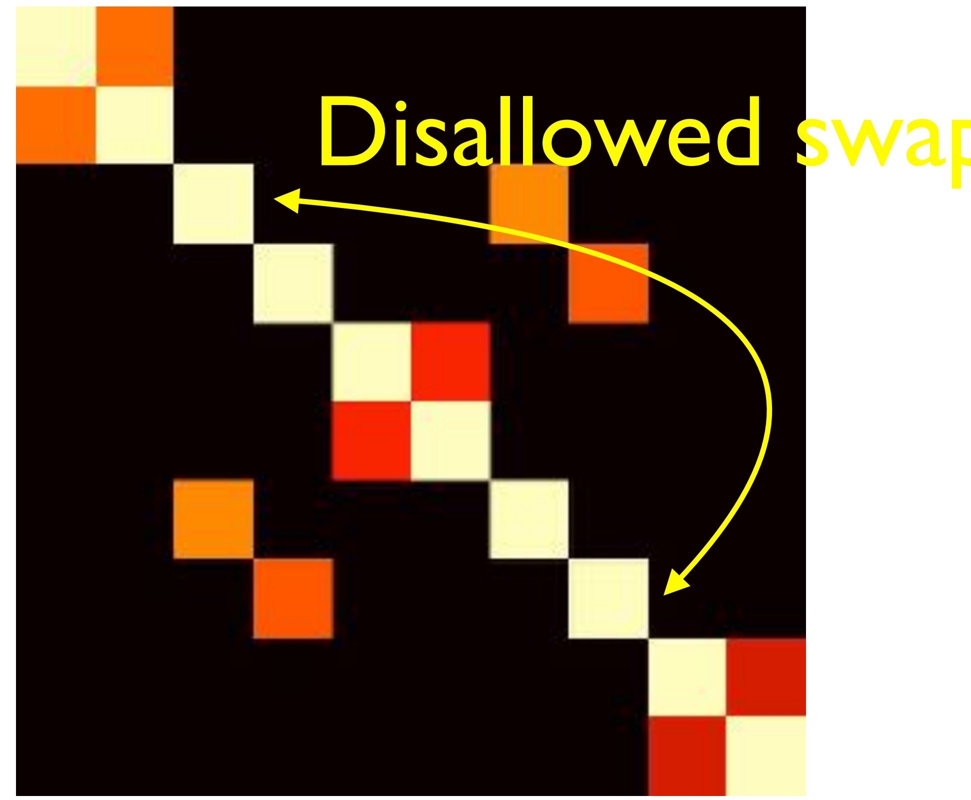
The implicit assumption here is that data from all subjects have the same uncertainty and that there is no dependence between subjects

Examples of exchangeability: Two groups paired



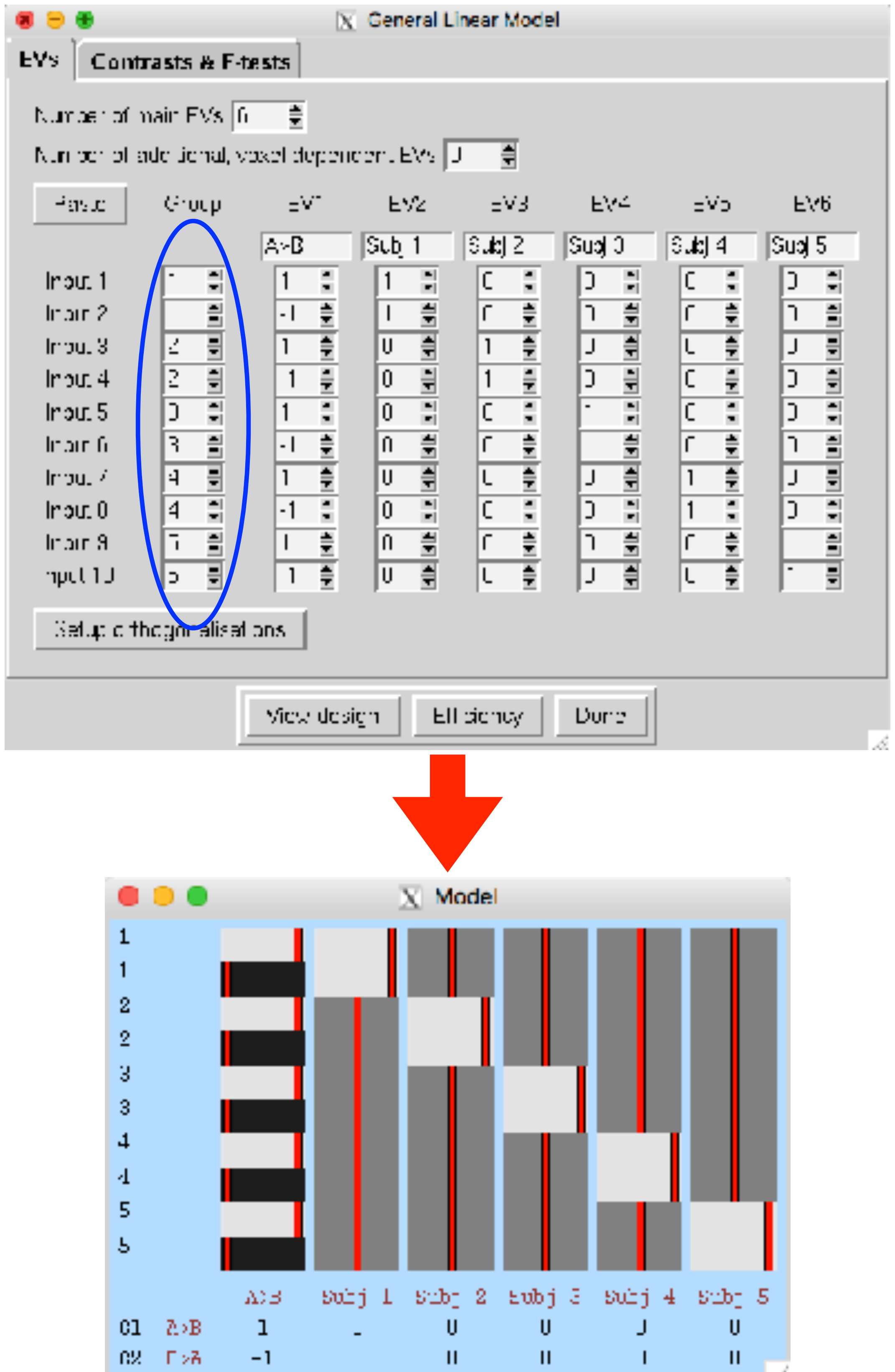


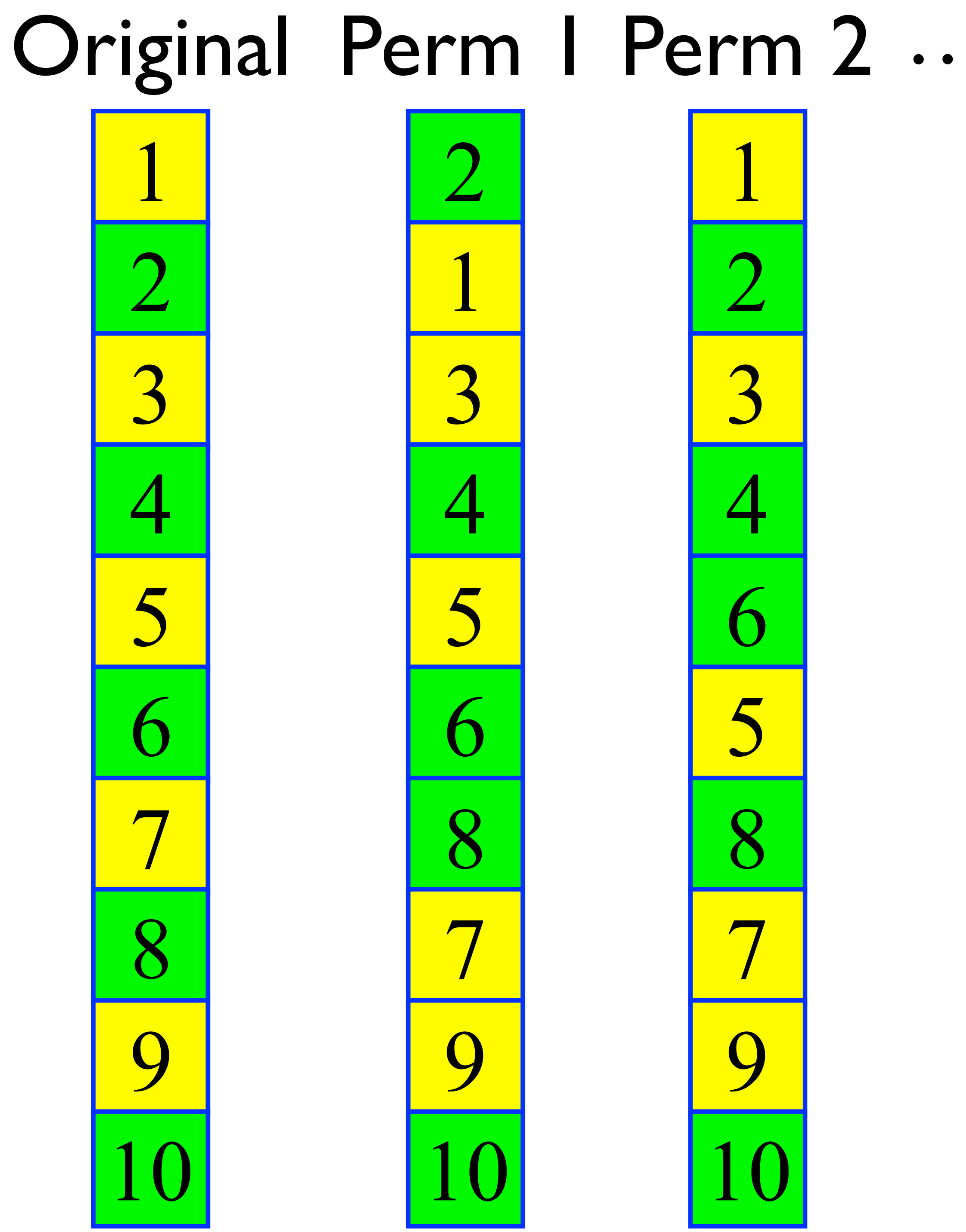
Assumed covariance matrix



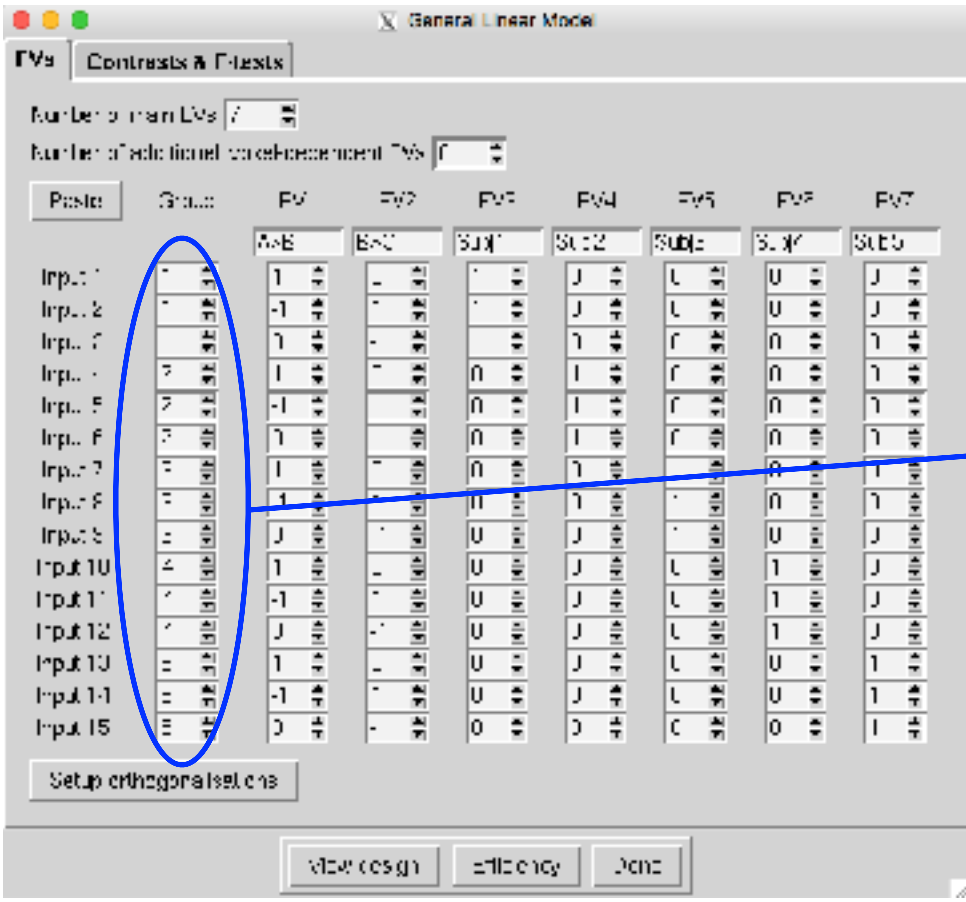
The implicit assumption here is that data from all subjects have the same uncertainty and that there is no dependence between subjects

Examples of exchangeability: Two groups paired

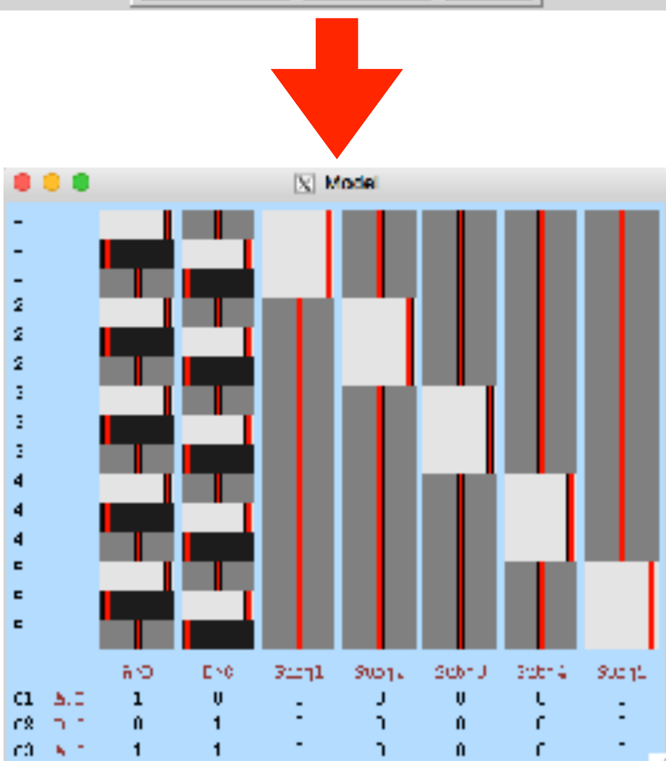




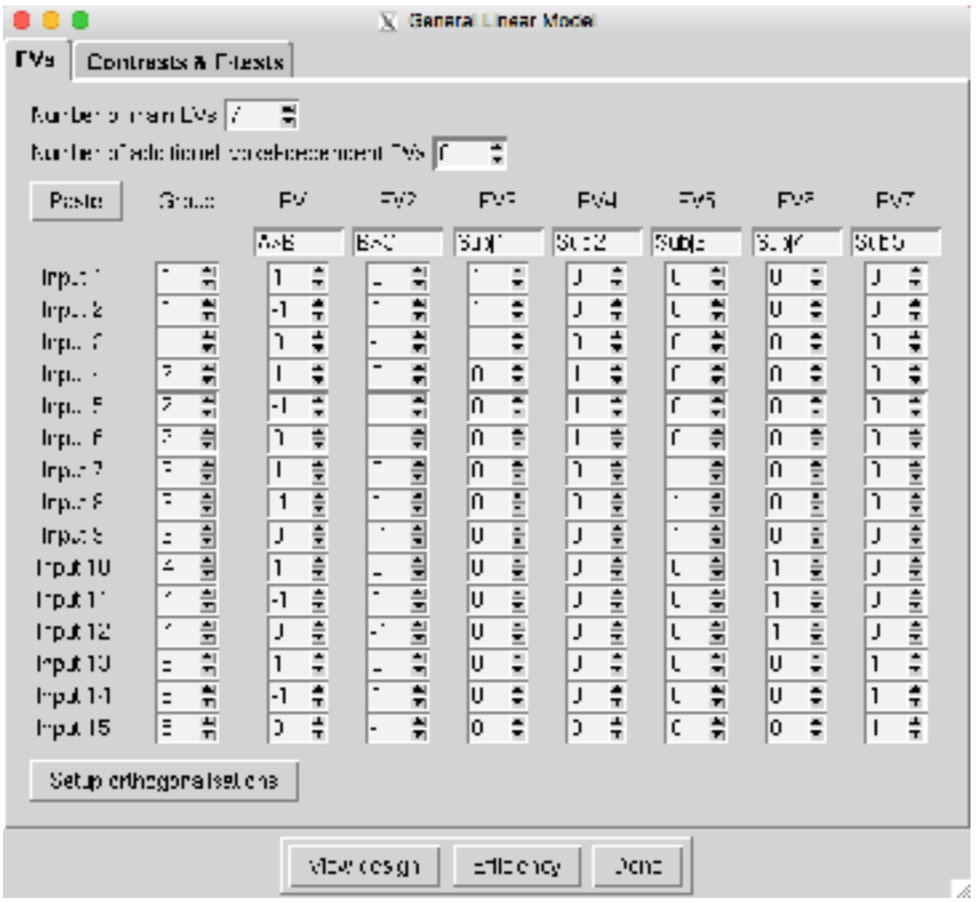
Examples of exchangeability: blocked ANOVA

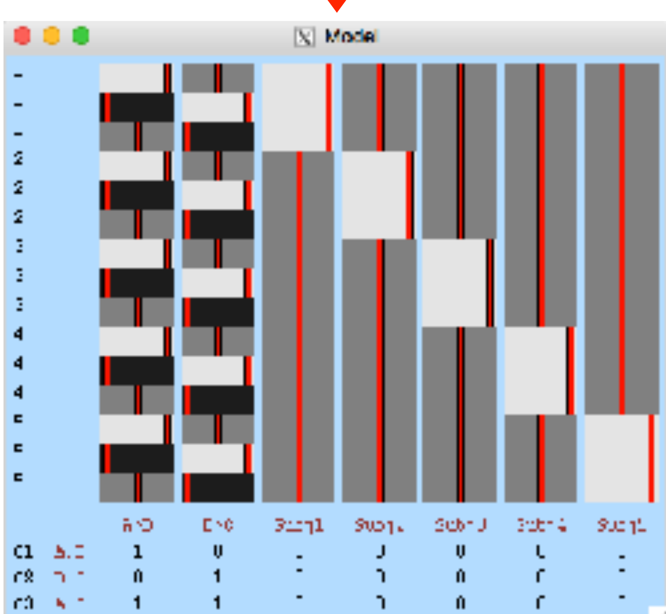


Same as previous: We can only swap labels within each subject

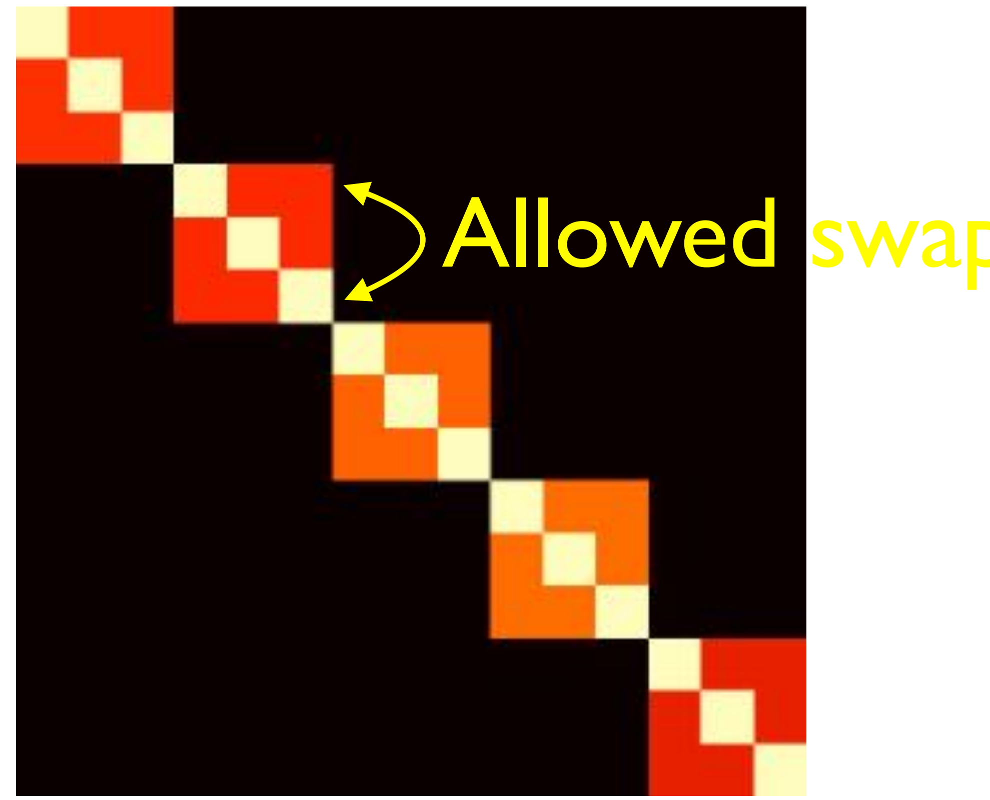


Examples of exchangeability: blocked ANOVA



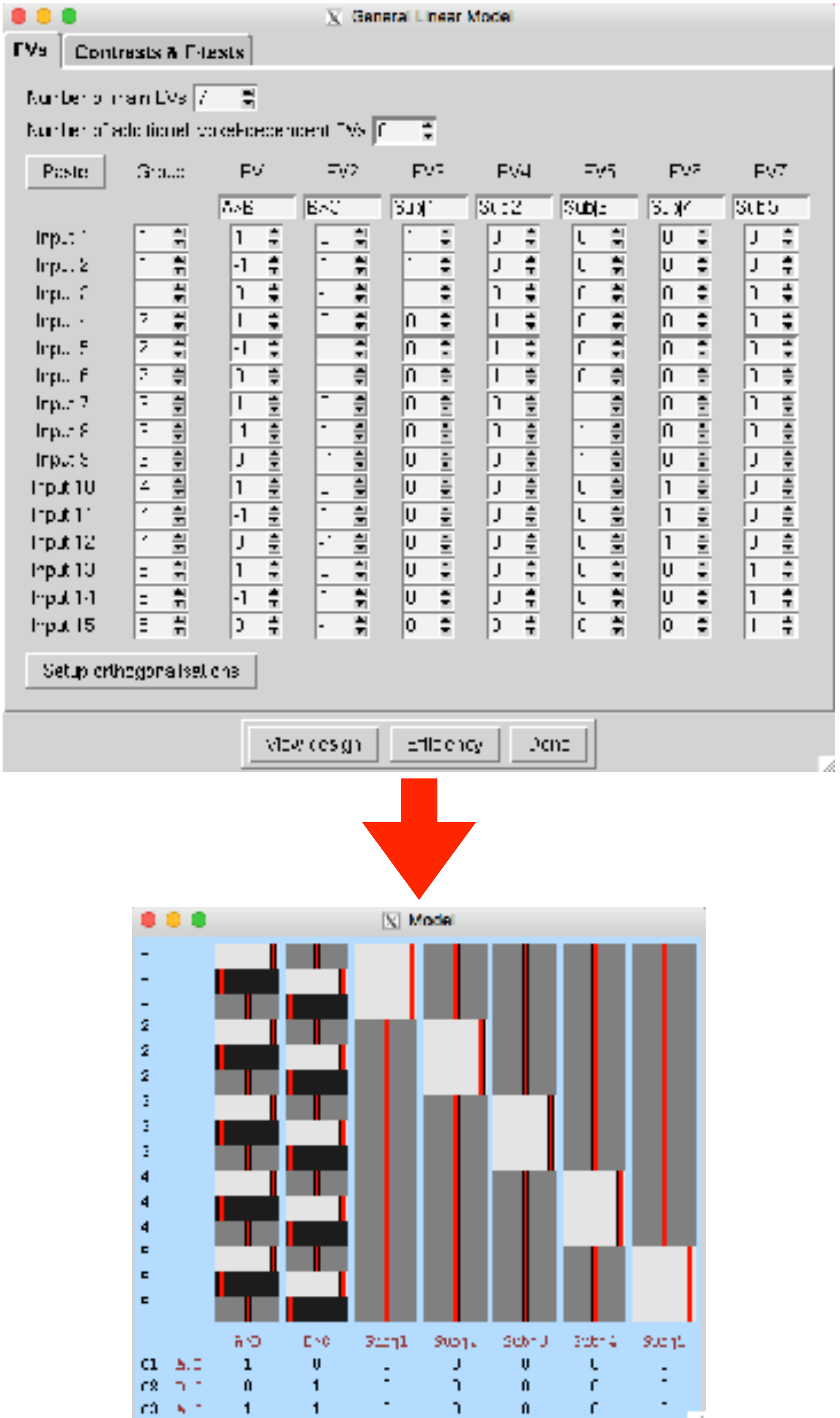


Assumed covariance matrix

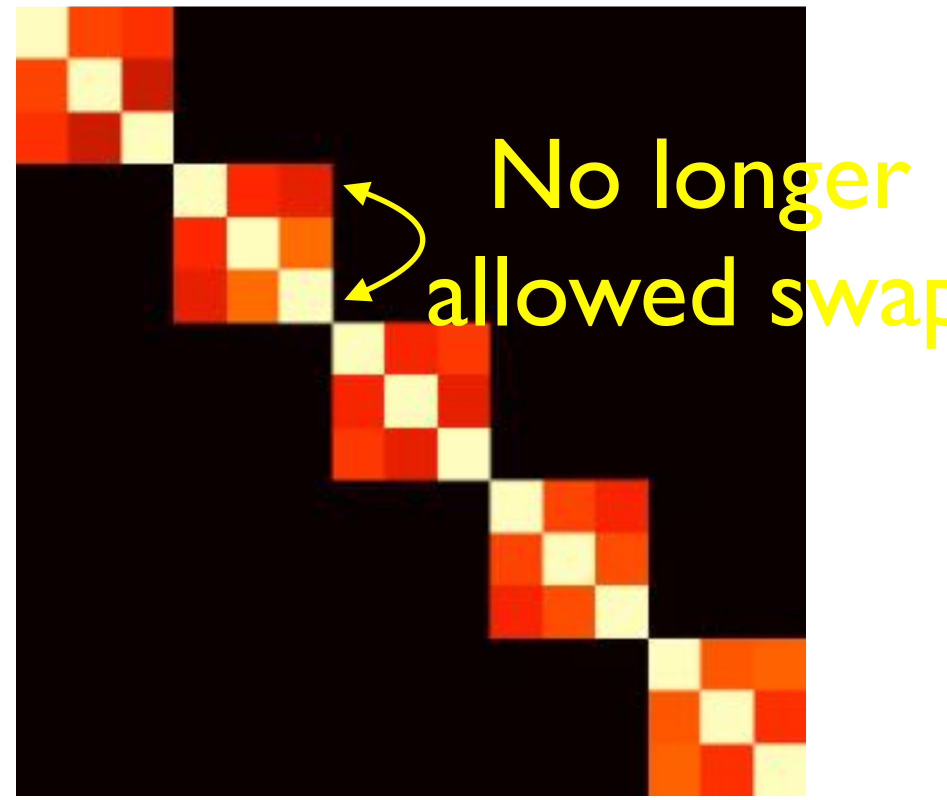


Assumptions: All subjects from the same "population", no dependence between subjects and "compound symmetry" within subjects

Examples of exchangeability: blocked ANOVA

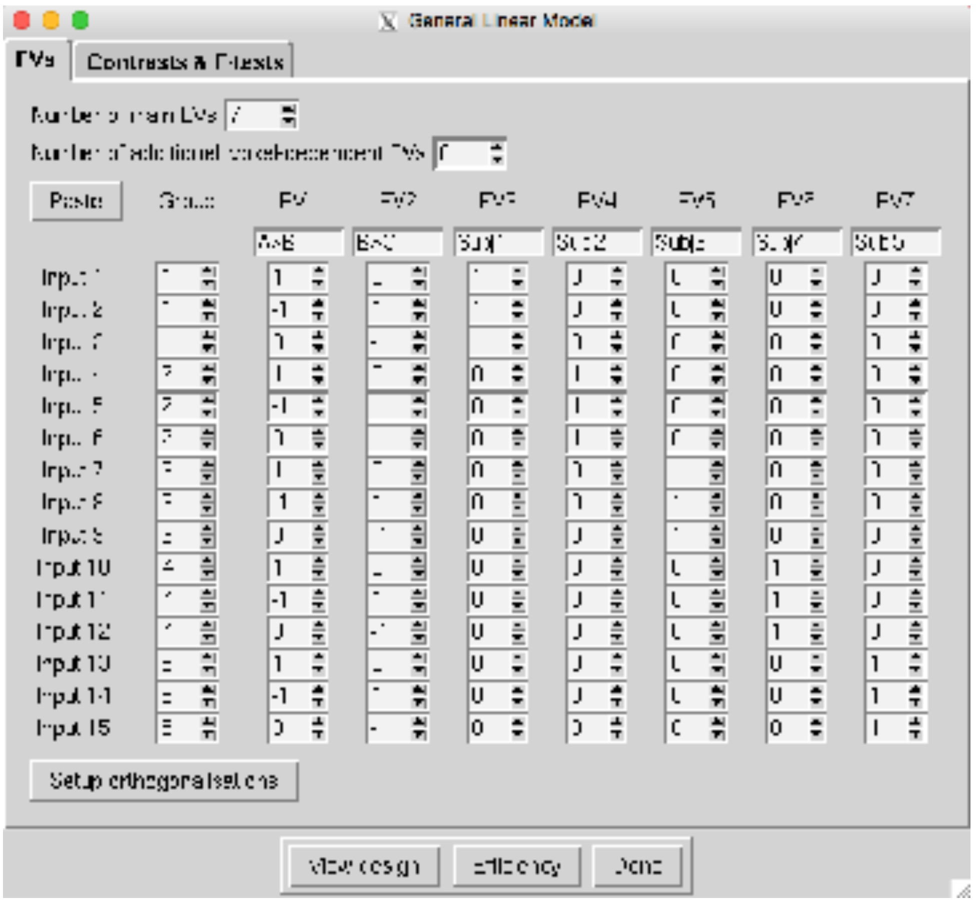


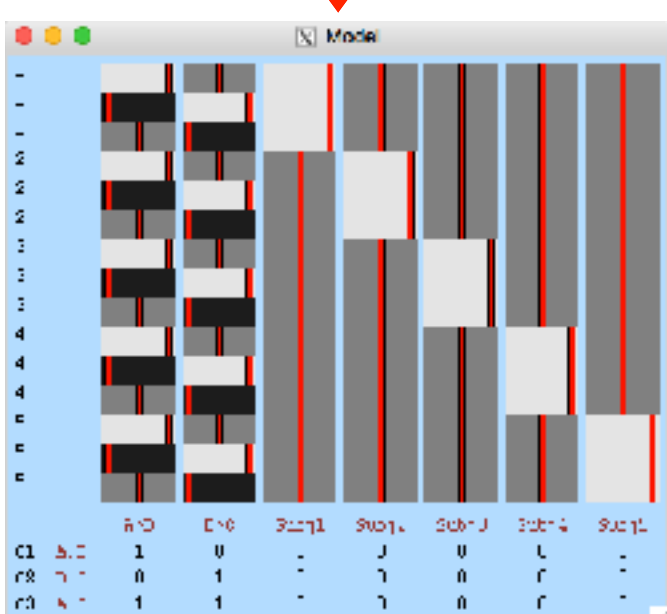
Assumed covariance matrix



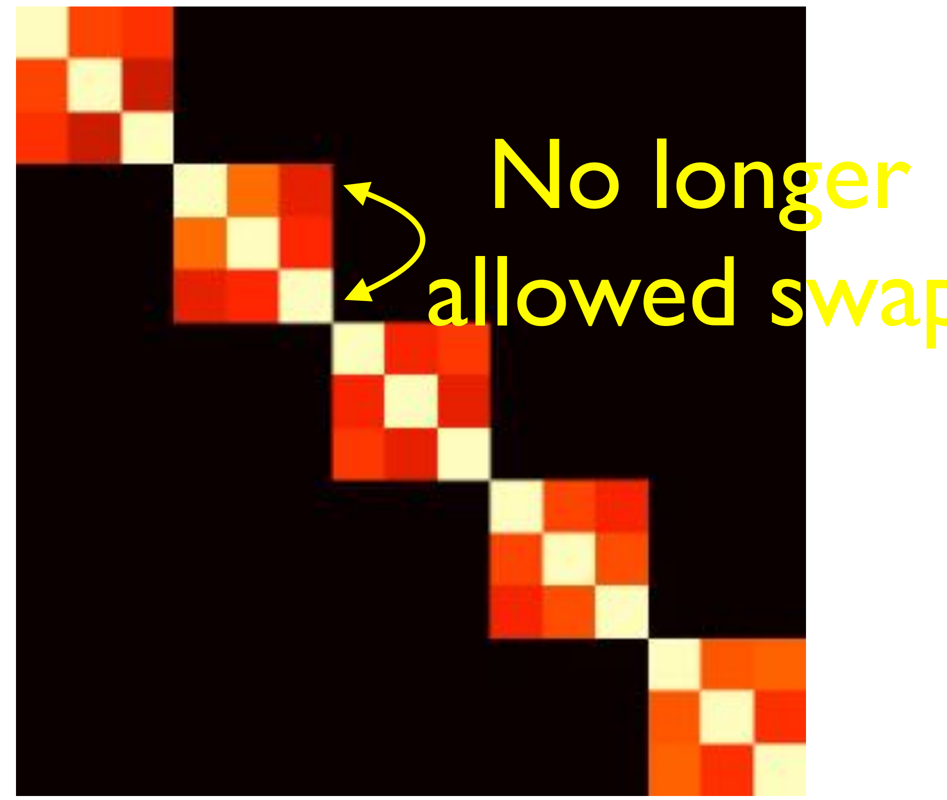
Assumptions: All subjects from the same "population", no dependence between subjects and "compound symmetry" within subjects

Examples of exchangeability: blocked ANOVA





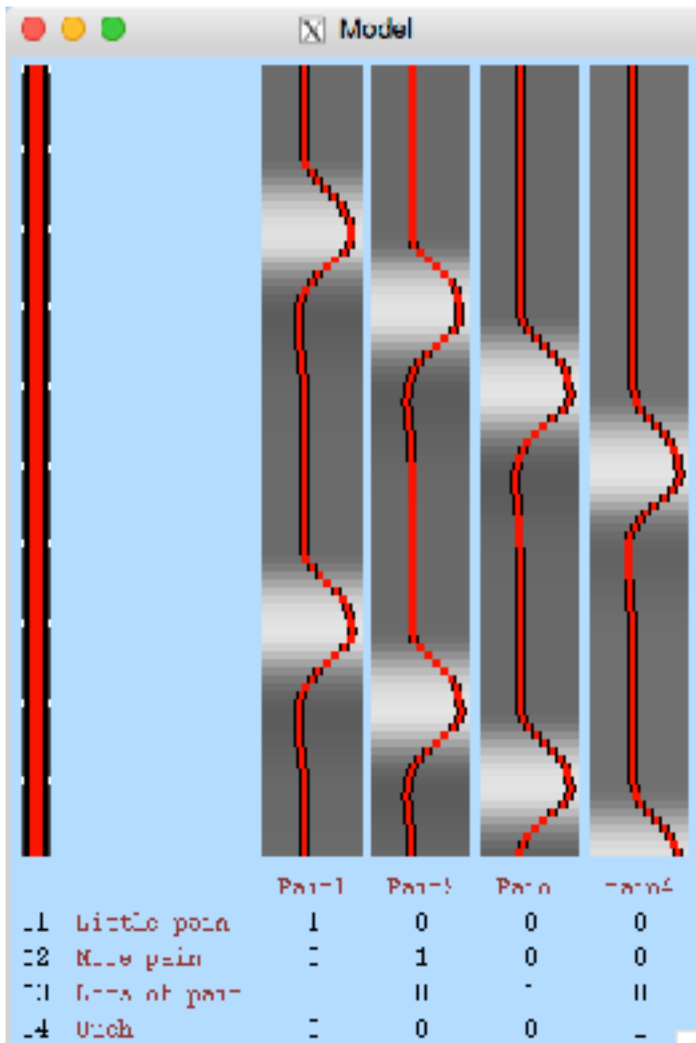
Assumed covariance matrix



Assumptions: All subjects from the same "population", no dependence between subjects and "compound symmetry" within subjects

My advice: Keep it simple!

Each subject scanned like this

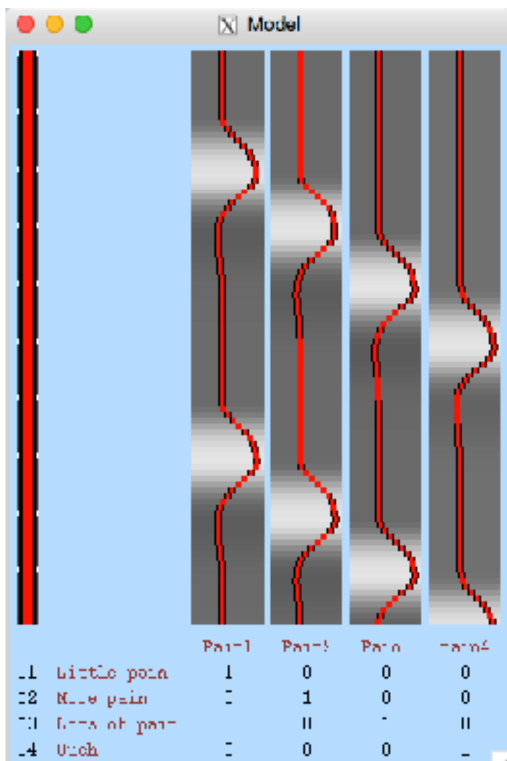


Taking 4 contrasts to 2nd level

We want to find areas that respond “linearly” to pain.

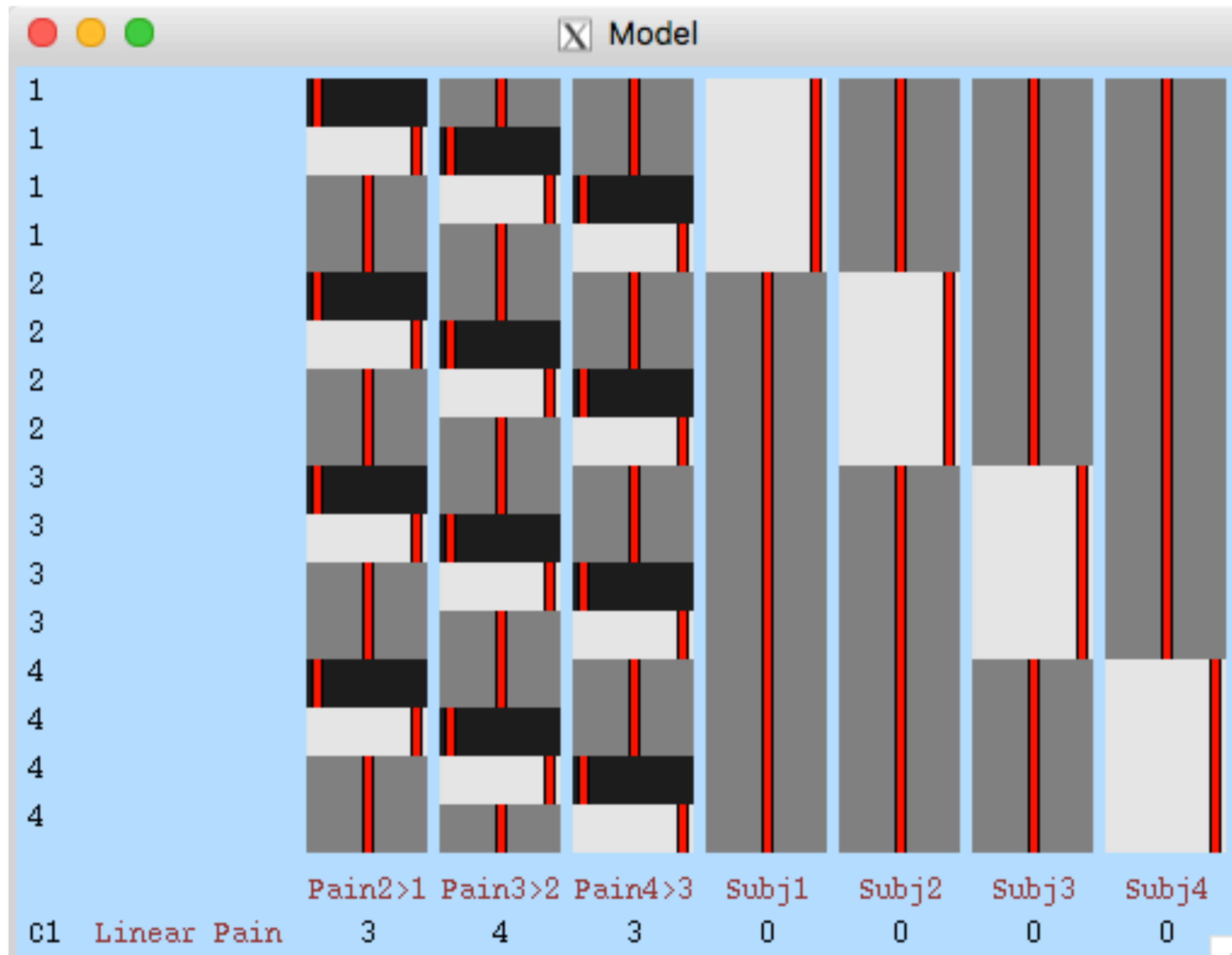
My advice: Keep it simple!

Each subject scanned like this

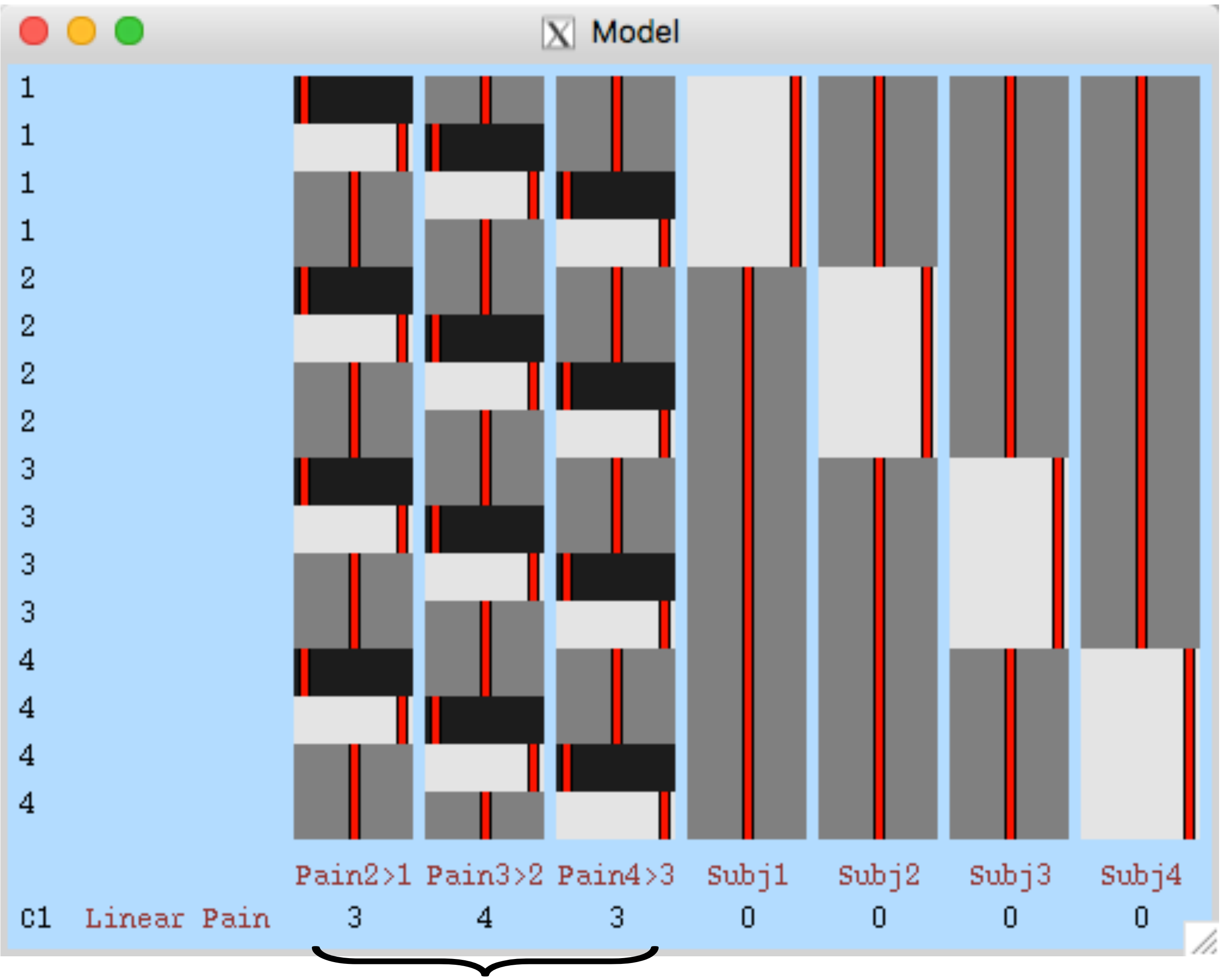


Taking 4 contrasts to 2nd level

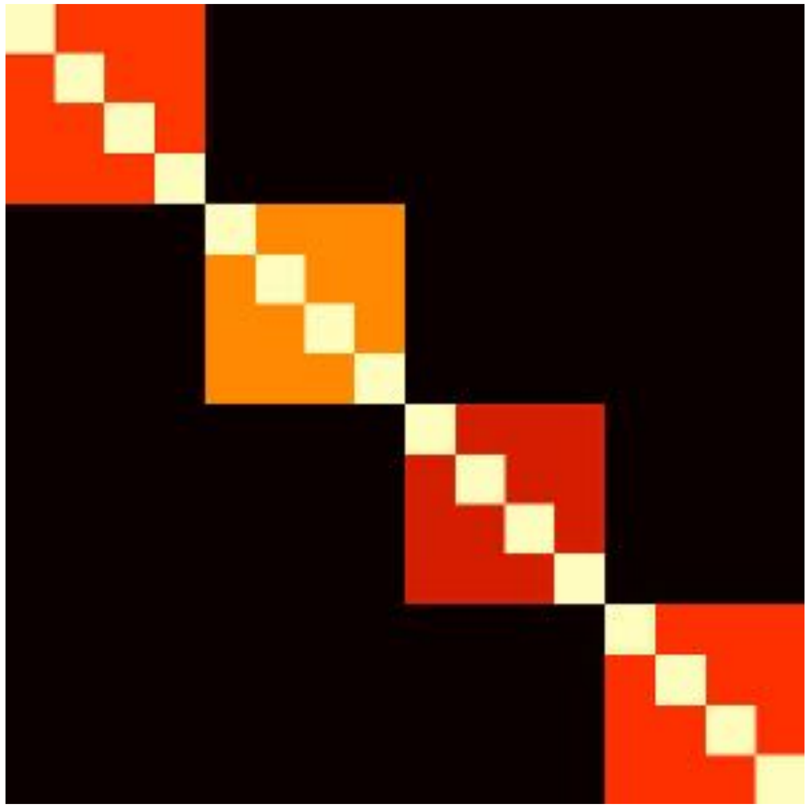
Repeating this for four subjects



My advice: Keep it simple!



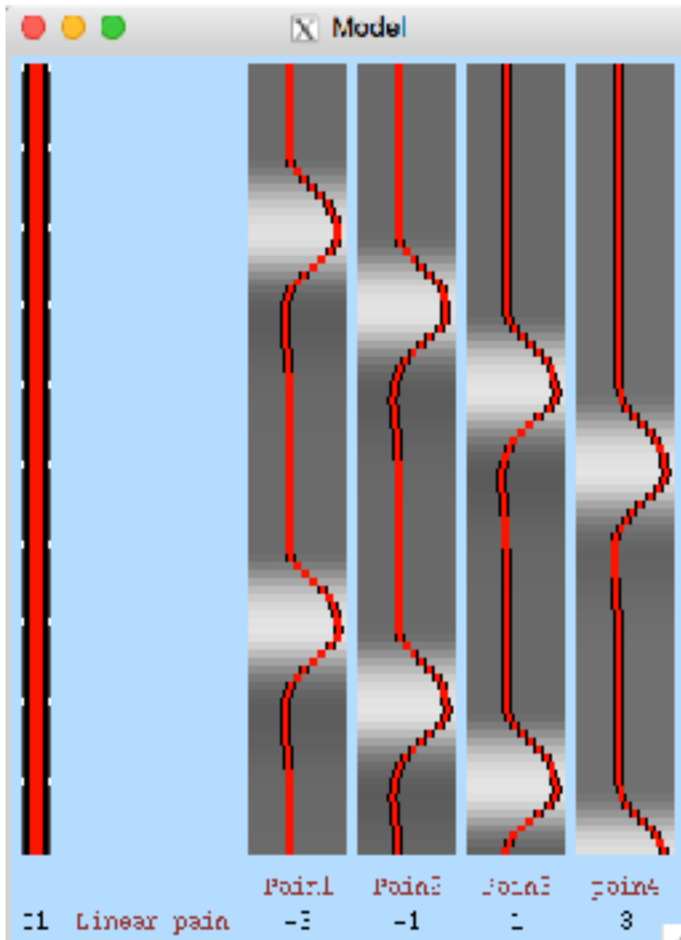
And figure out this contrast



You have to assume this covariance matrix

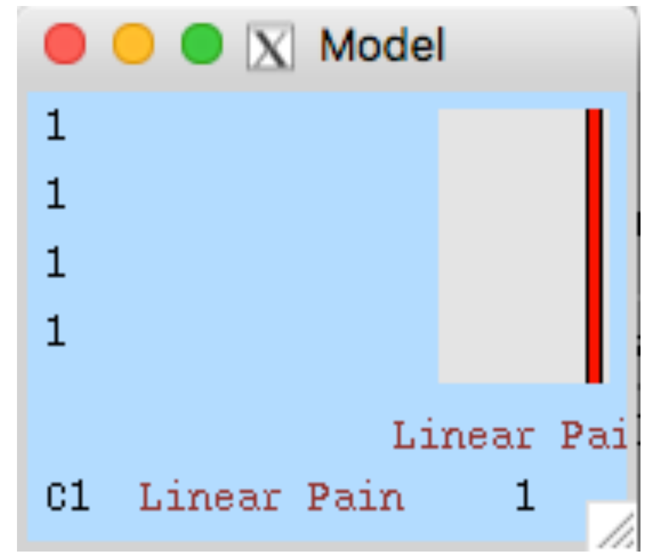
Why put yourself through all that pain?

My advice: Keep it simple!



When you can take a single contrast from the first level

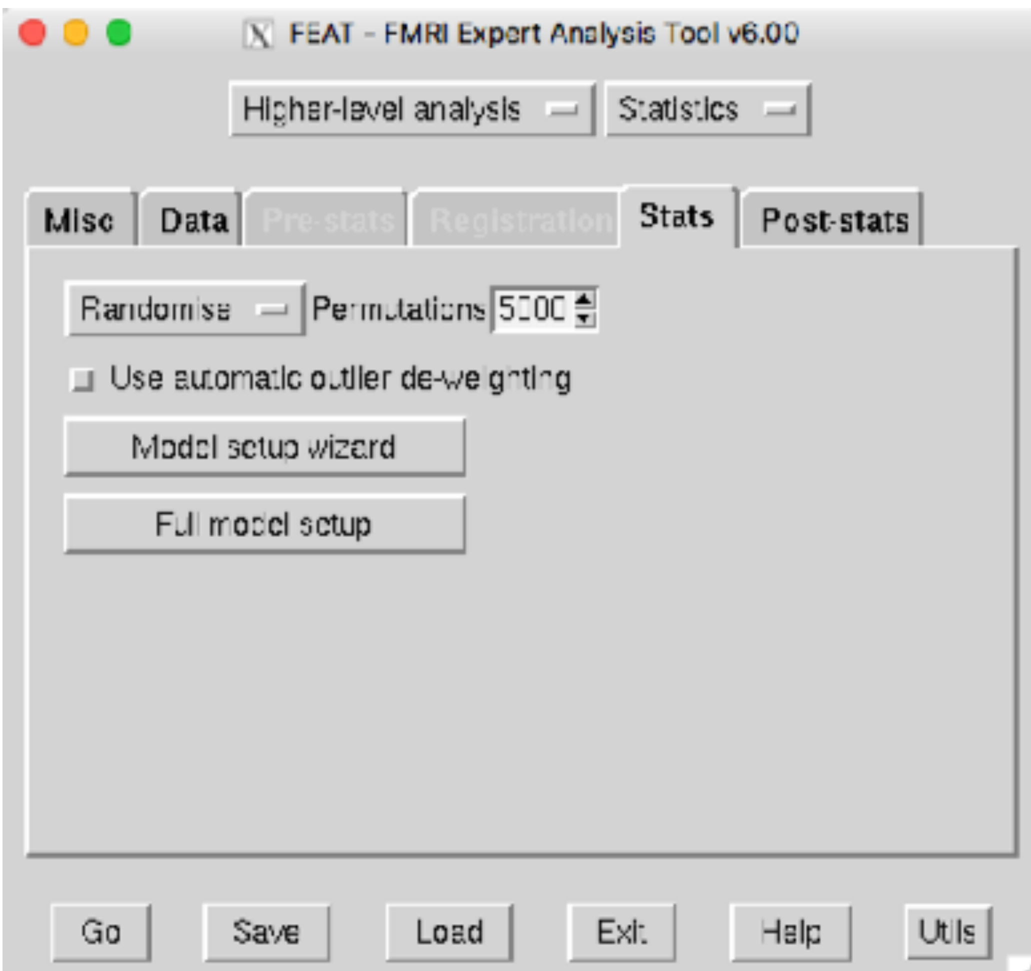
And get this at the second level

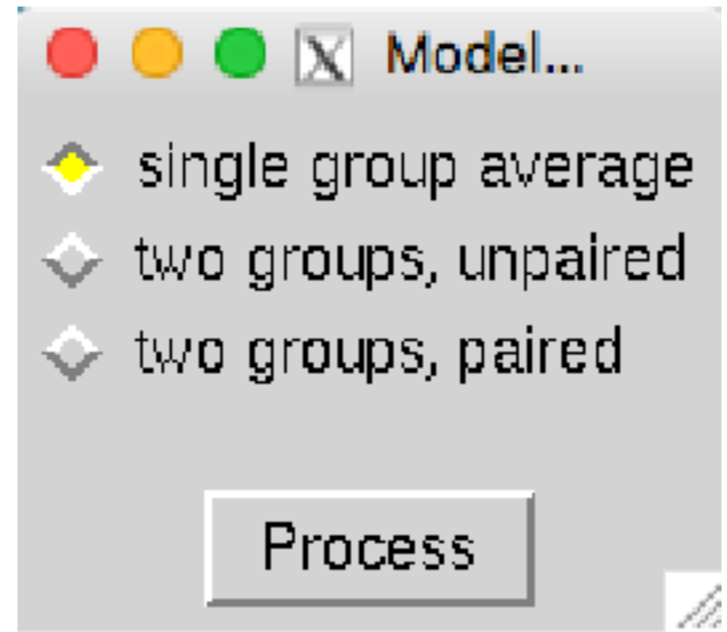


Assuming only symmetric errors

Much nicer, no?

Warning pertaining to FSL 6.0.1





Do not use the Model setup wizard together with Randomise in FSL 6.0.1