Inference 推断

how surprising is your statistic? (thresholding)

你的统计有多惊喜？（阈限）

中文翻译：王继源 孔亚卓

Outline大纲

- Null-hypothesis and Null-distribution 零假设和零分布
- Multiple comparisons and Family-wise error 多重比较和族错误率
- Different ways of being surprised 惊奇的不同方式
  - Voxel-wise inference (Maximum z) 体素推断（最大z）
  - Cluster-wise inference (Maximum size) 簇水平推断（最大尺寸）
- Parametric vs non-parametric tests 参数vs非参检验
- Enhanced clusters 增强的簇
- FDR - False Discovery Rate FDR-错误发现率

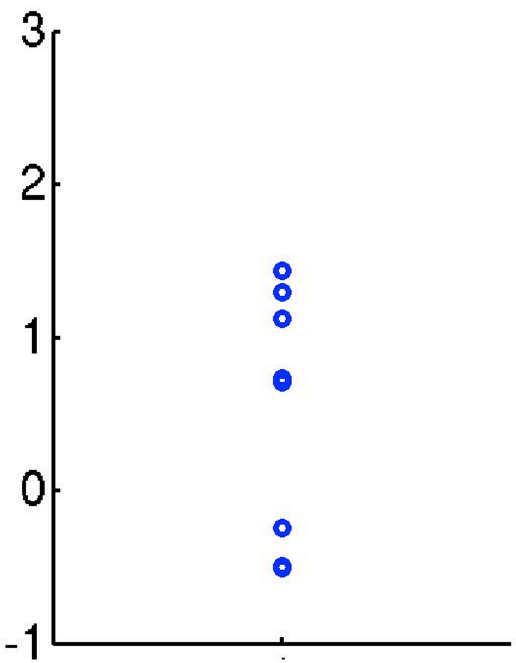
Outline大纲

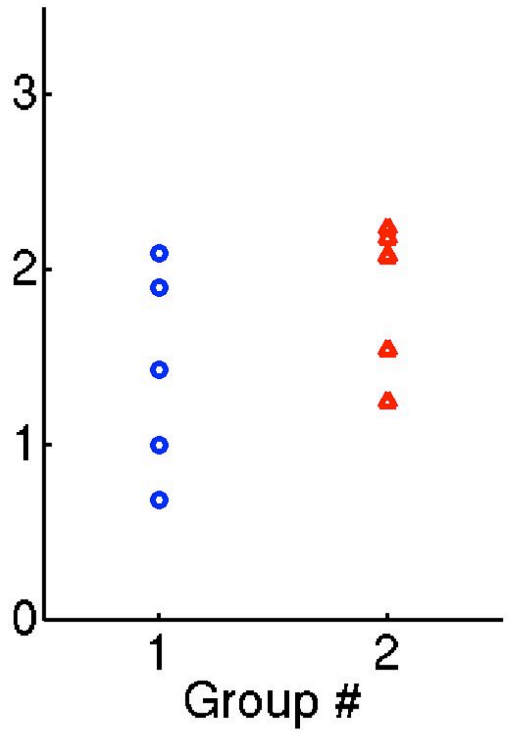
- Null-hypothesis and Null-distribution 零假设和零分布
- Multiple comparisons and Family-wise error 多重比较和族错误率
- Different ways of being surprised 惊奇的不同方式
  - Voxel-wise inference (Maximum z) 体素推断（最大z）
  - Cluster-wise inference (Maximum size) 簇水平推断（最大尺寸）
- Parametric vs non-parametric tests 参数vs非参检验
- Enhanced clusters 增强的簇
- FDR - False Discovery Rate FDR-错误发现率

The task of classical inference
经典推断的任务

- Given some data we want to know if (e.g.) a mean is different from zero or if two means are different

给一些数据我们想知道平均值是否不同于0，或者两个平均值是否不同





> 0 ?

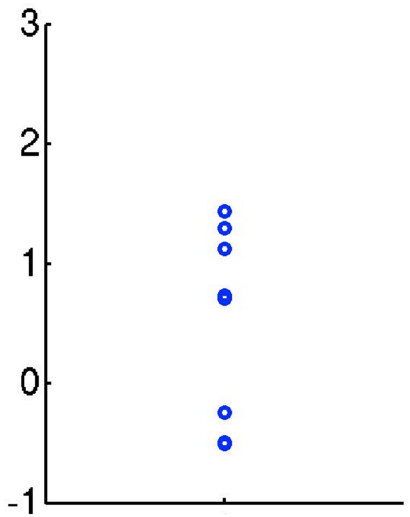
Different?不同?

Tools of classical inference 经典推断的工具

1. A null-hypothesis 零假设

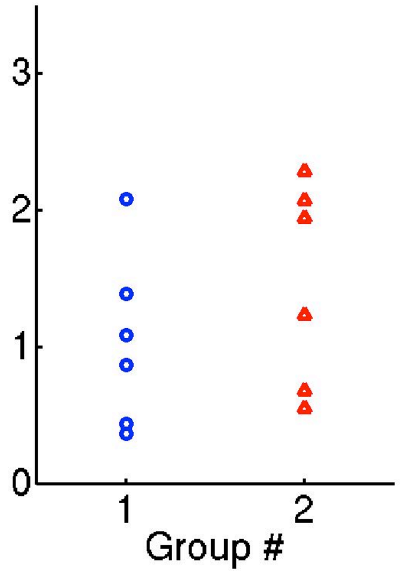
Typically the opposite of what we actually "hope", e.g.
通常是我们所"期望"的反面

There is **no** effect of treatment: $\mu = 0$
治疗无效$\mu = 0$



There is **no** difference between groups: $\mu_1 = \mu_2$

组之间无差$\mu 1 = \mu 2$

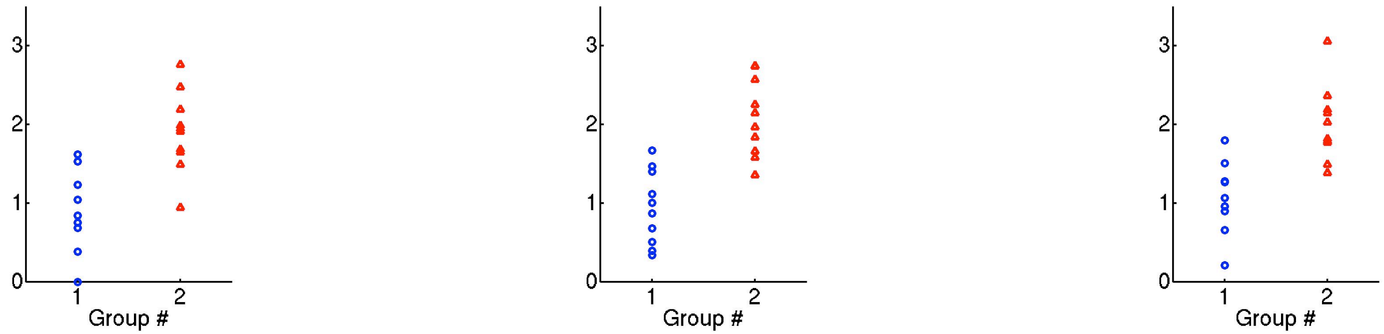


Tools of classical inference 经典推断的工具

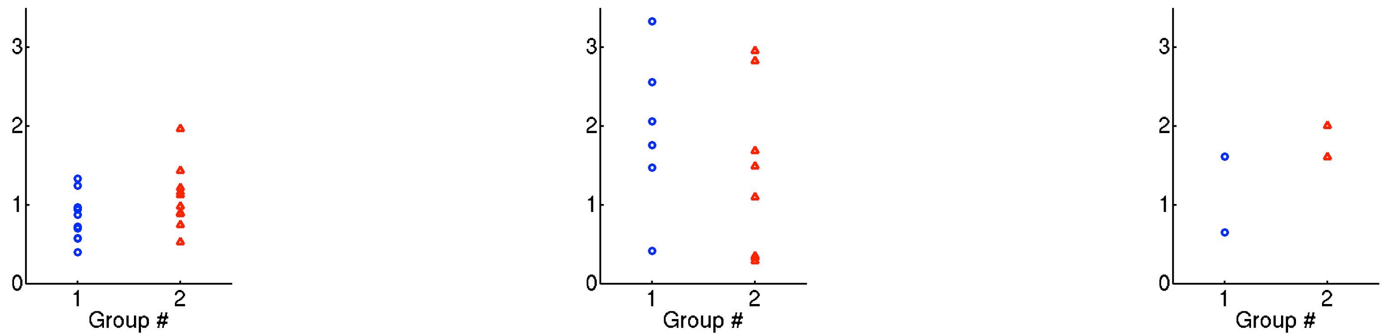
1. A null-hypothesis 零假设
2. A test-statistic 检验统计

Assesses “trustworthiness”
评估可信度

Trustworthy
可信的



Dodgy
不可靠的



Tools of classical inference 经典推断的工具

1. A null-hypothesis 零假设
2. A test-statistic 检验统计

Assesses "trustworthiness"评估可信度

A *t*-statistic reflects precisely this

t统计正好反映了这点

$$t = \sqrt{n}\frac{\overline{x_1} - \overline{x_2}}{\sqrt{\sigma^2}}$$

Large difference: Trustworthy

较大差异：可信

Many measurements: Trustworthy

多次测量：可信

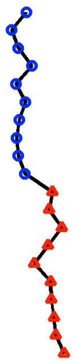
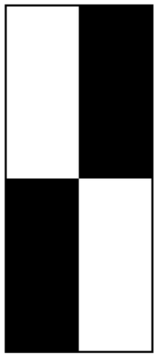
Small variability: Trustworthy

较小的变异：可信

Tools of classical inference 经典推断的工具

1. A null-hypothesis 零假设
2. A test-statistic 检验统计

Or expressed in GLM lingo
或者以GLM术语表示

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

Large difference: Trustworthy
较大差异：可信

$\bar{x}_1$ - $\bar{x}_2$

$$t = \frac{\mathbf{c}^T \widehat{\boldsymbol{\beta}}}{\sqrt{\sigma^2}\sqrt{\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}}}$$

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix}$$

Small variability: Trustworthy
小变异：可信

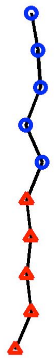
Many measurements: Trustworthy
很多测量：可信

Tools of classical inference 经典推断的工具

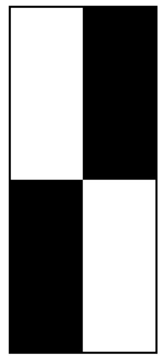
1. A null-hypothesis 零假设
2. A test-statistic 检验统计
3. A null-distribution 零分布

Let us assume there is no difference, i.e. the null-hypothesis is true.
假设没区别，零假设正确

We might then get these data
我们可能得到这些数据



$=$



$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \mathbf{e}$

$t = 2.19$

$\mathbf{c}^T\widehat{\boldsymbol{\beta}} = 1.17$

$$t = \frac{\mathbf{c}^T\widehat{\boldsymbol{\beta}}}{\sqrt{\sigma^2}\sqrt{\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}}}$$

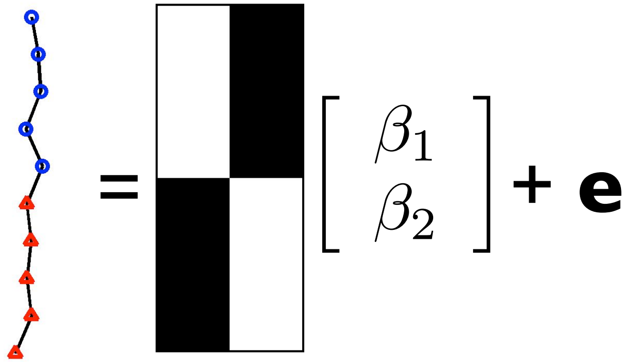
$\sigma^2 = 0.71$

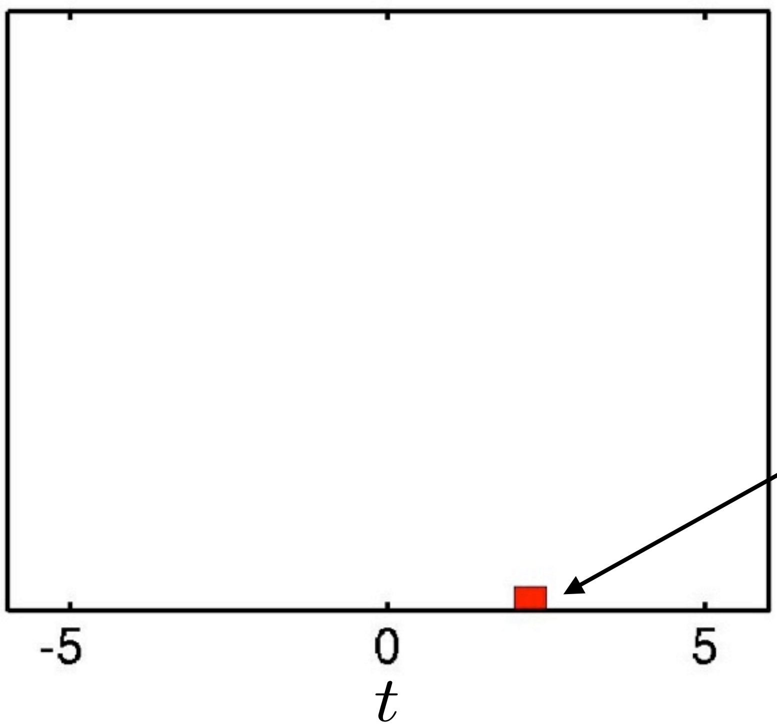
Constant 常量

Tools of classical inference 经典推断的工具

1. A null-hypothesis 零假设
2. A test-statistic 检验统计
3. A null-distribution 零分布

We might then get these data
我们可能得到这些数据





$t = 2.19$

$\mathbf{c}^T\widehat{\boldsymbol{\beta}} = 1.17$

$$t = \frac{\mathbf{c}^T\widehat{\boldsymbol{\beta}}}{\sqrt{\sigma^2}\sqrt{\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}}}$$

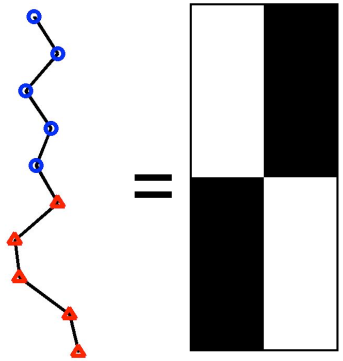
$\sigma^2 = 0.71$

Constant常量

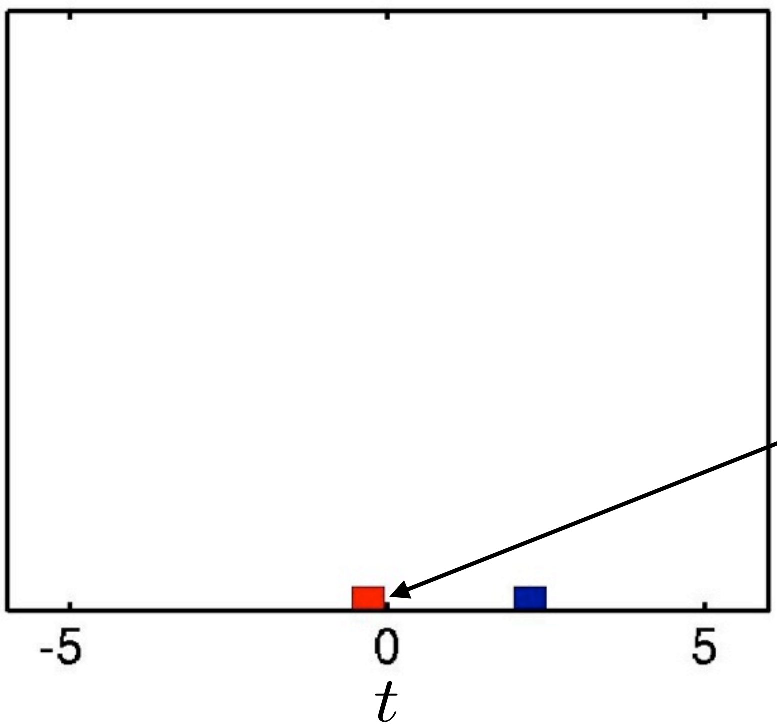
Tools of classical inference 经典推断的工具

1. A null-hypothesis 零假设
2. A test-statistic 检验统计
3. A null-distribution 零分布

or we could have gotten these
我们也可能获得这些数据



$$= \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \mathbf{e}$$



$$t = -0.51$$

$$\mathbf{c}^T\widehat{\boldsymbol{\beta}} = -0.37$$

$$t = \frac{\mathbf{c}^T\widehat{\boldsymbol{\beta}}}{\sqrt{\sigma^2}\sqrt{\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}}}$$

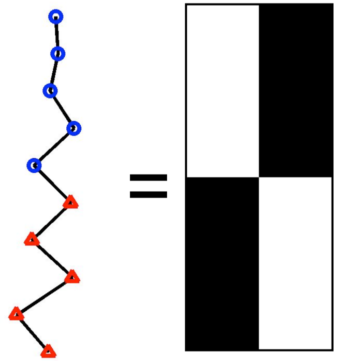
$$\sigma^2 = 1.28$$

Constant常量

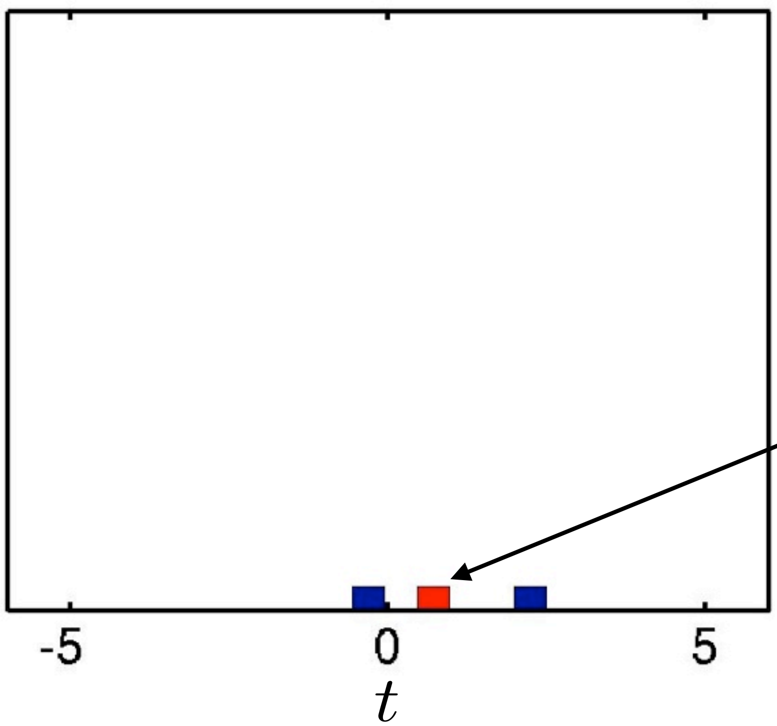
Tools of classical inference 经典推断的工具

1. A null-hypothesis 零假设
2. A test-statistic 检验统计
3. A null-distribution 零分布

maybe these或许这样



$$= \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \mathbf{e}$$



$$t = 0.49$$

$$\mathbf{c}^T\widehat{\boldsymbol{\beta}} = 0.31$$

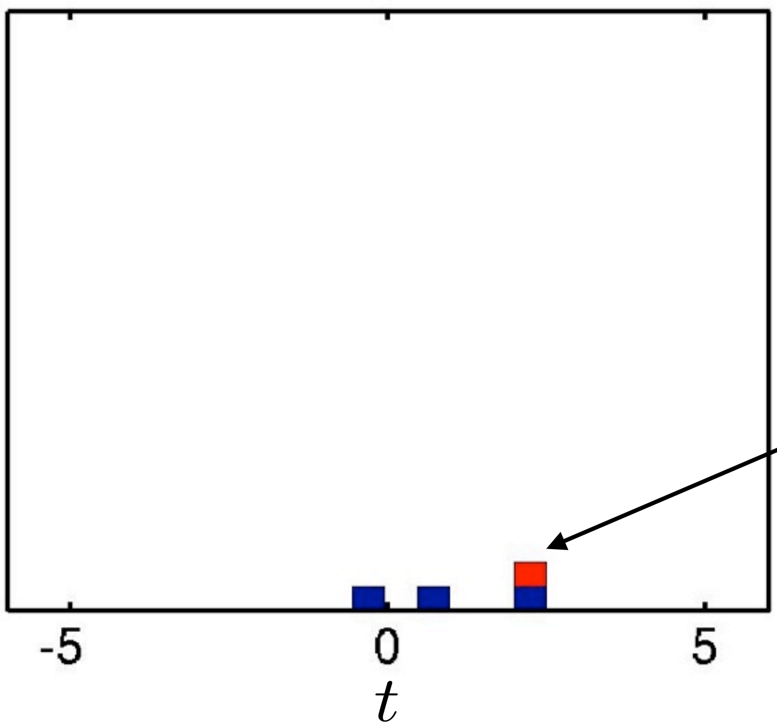
$$t = \frac{\mathbf{c}^T\widehat{\boldsymbol{\beta}}}{\sqrt{\sigma^2}\sqrt{\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}}}$$

$$\sigma^2 = 1.01$$

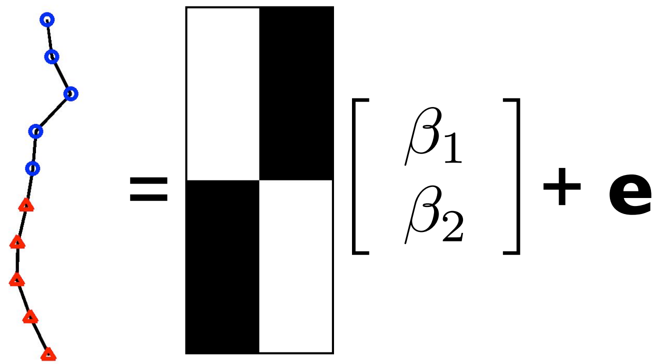
Constant常量

Tools of classical inference 经典推断的工具

1. A null-hypothesis 零假设
2. A test-statistic 检验统计
3. A null-distribution 零分布



or perhaps these
还有可能这样



$t = 2.19$

$\mathbf{c}^T\hat{\boldsymbol{\beta}} = 1.22$

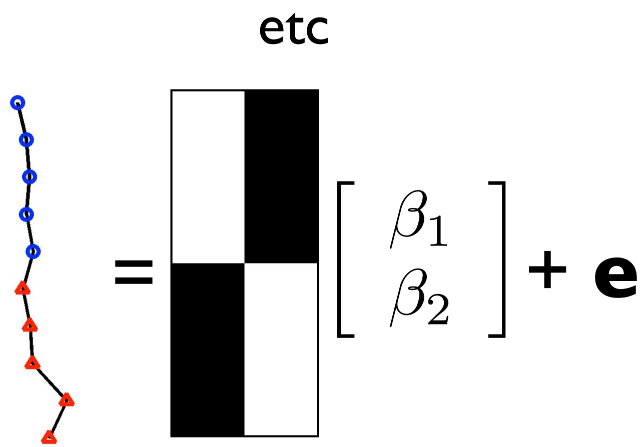
$$t = \frac{\mathbf{c}^T\hat{\boldsymbol{\beta}}}{\sqrt{\sigma^2}\sqrt{\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}}}$$

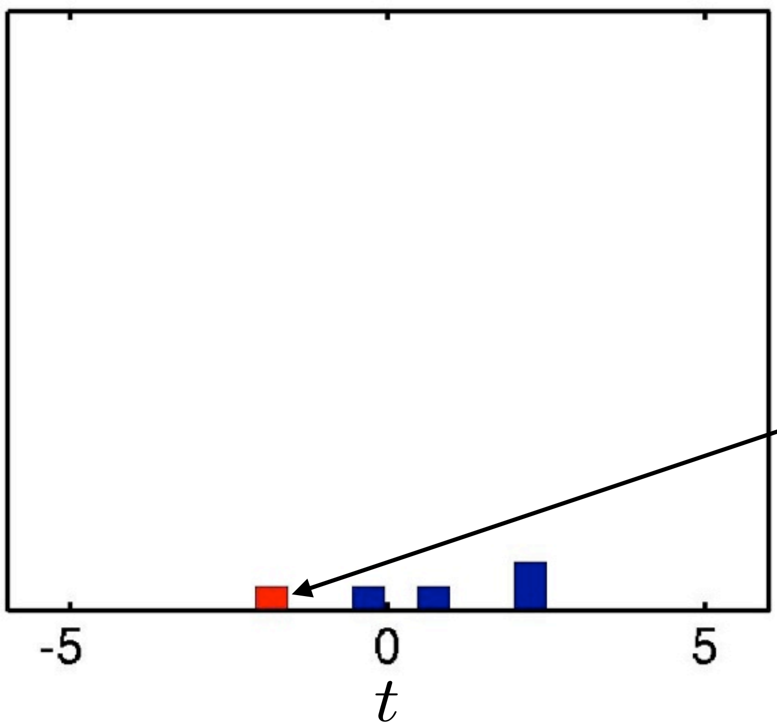
$\sigma^2 = 0.78$

Constant常量

Tools of classical inference 经典推断的工具

1. A null-hypothesis 零假设
2. A test-statistic 检验统计
3. A null-distribution 零分布





$$t = -1.66$$

$$\mathbf{c}^T\widehat{\boldsymbol{\beta}} = -0.69$$

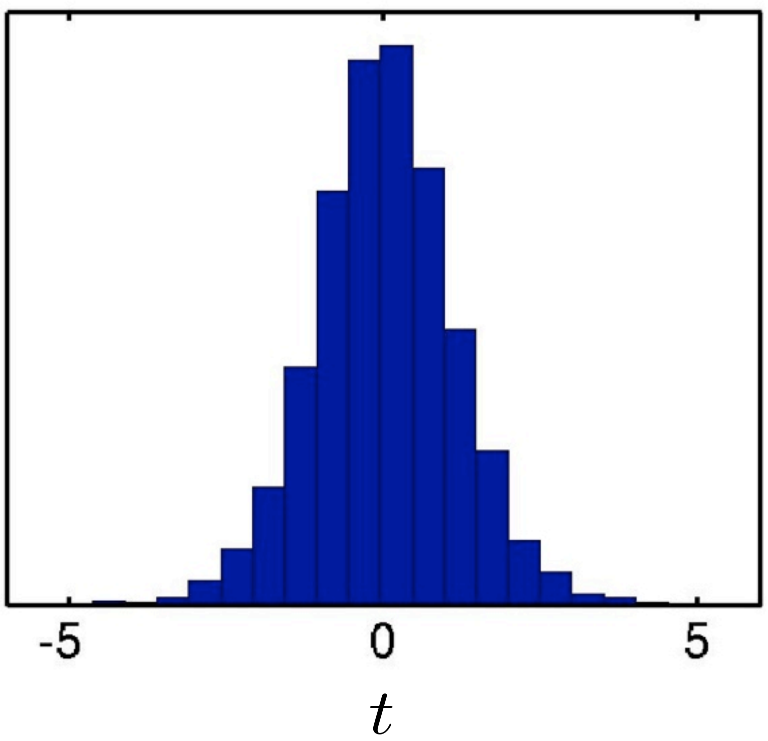
$$t = \frac{\mathbf{c}^T\widehat{\boldsymbol{\beta}}}{\sqrt{\sigma^2}\sqrt{\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}}}$$

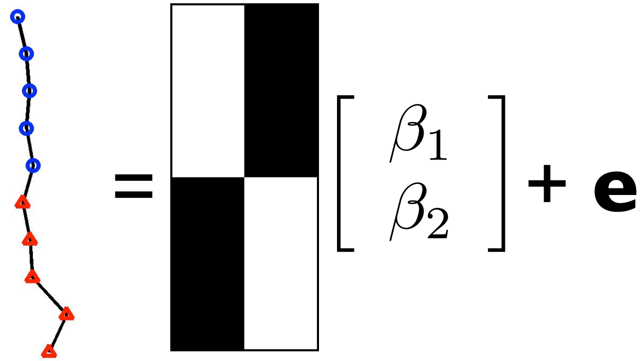
$$\sigma^2 = 0.44$$

Constant常量

Tools of classical inference 经典推断的工具

1. A null-hypothesis 零假设
2. A test-statistic 检验统计
3. A null-distribution 零分布

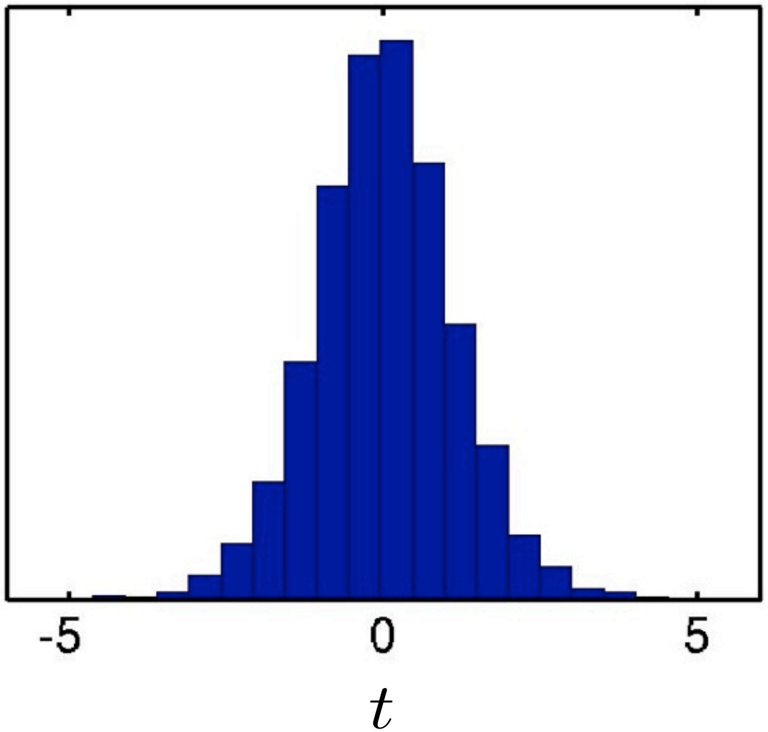




And if we do this til the cows come home
我们一直这么做直到花儿也谢了

Tools of classical inference 经典推断的工具

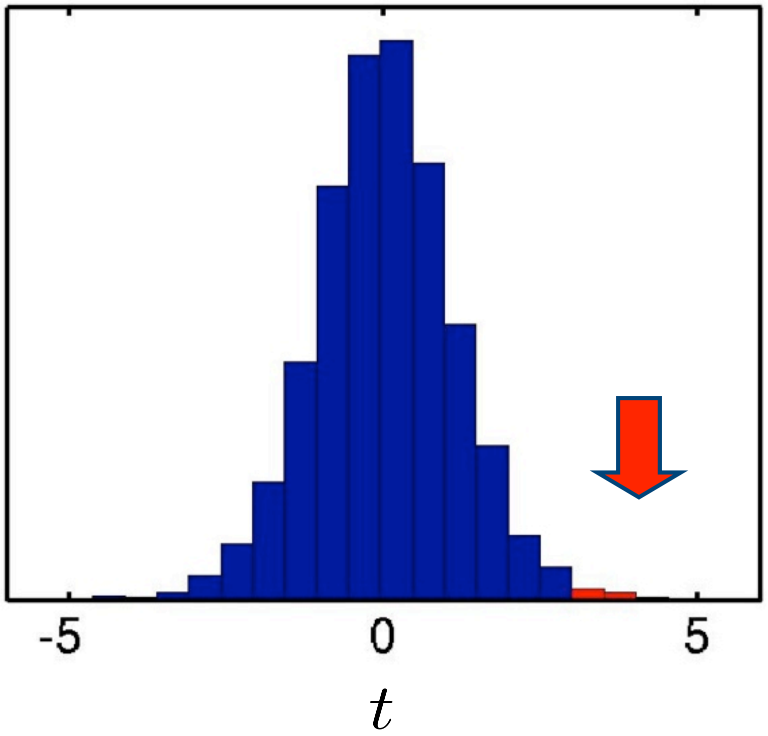
1. A null-hypothesis 零假设
2. A test-statistic 检验统计
3. A null-distribution 零分布



So, why is this helpful?
那么，为什么这样有帮助呢?

Tools of classical inference 经典推断的工具

1. A null-hypothesis 零假设
2. A test-statistic 检验统计
3. A null-distribution 零分布

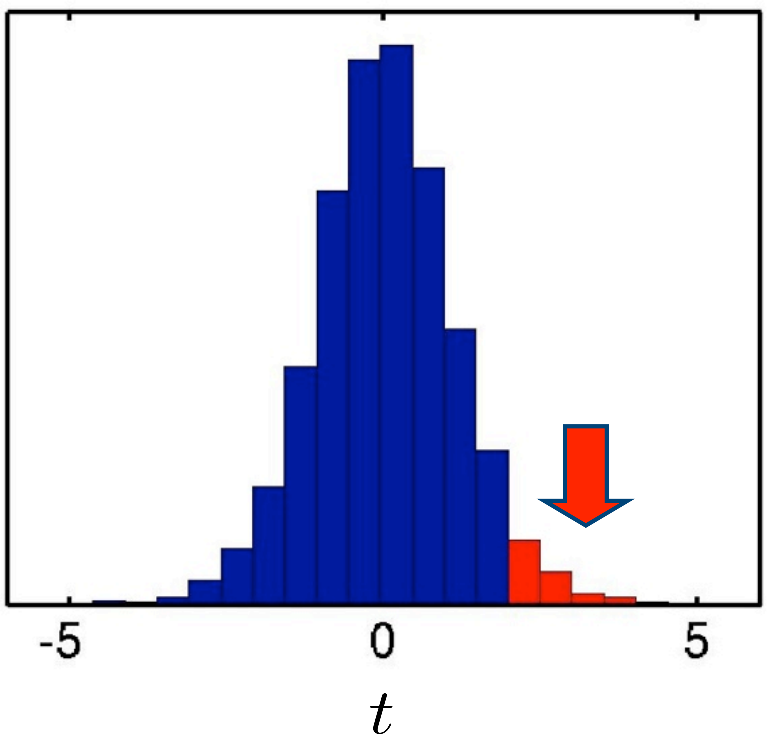


Well, it for example tells us that in ~1% of the cases $t >$ 3.00, even when the null-hypothesis is true.

好吧，他告诉我们大约1%的可能是t > 3.00，即零假设为真。

Tools of classical inference 经典推断的工具

1. A null-hypothesis 零假设
2. A test-statistic 检验统计
3. A null-distribution 零分布



Or that in ~5% of the cases t > 1.99.

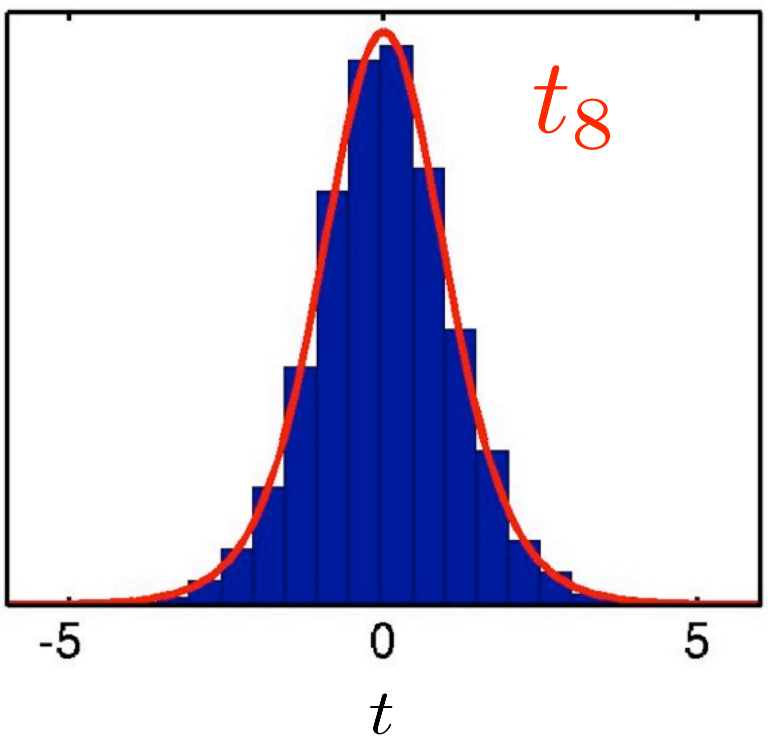
When the null-hypothesis is true.

或者在5%的情况下t> 1.99

当零假设为真

Tools of classical inference 经典推断的工具

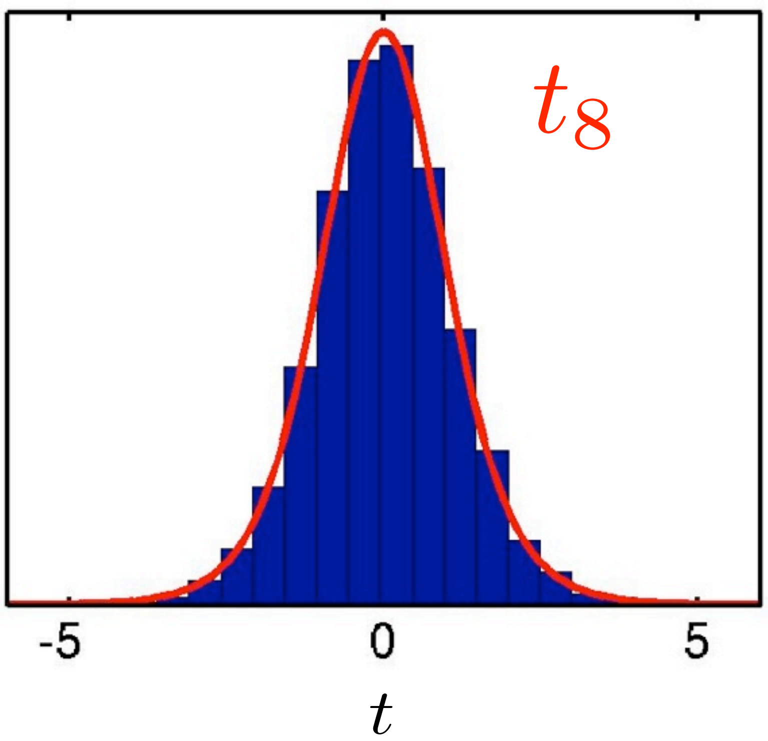
1. A null-hypothesis 零假设
2. A test-statistic 检验统计
3. A null-distribution 零分布



And best of all: This distribution is known *i.e.* one can calculate it.
Much as one can calculate sine or cosine
最重要的是：这种分布是已知的，即可以计算它。
就像每一个人可以计算正弦或余弦

Tools of classical inference 经典推断的工具

1. A null-hypothesis 零假设
2. A test-statistic 检验统计
3. A null-distribution 零分布



And best of all: This distribution is known *i.e.* one can calculate it.
Much as one can calculate sine or cosine
最重要的是：这种分布是已知的，即可以计算它。
就像每一个人可以计算正弦或余弦

Provided that $\mathbf{e} \sim N(0,\sigma^2)$
规定$\mathbf{e} \sim N(0,\sigma^2)$

An example experiment 举例实验

1. A null-hypothesis 零假设 H_0: $\overline{x}_1 = \overline{x}_2$, H_1: $\overline{x}_1 > \overline{x}_2$
2. A test-statistic 检验统计
3. A null-distribution 零分布

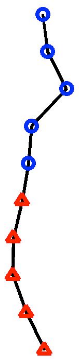
So, with these tools let us do an experiment
利用这些工具，我们来做个实验

An example experiment 举例实验

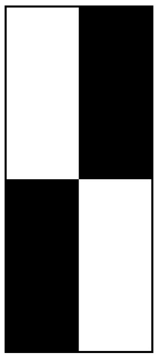
1. A null-hypothesis 零假设 H_0: $\overline{x}_1 = \overline{x}_2$, H_1: $\overline{x}_1 > \overline{x}_2$
2. A test-statistic 检验统计 $t_8 = 2.64$
3. A null-distribution 零分布

So, with these tools let us do an experiment

利用这些工具，我们来做个实验



=



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

$$t = \frac{\mathbf{c}^T \widehat{\boldsymbol{\beta}}}{\sqrt{\sigma^2}\sqrt{\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}}} = \frac{1.53}{\sqrt{0.85}\sqrt{0.4}} = 2.64$$

An example experiment 举例实验

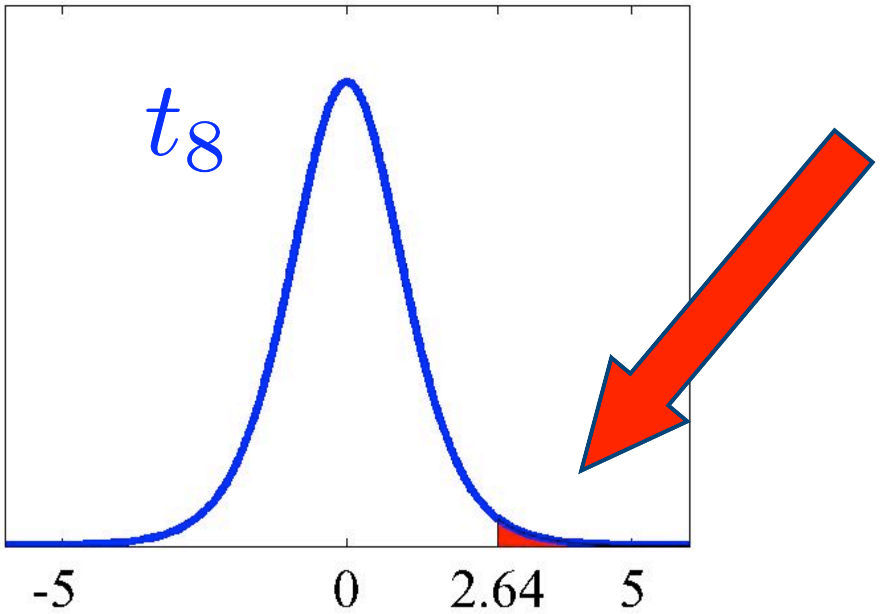
1. A null-hypothesis 零假设
2. A test-statistic 检验统计
3. A null-distribution 零分布

H_0: $\overline{x}_1 = \overline{x}_2$, H_1: $\overline{x}_1 > \overline{x}_2$

$t_8 = 2.64$

So, with these tools let us do an experiment

利用这些工具，我们来做个实验



If the null-hypothesis is true, we would expect to have a ~1.46% chance of finding a t-value this large or larger

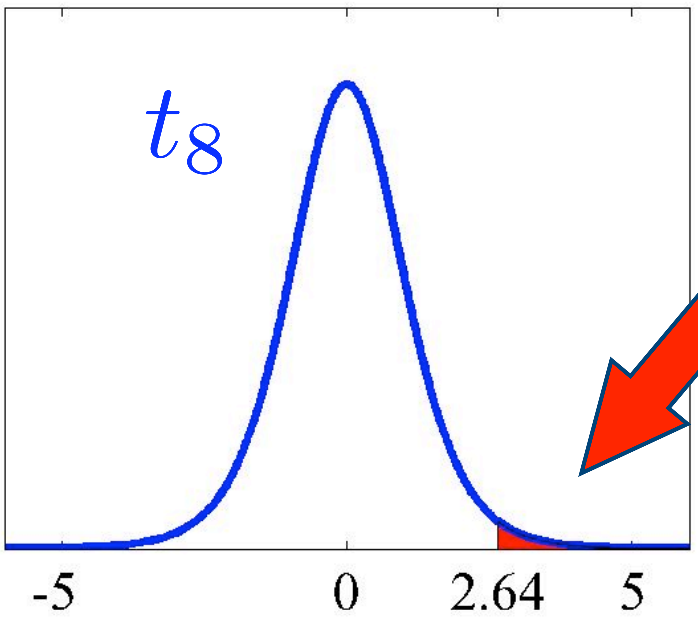
如果原假设为真，则我们期望有1.46%的机会找到如此大或更大的t值

An example experiment 举例实验

1. A null-hypothesis 零假设 H_0: $\overline{x}_1 = \overline{x}_2$, H_1: $\overline{x}_1 > \overline{x}_2$
2. A test-statistic 检验统计 $t_8 = 2.64$
3. A null-distribution 零分布 $t_8 = 2.64^*$

So, with these tools let us do an experiment

利用这些工具，我们来做个实验



There is ~1.46% risk that we reject the null-hypothesis (i.e. claim we found something) when the null is actually true. We can live with that (well, I can).

当零假设确实为真时，我们有1.46%的风险会拒绝零假设（即声称我们发现了某些东西）我们可以接受这种情况（嗯，至少我可以）

False positives/negatives 假阳性/阴性

- I am sure you have all heard about "false positives" and "false negatives". 我相信大家都听说过"假阳性"和"假阴性"。
- But what does that actually mean? 但这实际上是什么意思?

False positives/negatives 假阳性/阴性

- I am sure you have all heard about "false positives" and "false negatives". 我相信大家都听说过"假阳性"和"假阴性"。
- But what does that actually mean? 但这实际上是什么意思?

- We want to perform an experiment and as part of that we define a null-hypothesis, e.g. $H_0 : \ \mu = 0$
  我们想进行一个实验，并在其中定义一个零假设
- Now what can happen? 现在会发生什么?

False positives/negatives 假阳性/阴性

- I am sure you have all heard about "false positives" and "false negatives". 我相信大家都听说过"假阳性"和"假阴性"。
- But what does that actually mean? 但这实际上是什么意思?
- We want to perform an experiment and as part of that we define a null-hypothesis, e.g. $H_0 : \mu = 0$
  我们想进行一个实验，并在其中定义一个零假设
- Now what can happen? 现在会发生什么?

$$\left.\begin{array}{l}\text{H}_0\text{ is true真}\\ \text{H}_0\text{ is false假}\end{array}\right\}\text{True state of affairs 真实情况}$$

False positives/negatives 假阳性/阴性

- I am sure you have all heard about "false positives" and "false negatives". 我相信大家都听说过"假阳性"和"假阴性"。
- But what does that actually mean? 但这实际上是什么意思?
- We want to perform an experiment and as part of that we define a null-hypothesis, e.g. $H_0 : \mu = 0$
  我们想进行一个实验，并在其中定义一个零假设
- Now what can happen? 现在会发生什么?

H_0 is true真
H_0 is false假
} True state of affairs 真实情况

We don't reject H_0不拒绝H_0
We reject H_0拒绝H_0
} Our decision 我们的决定

False positives/negatives 假阳性/阴性

H_0 is true真
H_0 is false假
} True state of affairs真实情况

We don’t reject H_0不拒绝H_0
We reject H_0拒绝H_0
} Our decision我们的决定

	We don’t reject H_0不拒绝	We reject H_0拒绝
H_0 is true为真		
H_0 is false为假		

False positives/negatives 假阳性/阴性

H_0 is true真
H_0 is false假
True state of affairs真实情况

We don’t reject H_0不拒绝H_0
We reject H_0拒绝H_0
Our decision我们的决定

	We don’t reject H_0不拒绝	We reject H_0拒绝
H_0 is true为真	☺️	
H_0 is false为假		☺️

False positives/negatives 假阳性/阴性

H_0 is true真
H_0 is false假
} True state of affairs真实情况

We don't reject H_0不拒绝H_0
We reject H_0拒绝H_0
} Our decision我们的决定

	We don't reject H_0不拒绝	We reject H_0拒绝
H_0 is true为真	☺️	False positive假阳性
H_0 is false为假	False negative假阴性	☺️

False positives/negatives 假阳性/阴性

H_0 is true真
H_0 is false假
} True state of affairs真实情况

We don’t reject H_0不拒绝H_0
We reject H_0拒绝H_0
} Our decision我们的决定

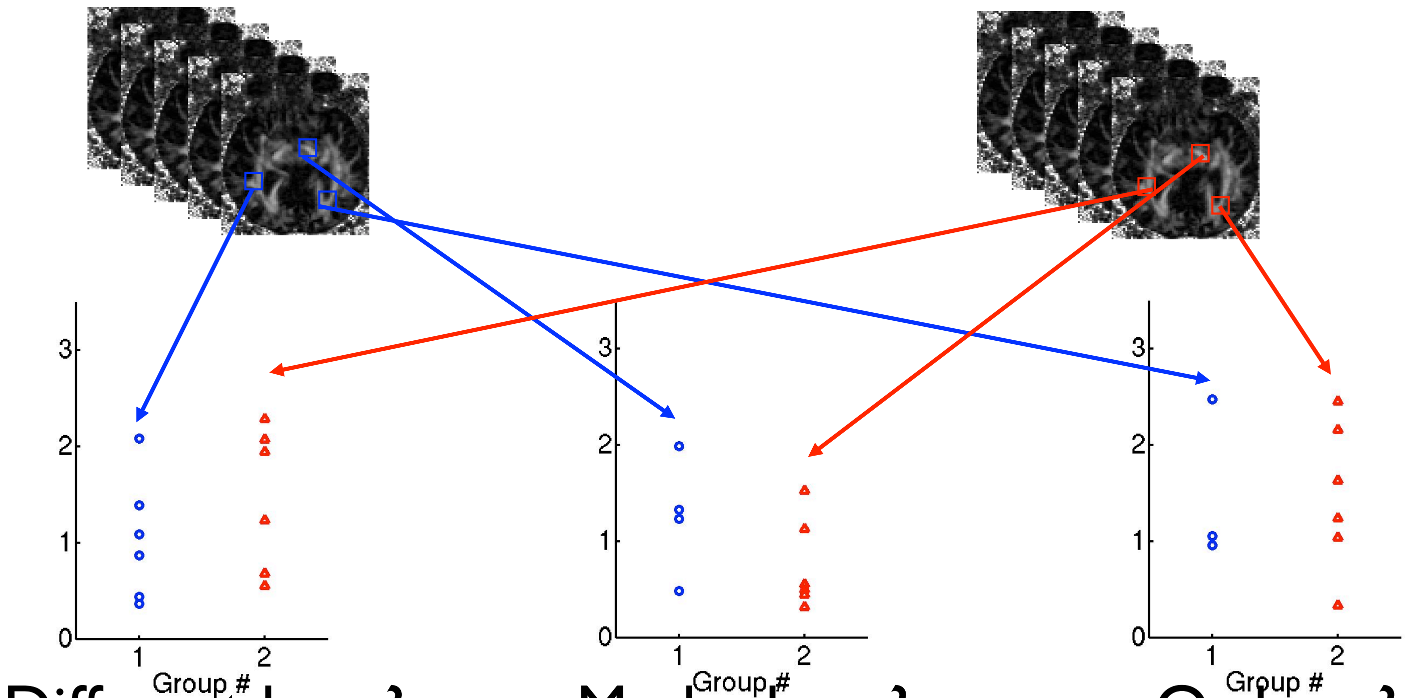
	We don’t reject H_0不拒绝	We reject H_0拒绝
H_0 is true为真	☺️	False positive Type I error 假阳性，I类误差
H_0 is false为假	False negative Type II error 假阴性，II类误差	☺️

Outline大纲

- Null-hypothesis and Null-distribution 零假设和零分布
- Multiple comparisons and Family-wise error 多重比较和族错误率
- Different ways of being surprised 惊奇的不同方式
  - Voxel-wise inference (Maximum z) 体素推断（最大z）
  - Cluster-wise inference (Maximum size) 簇水平推断（最大尺寸）
- Parametric vs non-parametric tests 参数vs非参检验
- Enhanced clusters 增强的簇
- FDR - False Discovery Rate FDR-错误发现率

Multiple Comparisons 多重比较

- In neuroimaging we typically perform <u>many</u> tests as part of a study 在神经影像学中，我们通常会执行许多检验，作为研究的一部分



Different here? 这有差?

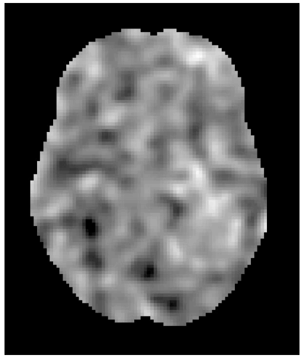
Maybe here? 或者在这?

Or here? 还是这?

• • •

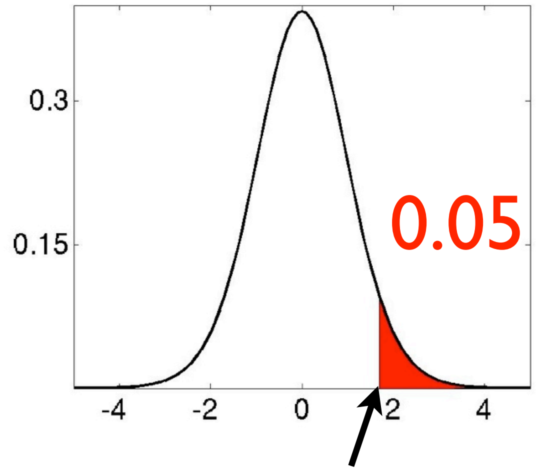
• • •

What happens when we apply this to imaging data? 我们把这个运用到图像数据会发生啥?



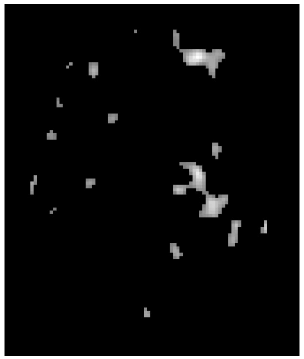
z-map where each voxel ~*N*.
Null-hypothesis true everywhere, i.e. NO ACTIVATIONS
每个体素正态分布的z图。
零假设在所有位置都对，也就是没激活

z



1.64

z-map thresholded at 1.64
阈值在1.64



16 clusters
288 voxels
~5.5% of the voxels
差不多有16个簇，288个体素，
约占这些体素的5.5%

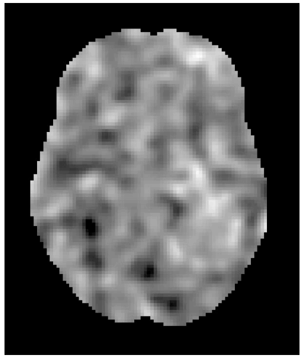
That's a LOT of false positives
有很多误报（假阳性）

Italians doing maths:
The Bonferroni correction

意大利人的数学：Bonferroni校正

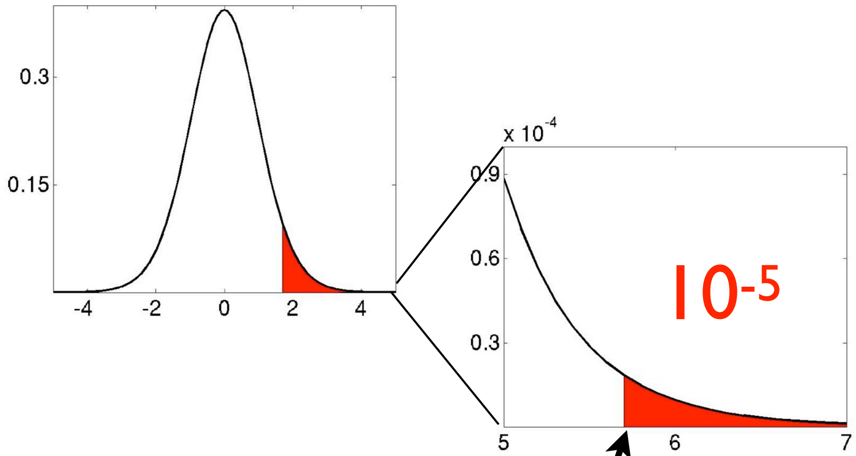
Bonferroni says threshold at α divided by # of tests

阈值应该是α除以测试次数



5255 voxels

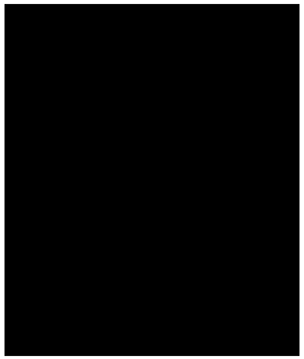
$0.05/5255 \approx 10^{-5}$



5.65

z-map
thresholded at
5.65

阈值在5.65



No false positives.
Hurrah for Italy!

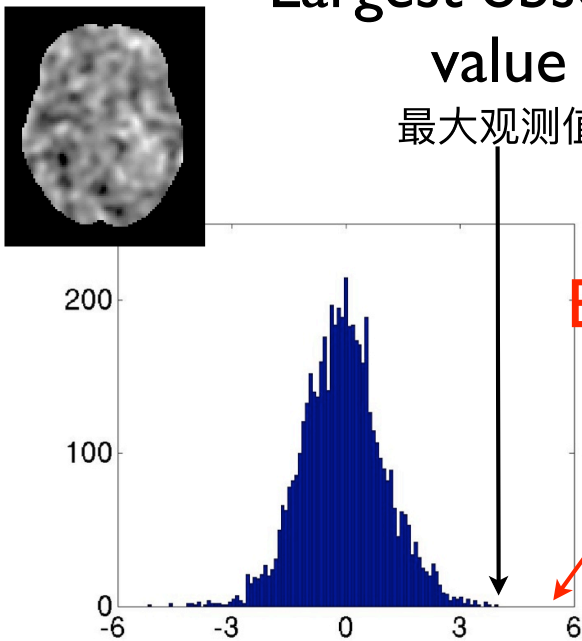
无误报。 意大利万岁！

But … doesn’t 5.65 sound very high?

但是5.65听起来不是很高么

Largest observed value

最大观测值



Bonferroni threshold

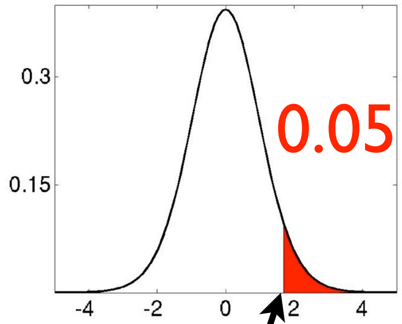
阈限

Observed values in the z-map

观测值的z图

Too lenient

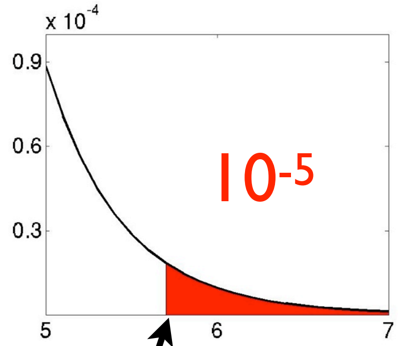
太宽容



1.64

Too harsh

太苛刻



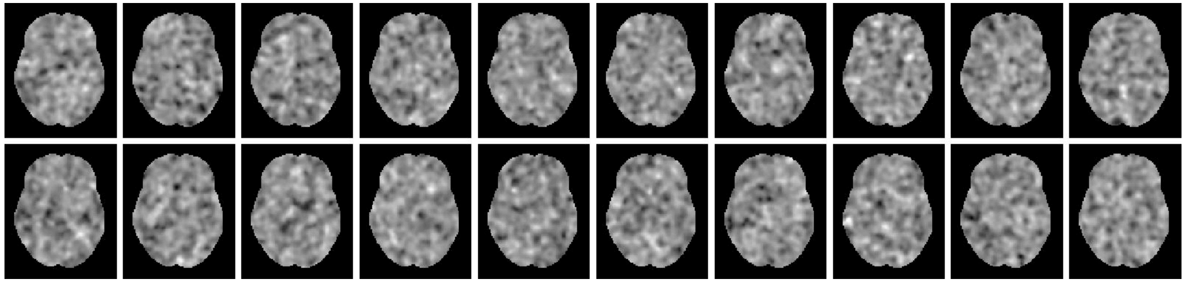
5.65

So what do we want then?

我们到底想要什么?

Family-wise error 族错误率

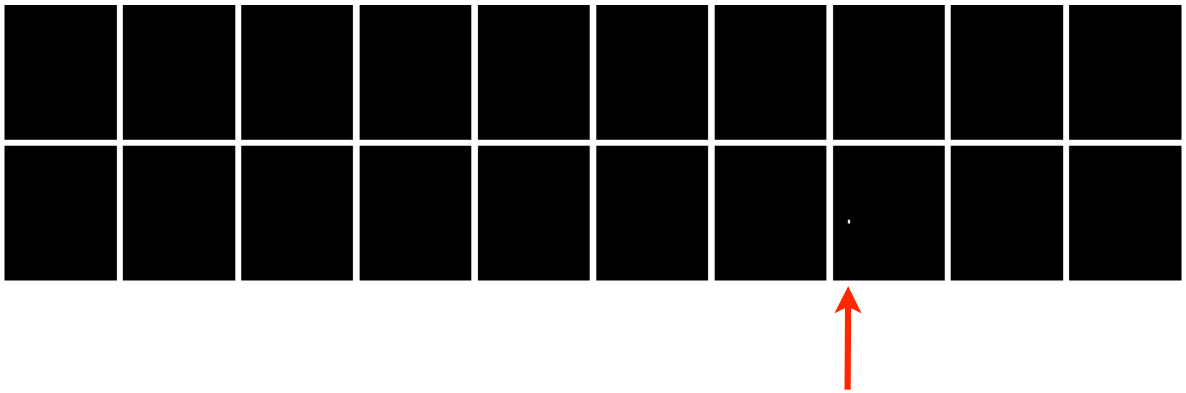
Let’s say we perform a series of identical studies
假设我们进行了一系列完全相同的研究



Each z-map is the end result of a study 每个z图都是研究的最终结果

Let us further say that the null-hypothesis is true.
我们进一步说零假设正确。

We want to threshold the data so that only once in 20 studies do we find a voxel above this threshold.
我们想对数据进行阈值处理，以便20个研究中只有一次发现体素高于此阈值。



But how do we find such a threshold?
但我们怎么找到这个阈值呢

Outline大纲

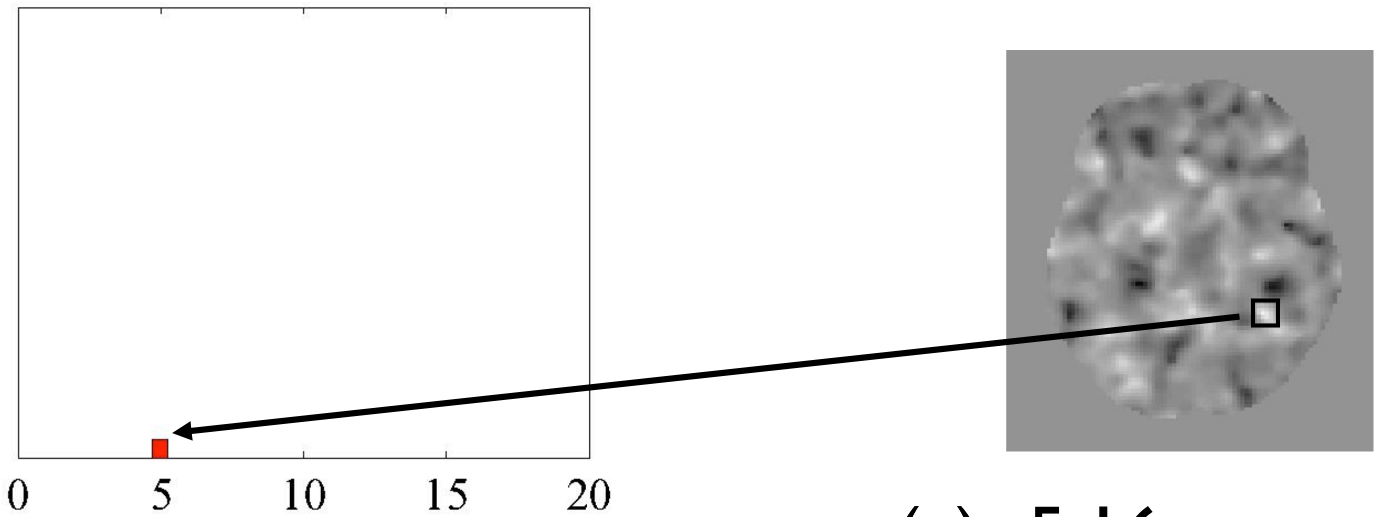
- Null-hypothesis and Null-distribution 零假设和零分布
- Multiple comparisons and Family-wise error 多重比较和族错误率
- Different ways of being surprised 惊奇的不同方式
  - Voxel-wise inference (Maximum z) 体素推断（最大z）
  - Cluster-wise inference (Maximum size) 簇水平推断（最大尺寸）
- Parametric vs non-parametric tests 参数vs非参检验
- Enhanced clusters 增强的簇
- FDR - False Discovery Rate FDR-错误发现率

Maximum z 最大z值

- When we want to control "family-wise error", what do we in practice want?
我们什么时候想控制族错误率，我们实际想要什么呢?

- If the null-hypothesis is true (no activation) we want to reject it no more than 5% of the time. 如果零假设为真（无激活），我们想要不大于5%的时候拒绝他.

- And if we reject anything, we will definitely reject the most "extreme" value (max(z)) in the brain. 如果我们拒绝了所有，我们肯定会拒绝大脑中的最极端的值(max(z))



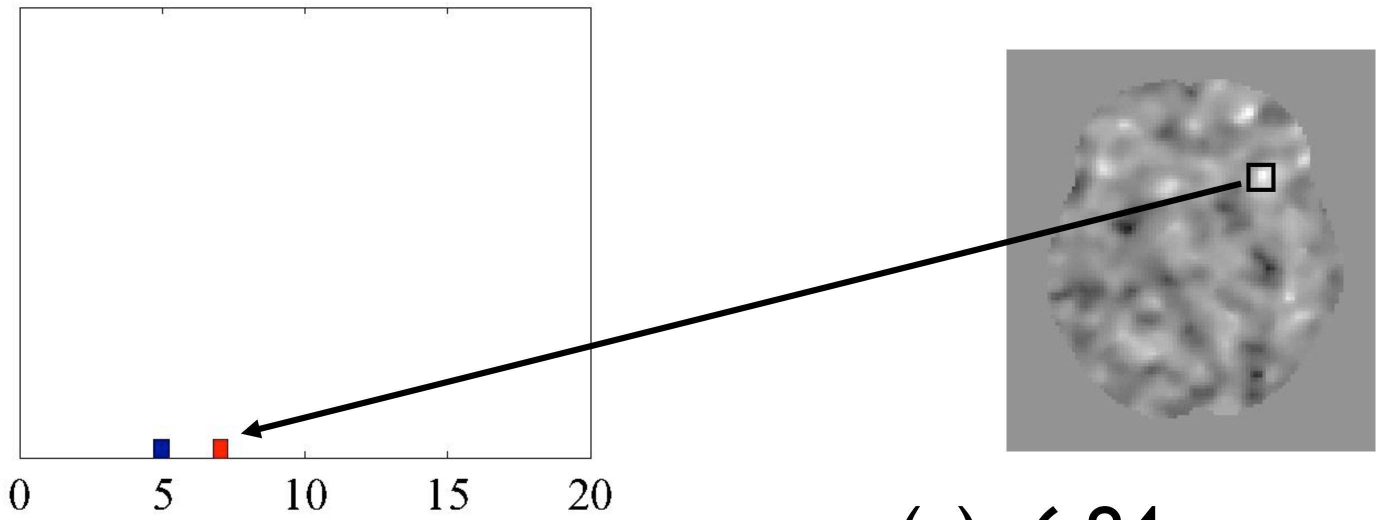
max(z)=5.16

Maximum z 最大z值

- When we want to control "family-wise error", what do we in practice want?
  我们什么时候想控制族错误率，我们实际想要什么呢?

- If the null-hypothesis is true (no activation) we want to reject it no more than 5% of the time. 如果零假设为真（无激活），我们想要不大于5%的时候拒绝他.

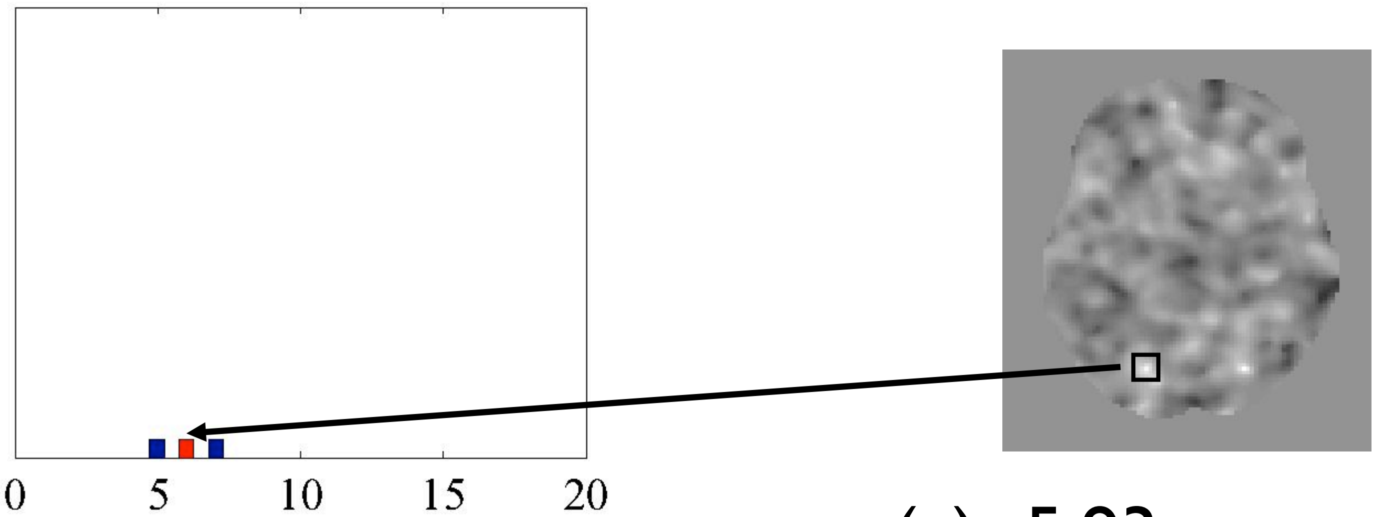
- And if we reject anything, we will definitely reject the most "extreme" value (max(z)) in the brain. 如果我们拒绝了所有，我们肯定会拒绝大脑中的最极端的值(max(z))



max(z)=6.84

Maximum z 最大z值

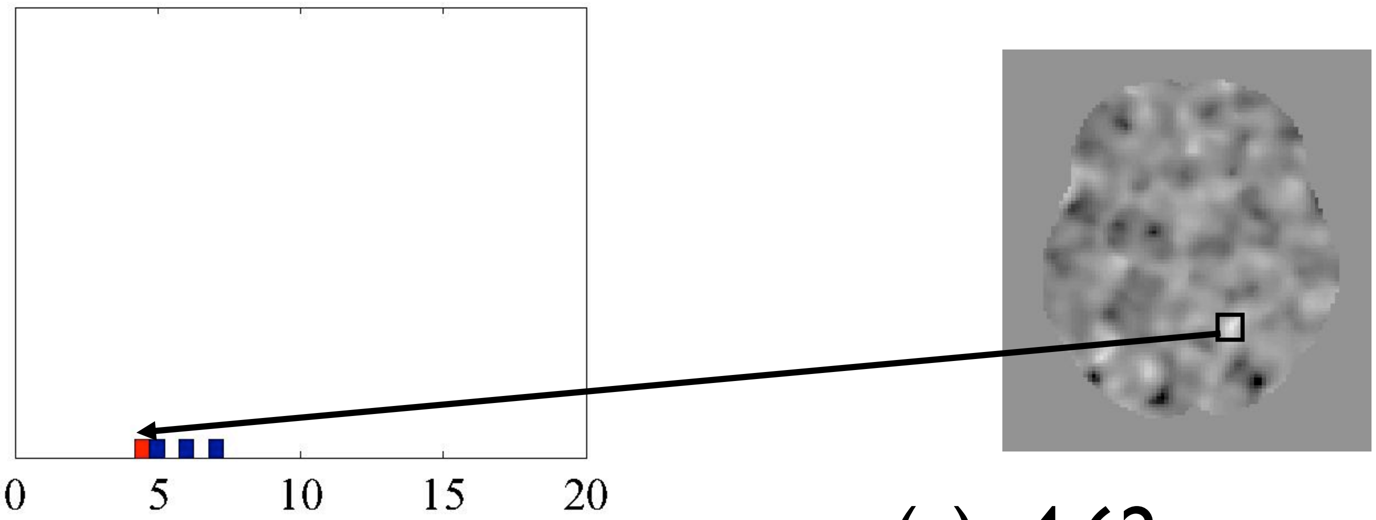
- When we want to control "family-wise error", what do we in practice want?
  我们什么时候想控制族错误率，我们实际想要什么呢?
- If the null-hypothesis is true (no activation) we want to reject it no more than 5% of the time. 如果零假设为真（无激活），我们想要不大于5%的时候拒绝他.
- And if we reject anything, we will definitely reject the most "extreme" value (max(z)) in the brain. 如果我们拒绝了所有，我们肯定会拒绝大脑中的最极端的值(max(z))



max(z)=5.93

Maximum z 最大z值

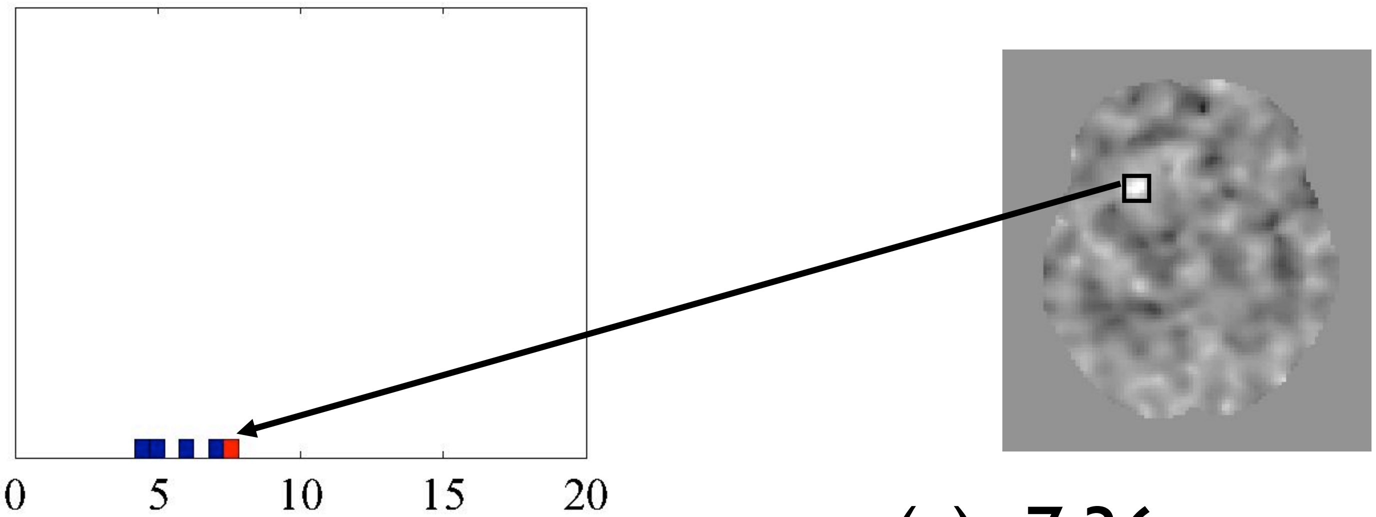
- When we want to control "family-wise error", what do we in practice want?
  我们什么时候想控制族错误率，我们实际想要什么呢?
- If the null-hypothesis is true (no activation) we want to reject it no more than 5% of the time. 如果零假设为真（无激活），我们想要不大于5%的时候拒绝他.
- And if we reject anything, we will definitely reject the most "extreme" value (max(z)) in the brain. 如果我们拒绝了所有，我们肯定会拒绝大脑中的最极端的值(max(z))



max(z)=4.62

Maximum z 最大z值

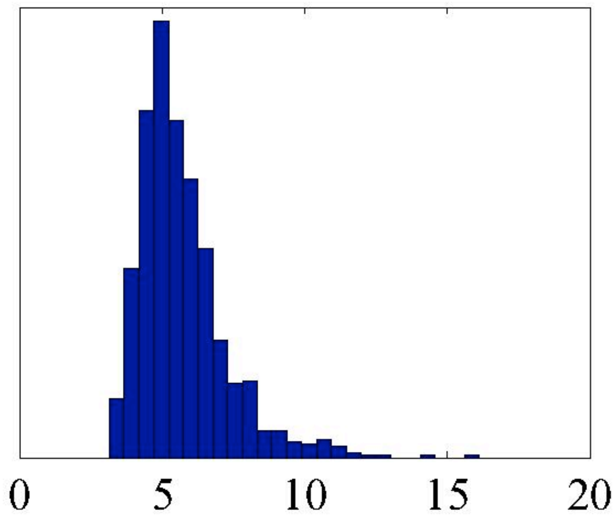
- When we want to control "family-wise error", what do we in practice want?
  我们什么时候想控制族错误率，我们实际想要什么呢?
- If the null-hypothesis is true (no activation) we want to reject it no more than 5% of the time. 如果零假设为真（无激活），我们想要不大于5%的时候拒绝他.
- And if we reject anything, we will definitely reject the most "extreme" value (max(z)) in the brain. 如果我们拒绝了所有，我们肯定会拒绝大脑中的最极端的值(max(z))



max(z)=7.36

Maximum z 最大z值

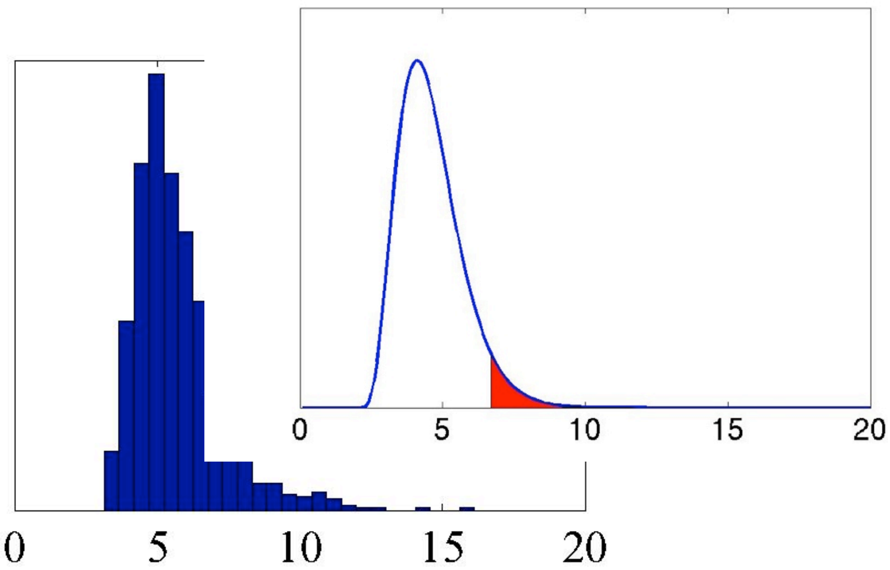
- When we want to control "family-wise error", what do we in practice want?
  我们什么时候想控制族错误率，我们实际想要什么呢?
- If the null-hypothesis is true (no activation) we want to reject it no more than 5% of the time. 如果零假设为真（无激活），我们想要不大于5%的时候拒绝他.
- And if we reject anything, we will definitely reject the most "extreme" value (max(z)) in the brain. 如果我们拒绝了所有，我们肯定会拒绝大脑中的最极端的值(max(z))



Etc…

Maximum z 最大z值

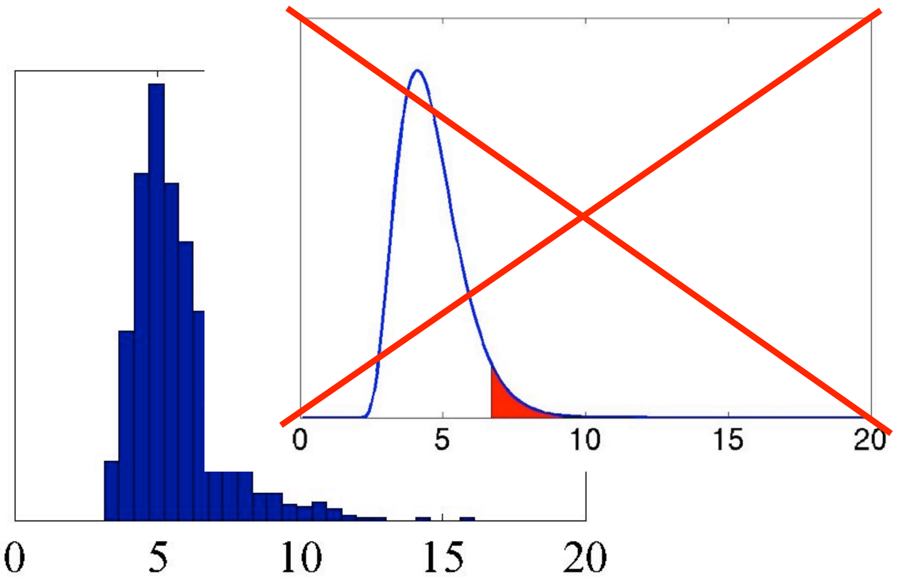
- When we want to control "family-wise error", what do we in practice want?
  我们什么时候想控制族错误率，我们实际想要什么呢?
- If the null-hypothesis is true (no activation) we want to reject it no more than 5% of the time. 如果零假设为真（无激活），我们想要不大于5%的时候拒绝他.
- And if we reject anything, we will definitely reject the most "extreme" value (max(z)) in the brain. 如果我们拒绝了所有，我们肯定会拒绝大脑中的最极端的值(max(z))



This is the distribution we want to use for our FWE control.
这是我们要用于FWE控制的分布。

Maximum z 最大z值

- When we want to control "family-wise error", what do we in practice want?
  我们什么时候想控制族错误率，我们实际想要什么呢?
- If the null-hypothesis is true (no activation) we want to reject it no more than 5% of the time. 如果零假设为真（无激活），我们想要不大于5%的时候拒绝他.
- And if we reject anything, we will definitely reject the most "extreme" value (max(z)) in the brain. 如果我们拒绝了所有，我们肯定会拒绝大脑中的最极端的值(max(z))



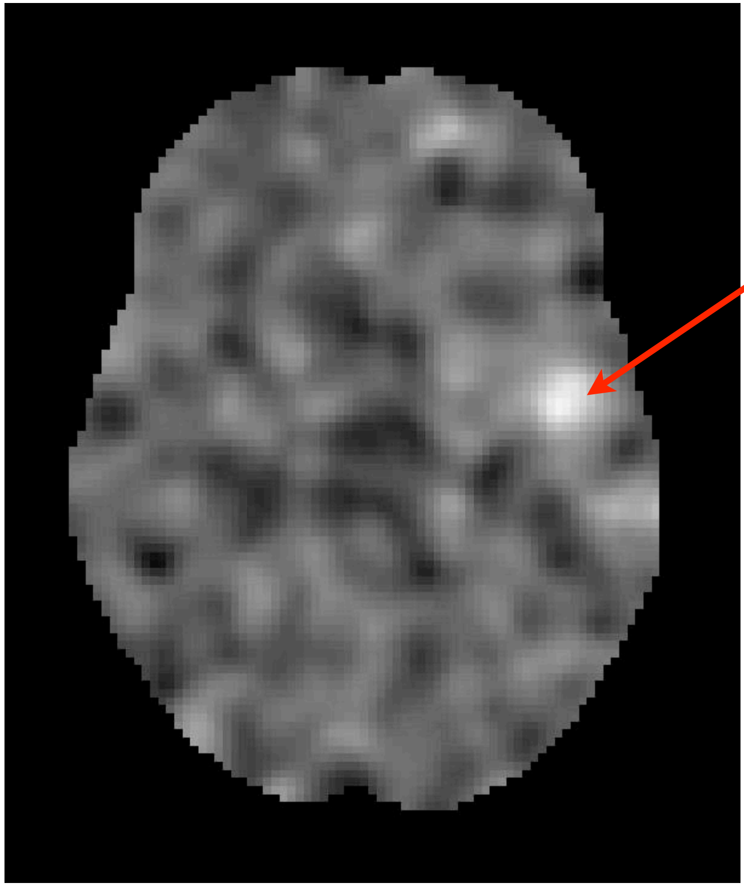
This is the distribution we want to use for our FWE control.
But there is no known expression for it! ☹
这是我们要用于FWE控制的分布。但是没有已知的表达方式！☹

Outline大纲

- Null-hypothesis and Null-distribution 零假设和零分布
- Multiple comparisons and Family-wise error 多重比较和族错误率
- Different ways of being surprised 惊奇的不同方式
  - Voxel-wise inference (Maximum z) 体素推断（最大z)
  - Cluster-wise inference (Maximum size) 簇水平推断（最大尺寸）
- Parametric vs non-parametric tests 参数vs非参检验
- Enhanced clusters 增强的簇
- FDR - False Discovery Rate FDR-错误发现率

Spatial extent: another way to be surprised
空间范围：另一个让人惊奇的方式

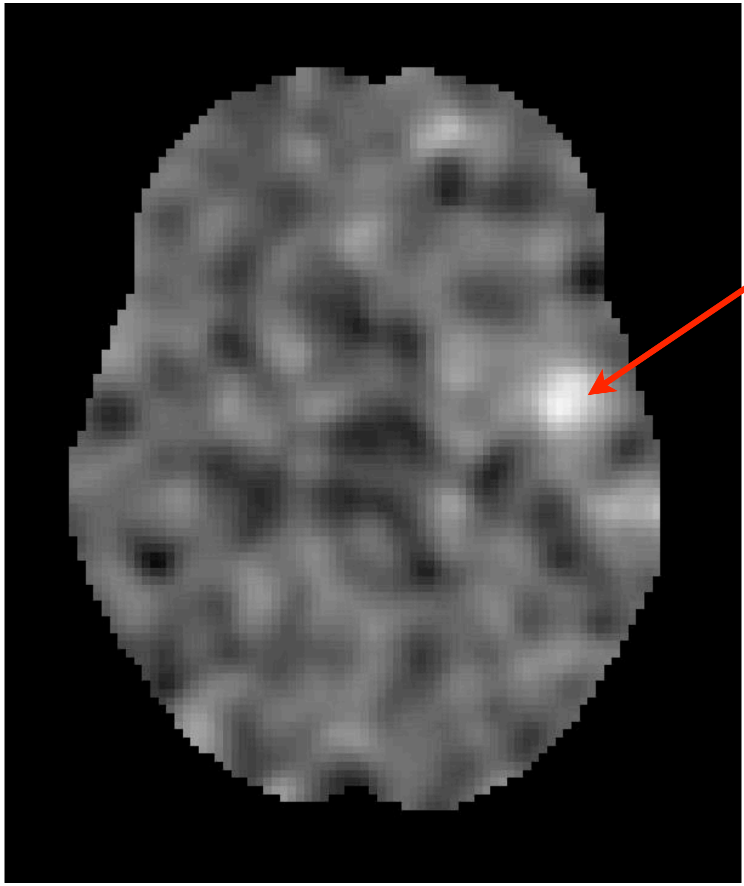
This far we have talked about voxel-based tests
目前为止我们讨论了基于体素的检验。



We say: Look! A z-value of 7. That is so surprising (under the null-hypothesis) that I will have to reject it. (Though we are of course secretly delighted to do so)
看z值等于7，太令人意外了（在零假设下），我不得不拒绝他。（虽然我们当然很乐意这么做。）

Spatial extent: another way to be surprised
空间范围：另一个让人惊奇的方式

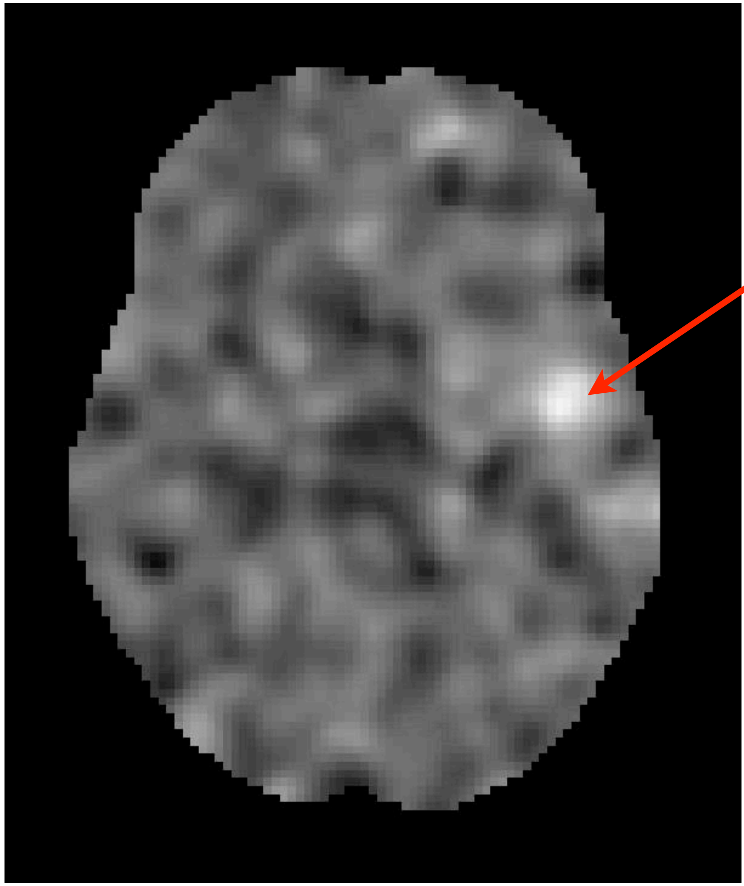
This far we have talked about voxel-based tests
目前为止我们讨论了基于体素的检验。



We say: Look! A z-value of 7. That is so surprising (under the null-hypothesis) that I will have to reject it. (Though we are of course secretly delighted to do so)
看z值等于7，太令人意外了（在零假设下），我不得不拒绝他。（虽然我们当然很乐意这么做。）

Spatial extent: another way to be surprised
空间范围：另一个让人惊奇的方式

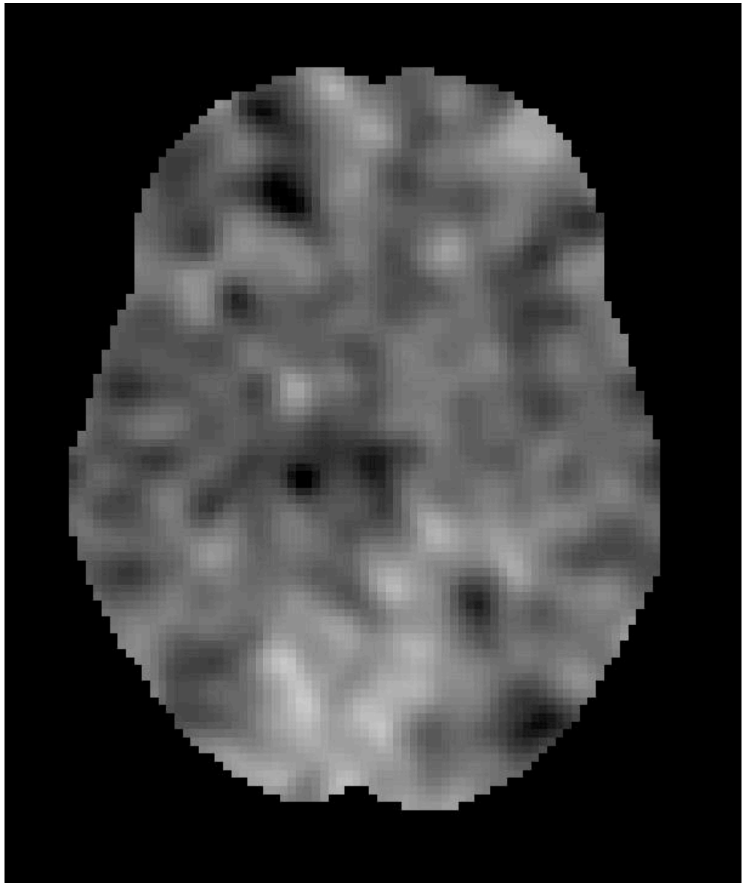
This far we have talked about voxel-based tests
目前为止我们讨论了基于体素的检验。



We say: Look! A z-value of 7. That is so surprising (under the null-hypothesis) that I will have to reject it. (Though we are of course secretly delighted to do so)
看z值等于7，太令人意外了（在零假设下），我不得不拒绝他。（虽然我们当然很乐意这么做。）

Spatial extent: another way to be surprised
空间范围：另一个让人惊奇的方式

But sometimes our data just aren't that surprising.
但有时候我们的数据就没这么意外了。



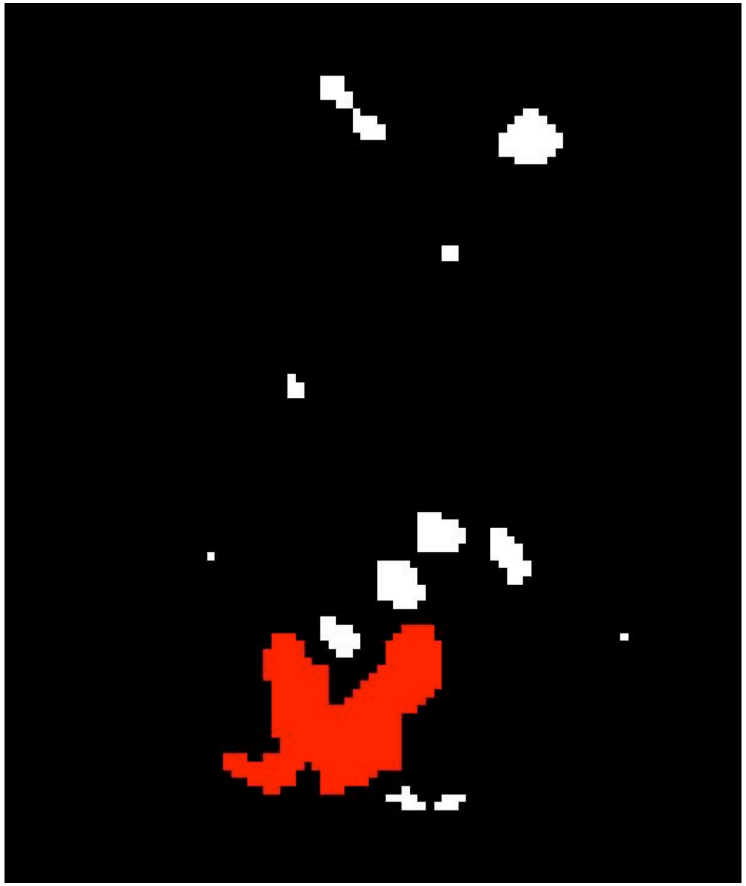
Nothing surprising here! The largest z-value is ~4. We cannot reject the null-hypothesis, and we are **devastated**.
这里不足为奇！最大的z值为～4。 我们不能拒绝零假设，我们为此感到伤心绝望。

Spatial extent: another way to be surprised

空间范围：另一个让人惊奇的方式

So we threshold the z-map at 2.3 (arbitrary threshold) and look at the spatial extent of clusters

因此，我们将z-map的阈值定为2.3（任意阈值），然后查看簇的空间范围



We say: Look at that whopper! 301 connected voxels all with z-values > 2.3. That is really surprising (under the null-hypothesis). I will have to reject it.

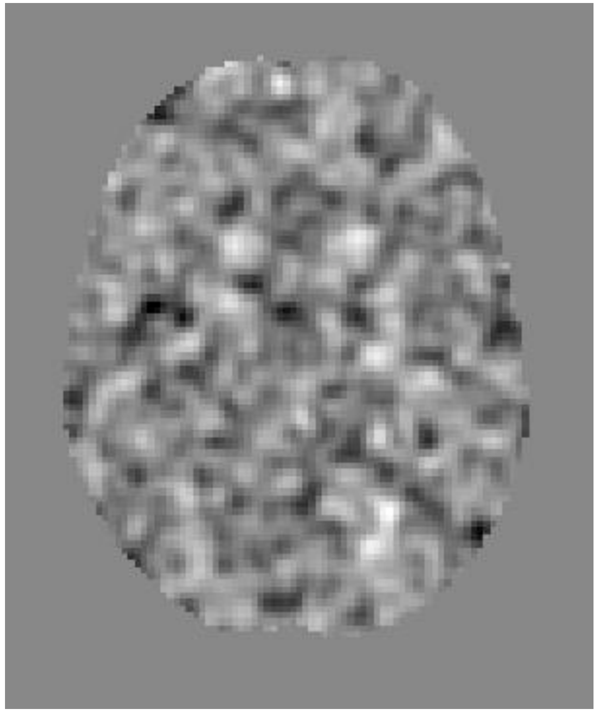
看看那个弥天大谎！301个连接的体素均具有z值> 2.3。这确实令人惊讶（在原假设下）。我将不得不拒绝它。

Distribution of Max Cluster Size

最大簇大小分布

As with the *z*-values we need a "null-distribution". What would that look like in this case?

与z值一样，我们要零分布。这种情况下，看起来是啥样?



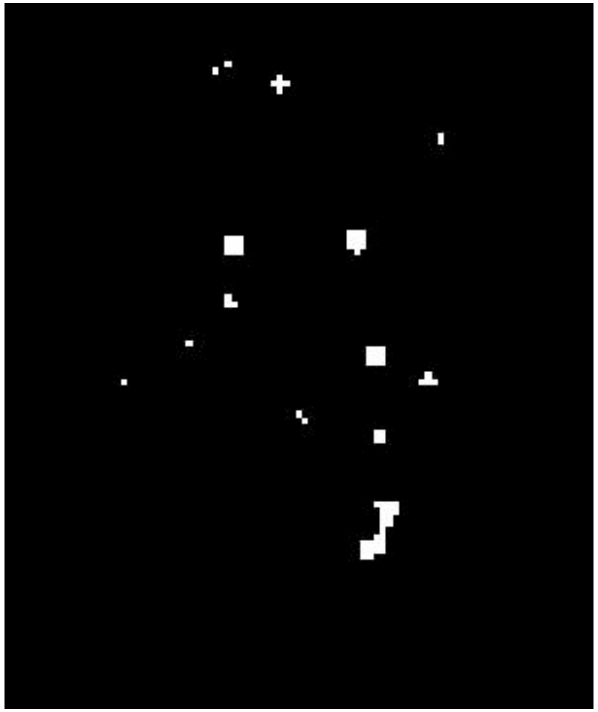
Let's say we have acquired some data

我们采集了一堆数据

Distribution of Max Cluster Size
最大簇大小分布

If we reject any cluster we will reject the largest. So what we want is the distribution of the largest cluster, under the null-hypothesis.

如果我们拒绝任何集群，我们将拒绝最大的集群。 所以我们想要的是零假设下最大簇的分布。



Threshold the z-map at 2.3 (arbitrary)

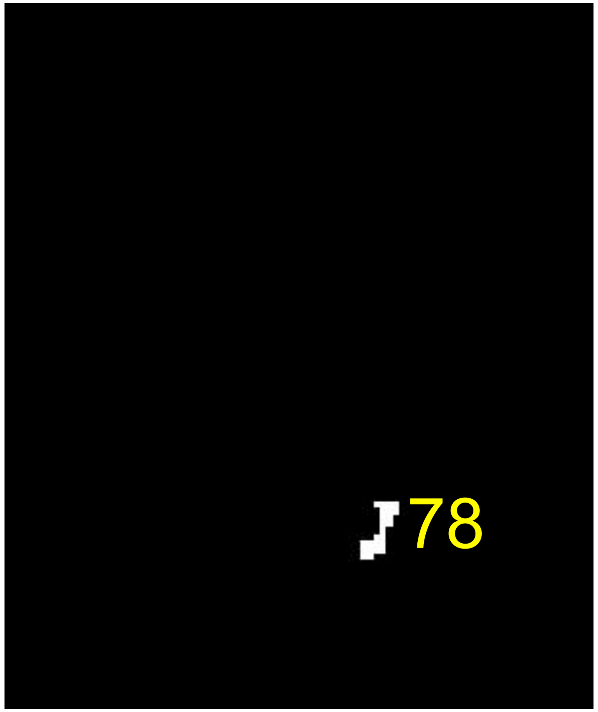
武断地把阈值设在2.3

Distribution of Max Cluster Size

最大簇大小分布

If we reject any cluster we will reject the largest. So what we want is the distribution of the largest cluster, under the null-hypothesis.

如果我们拒绝任何集群，我们将拒绝最大的集群。 所以我们想要的是零假设下最大簇的分布。



Locate the largest cluster anywhere in the brain.

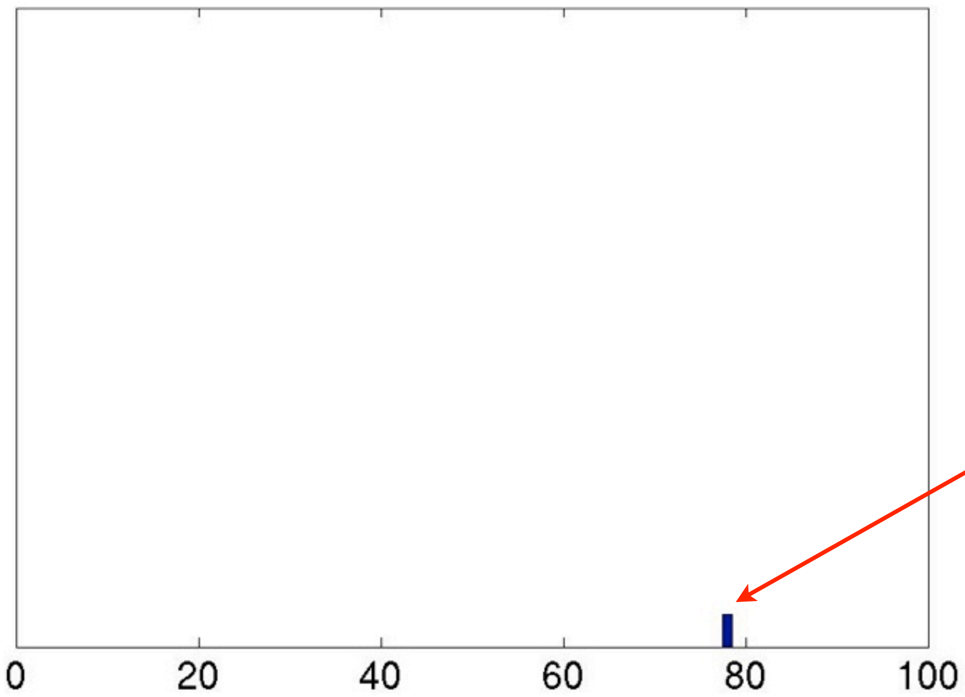
找到大脑中任何位置的最大簇

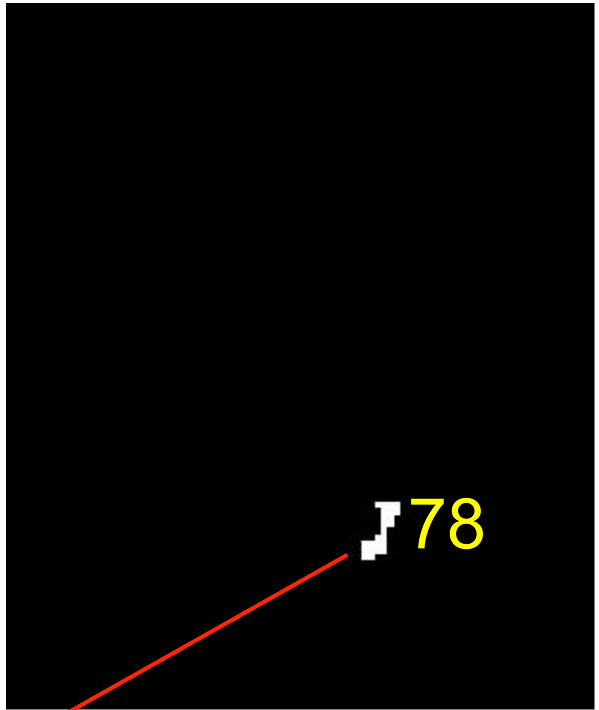
Distribution of Max Cluster Size

最大簇大小分布

If we reject any cluster we will reject the largest. So what we want is the distribution of the largest cluster, under the null-hypothesis.

如果我们拒绝任何集群，我们将拒绝最大的集群。 所以我们想要的是零假设下最大簇的分布。





And record how large it is.

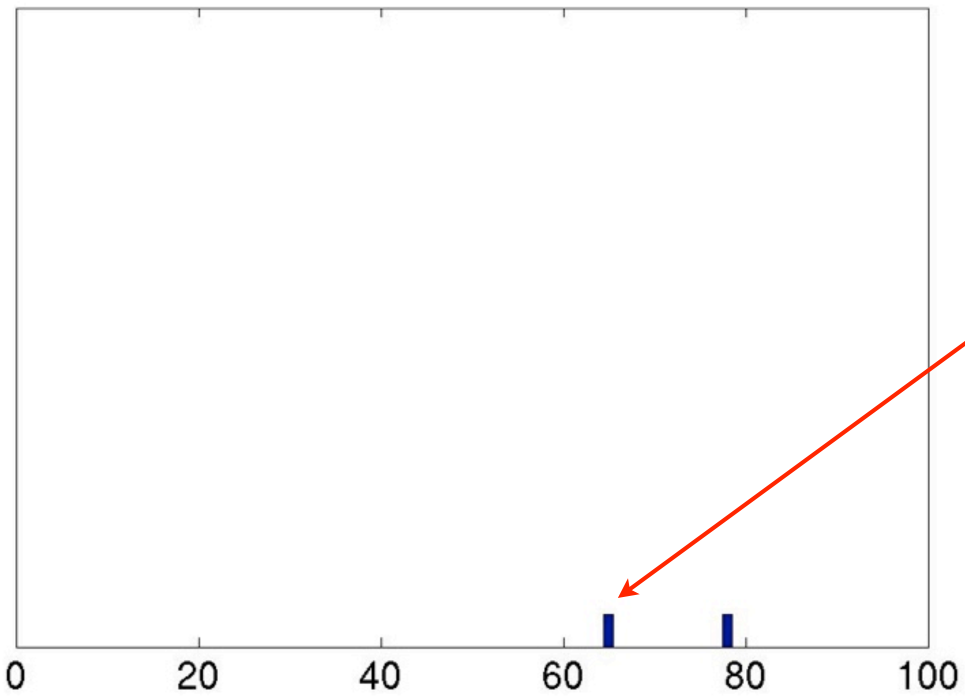
记录下有多大。

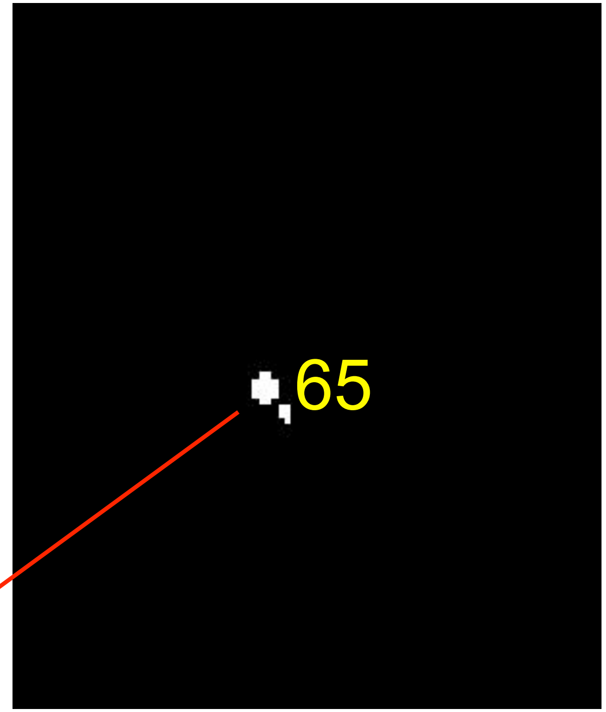
Distribution of Max Cluster Size

最大簇大小分布

If we reject any cluster we will reject the largest. So what we want is the distribution of the largest cluster, under the null-hypothesis.

如果我们拒绝任何集群，我们将拒绝最大的集群。 所以我们想要的是零假设下最大簇的分布。





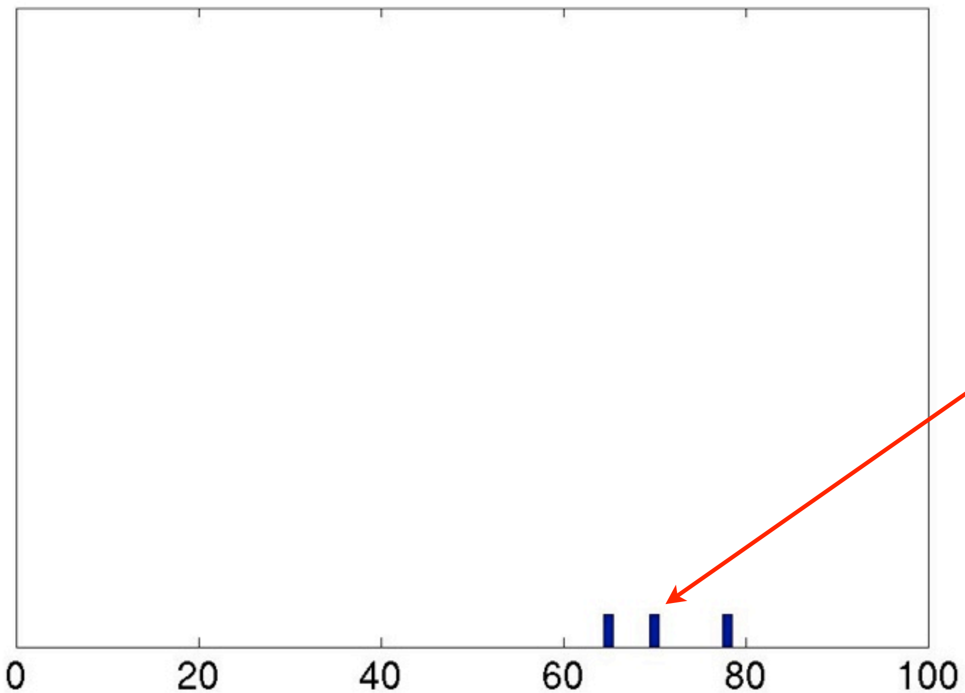
And do the same for another experiment…

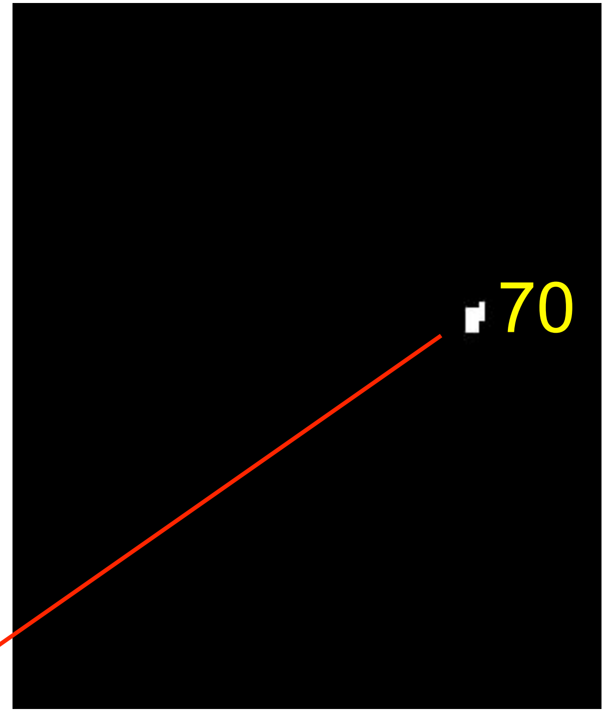
对另一个实验也如此…

Distribution of Max Cluster Size
最大簇大小分布

If we reject any cluster we will reject the largest. So what we want is the distribution of the largest cluster, under the null-hypothesis.

如果我们拒绝任何集群，我们将拒绝最大的集群。 所以我们想要的是零假设下最大簇的分布。





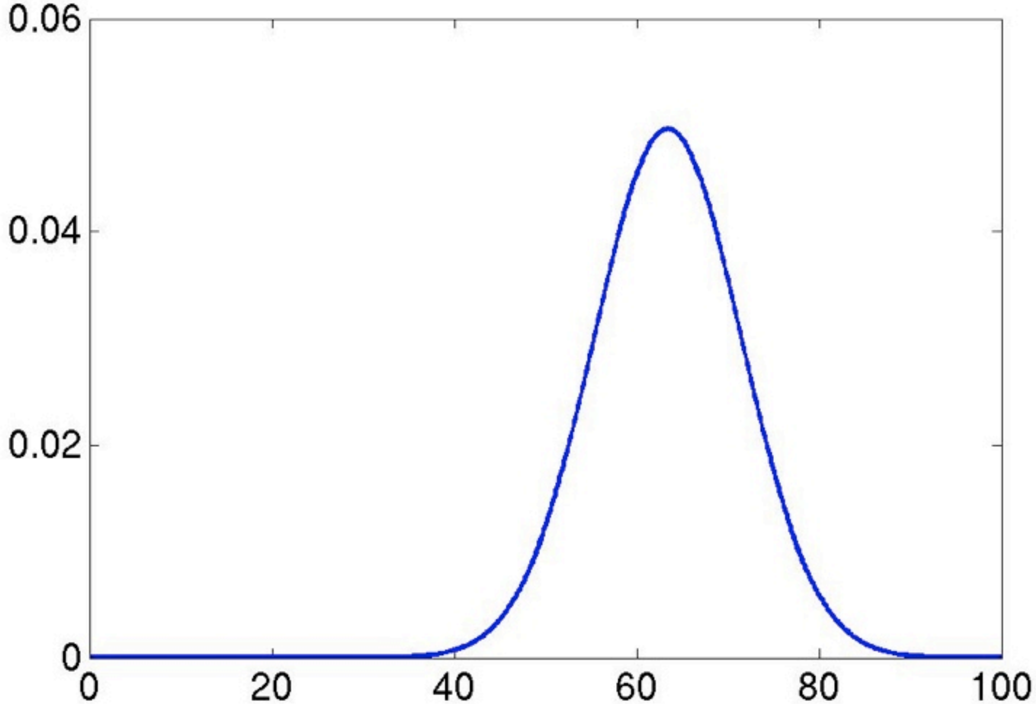
Etc …

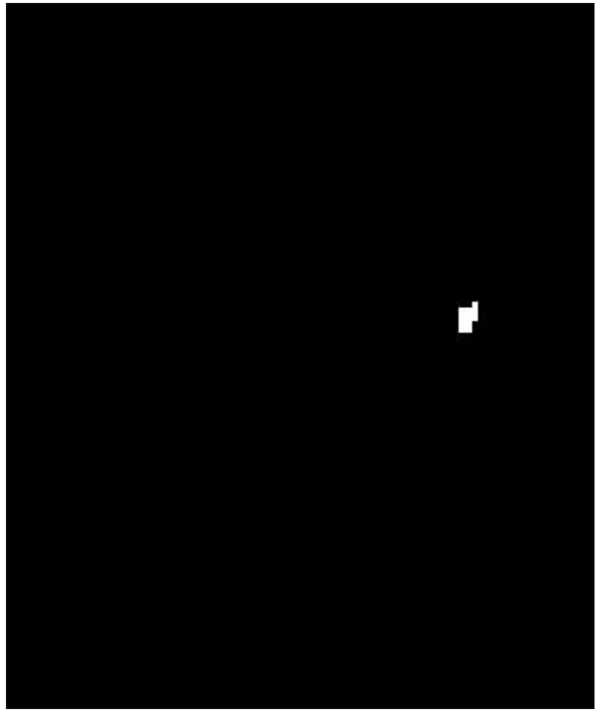
继续。。。

Distribution of Max Cluster Size
最大簇大小分布

If we reject any cluster we will reject the largest. So what we want is the distribution of the largest cluster, under the null-hypothesis.

如果我们拒绝任何集群，我们将拒绝最大的集群。 所以我们想要的是零假设下最大簇的分布。



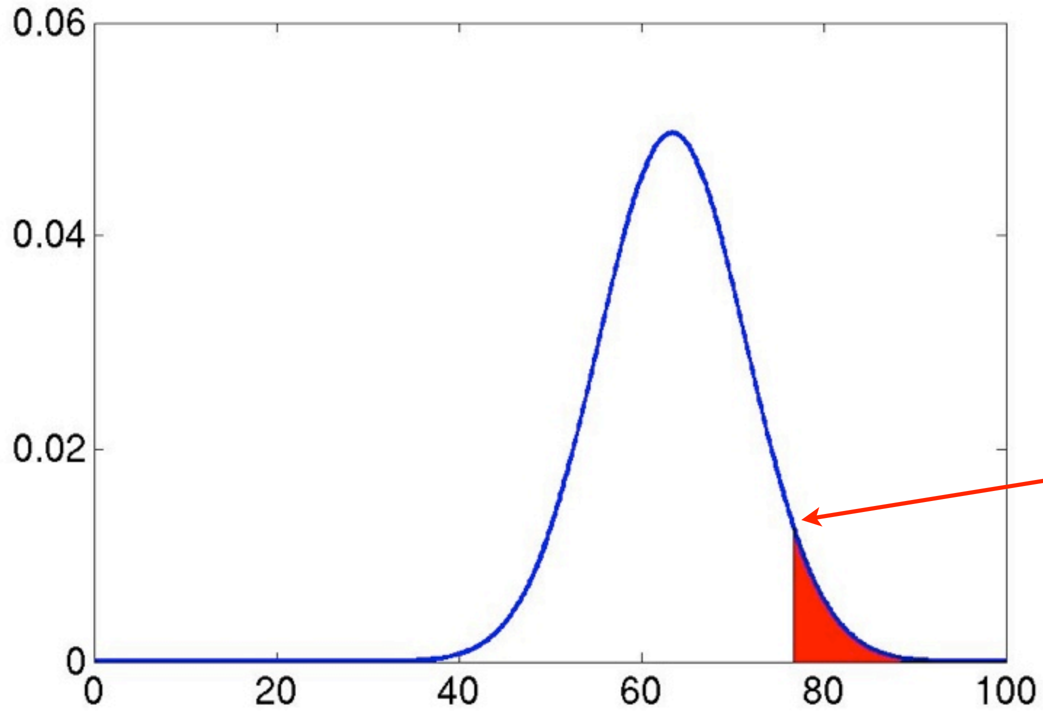


Until we have …

直到我们…

Distribution of Max Cluster Size
最大簇大小分布

If we reject any cluster we will reject the largest. So what we want is the distribution of the largest cluster, under the null-hypothesis.
如果我们拒绝任何集群，我们将拒绝最大的集群。 所以我们想要的是零假设下最大簇的分布。



If we find a cluster larger than 76 voxels we reject the null-hypothesis.
如果我们找到一个簇比76个体素多，我们拒绝零假设。

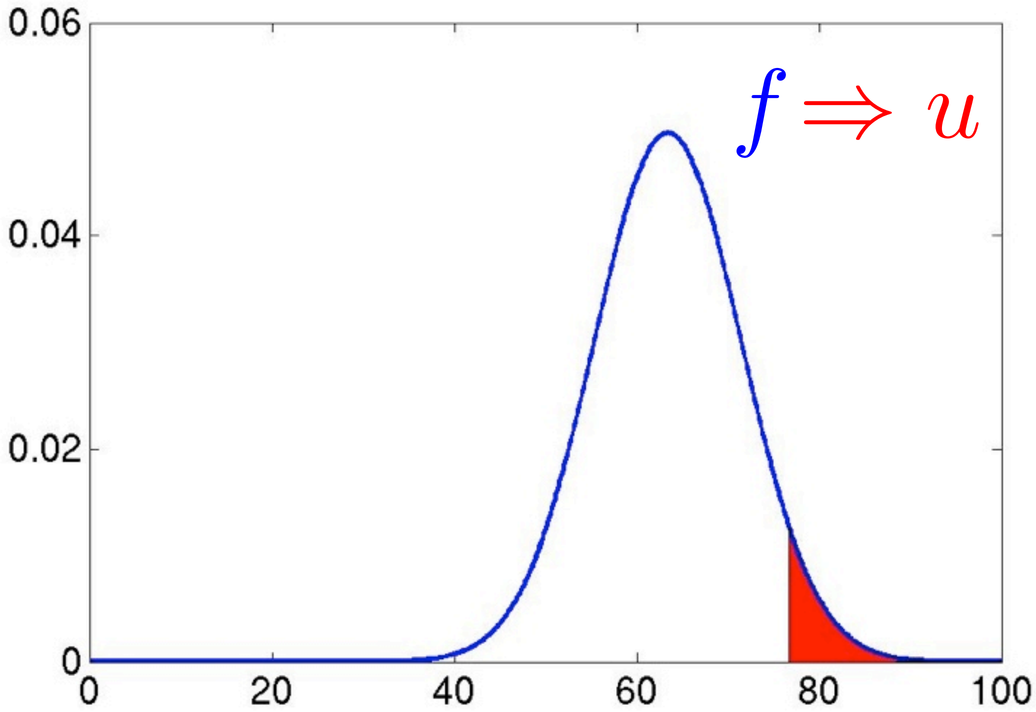
And this (76) is the level we want to threshold at
这个数就是我们想要的阈限。

Distribution of Max Cluster Size

最大簇大小分布

So, just as was the case for the t-values, we now have a distribution f that allows us to calculate a Family Wise threshold u pertaining to cluster size.

因此，就像t值一样，我们现在有一个分布f它使我们能够计算与簇大小有关的总体的阈值 u。



But what does f and u crucially depend on?

但f和 u关键取决于什么？

Distribution of Max Cluster Size

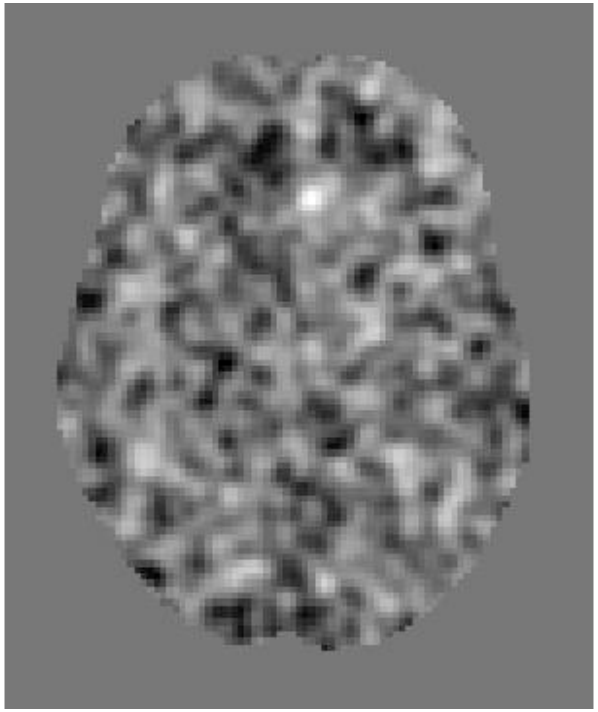
最大簇大小分布

So, just as was the case for the t-values, we now have a distribution f that allows us to calculate a Family Wise threshold u pertaining to cluster size.

因此，就像t值一样，我们现在有一个分布f它使我们能够计算与簇大小有关的总体的阈值 u。

f depends crucially on the initial “cluster-forming” threshold?

f关键取决于初始“簇形成”的阈值?



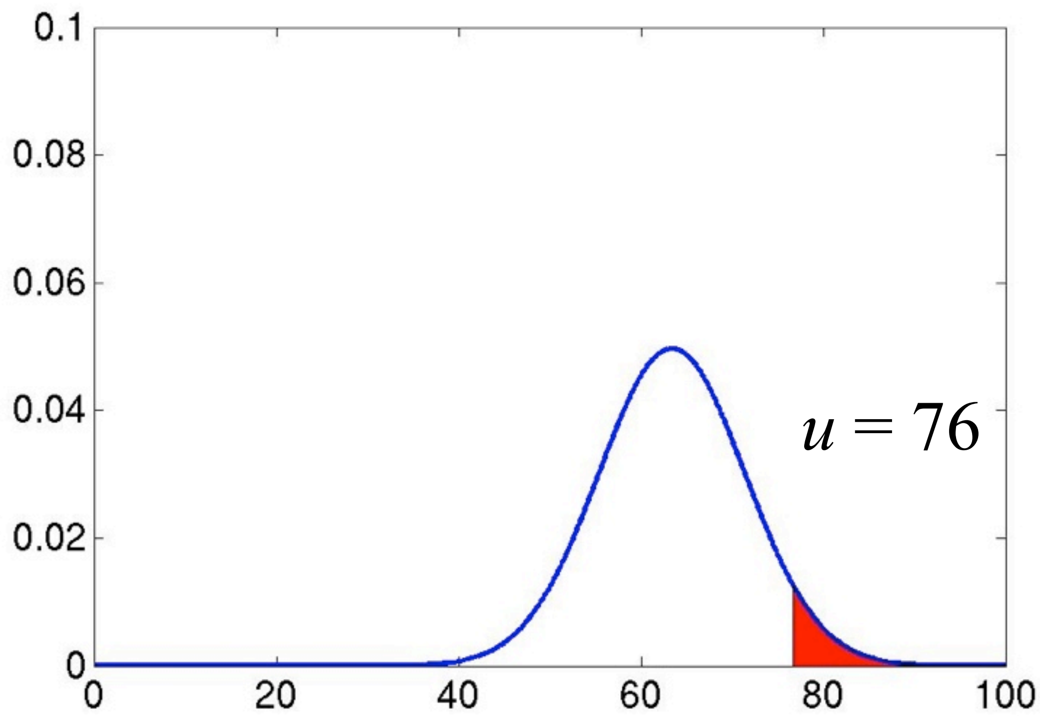
$z = 2.3$

Distribution of Max Cluster Size

最大簇大小分布

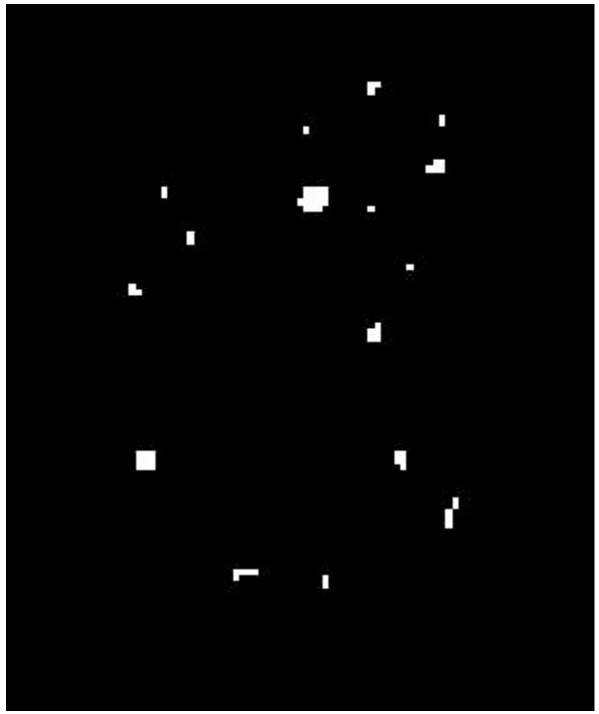
So, just as was the case for the t-values, we now have a distribution f that allows us to calculate a Family Wise threshold u pertaining to cluster size.

因此，就像t值一样，我们现在有一个分布f它使我们能够计算与簇大小有关的总体的阈值 u。



f depends crucially on the initial “cluster-forming” threshold?

f关键取决于初始“簇形成”的阈值?

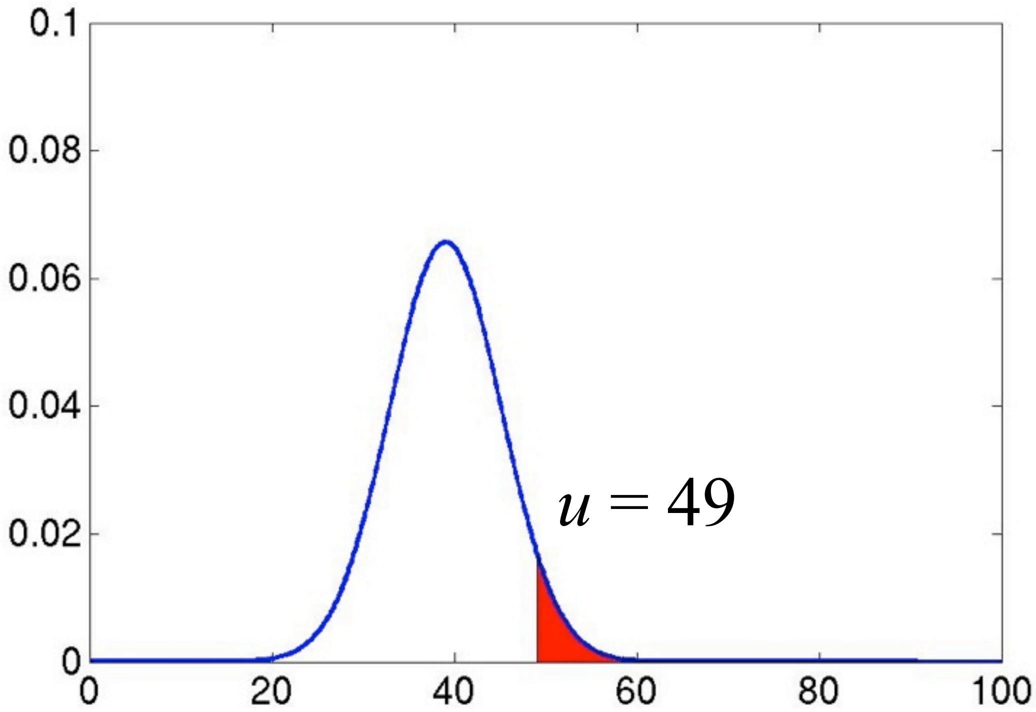


$z = 2.3$

Distribution of Max Cluster Size
最大簇大小分布

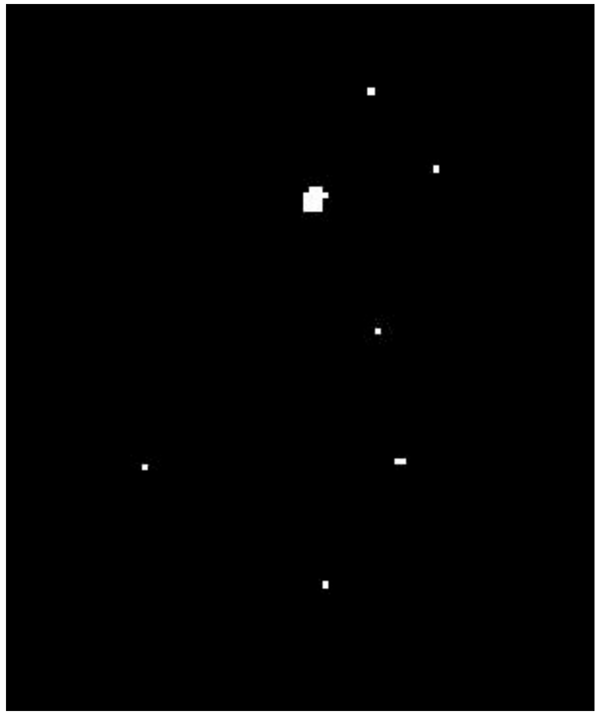
So, just as was the case for the t-values, we now have a distribution f that allows us to calculate a Family Wise threshold u pertaining to cluster size.

因此，就像t值一样，我们现在有一个分布f它使我们能够计算与簇大小有关的总体的阈值 u。



f depends crucially on the initial “cluster-forming” threshold?

f关键取决于初始“簇形成”的阈值?



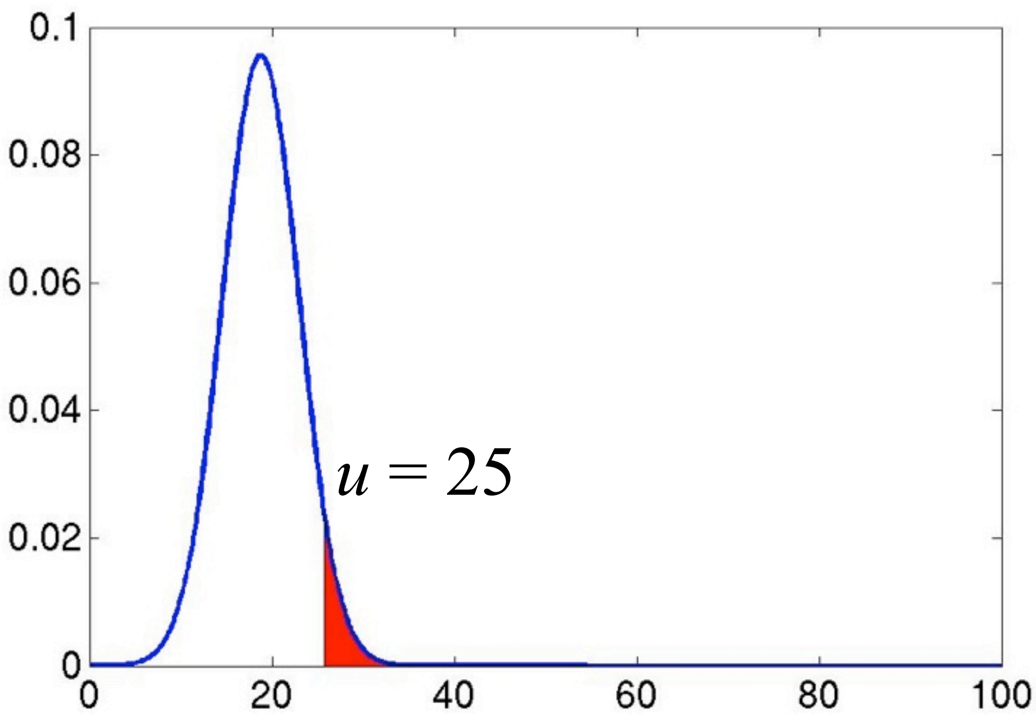
$z = 2.7$

Distribution of Max Cluster Size

最大簇大小分布

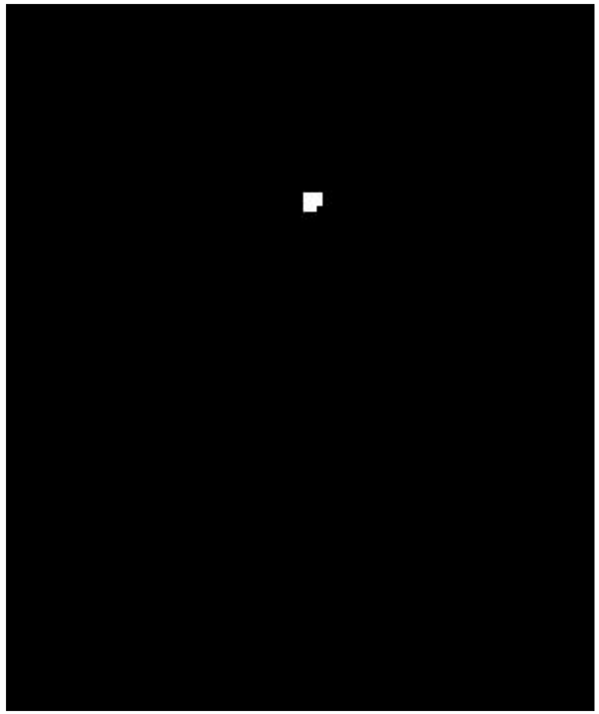
So, just as was the case for the t-values, we now have a distribution f that allows us to calculate a Family Wise threshold u pertaining to cluster size.

因此，就像t值一样，我们现在有一个分布f它使我们能够计算与簇大小有关的总体的阈值 u。



f depends crucially on the initial “cluster-forming” threshold?

f关键取决于初始“簇形成”的阈值?



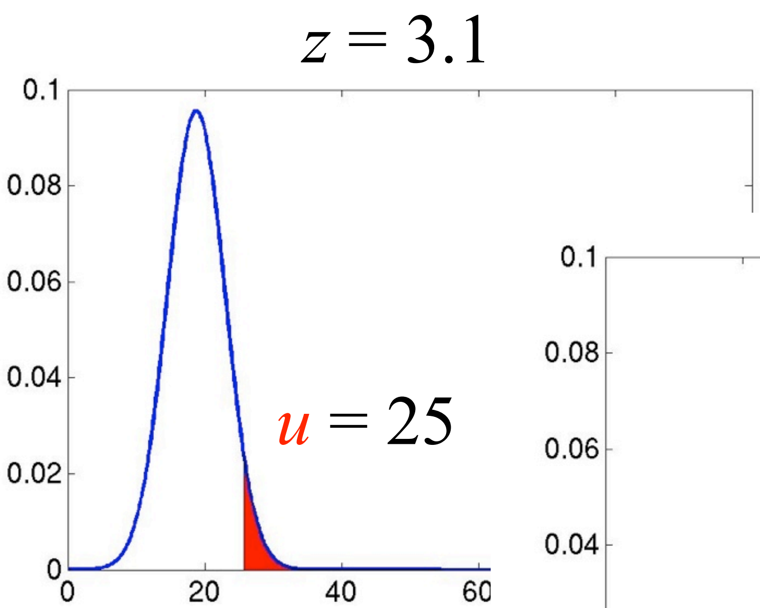
$z = 3.1$

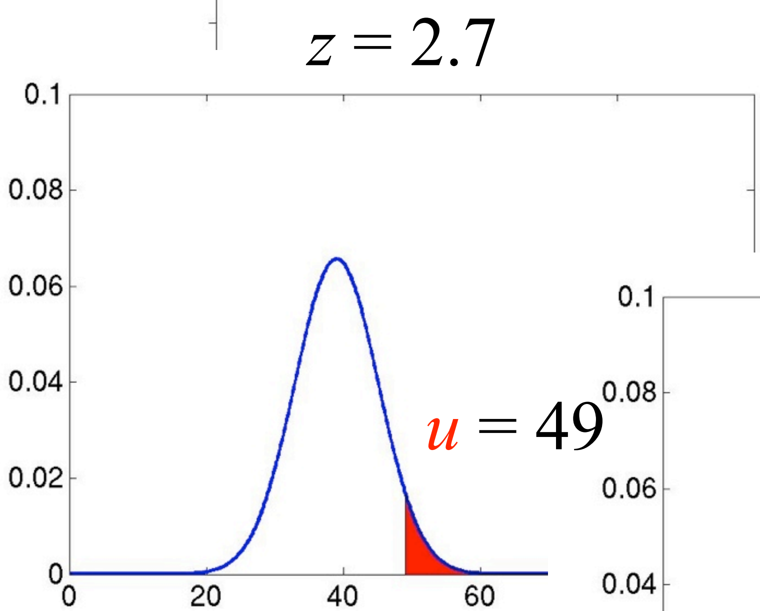
Distribution of Max Cluster Size

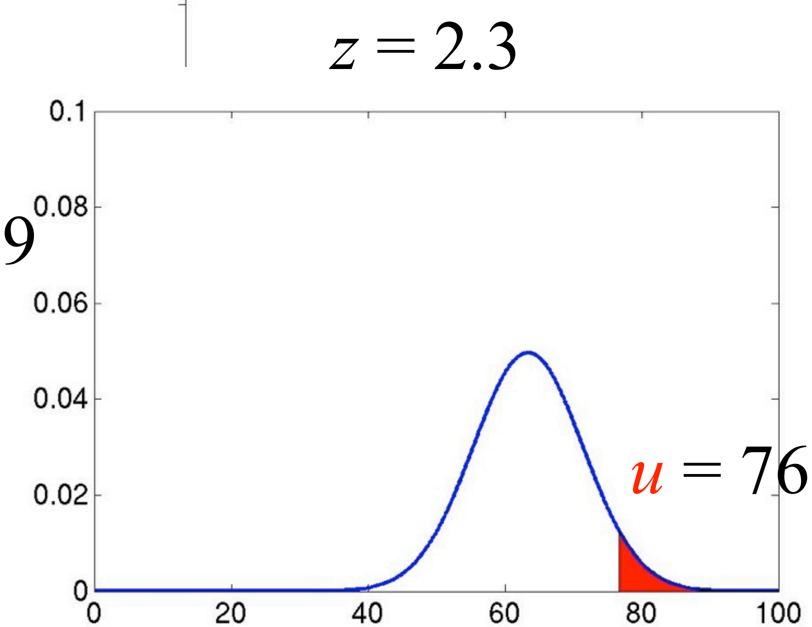
最大簇大小分布

Hence the distribution for the cluster size should really be written $f(z)$ and the same for $u(z)$

因此簇大小的分布应该写成$f(z)$，同样的$u(z)$也是。







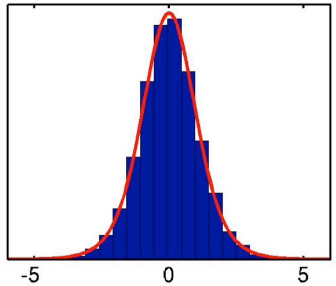
But as before we don’t have an expression for these distributions.

就像之前一样，我们不知道这些分布的表达式。

Outline大纲

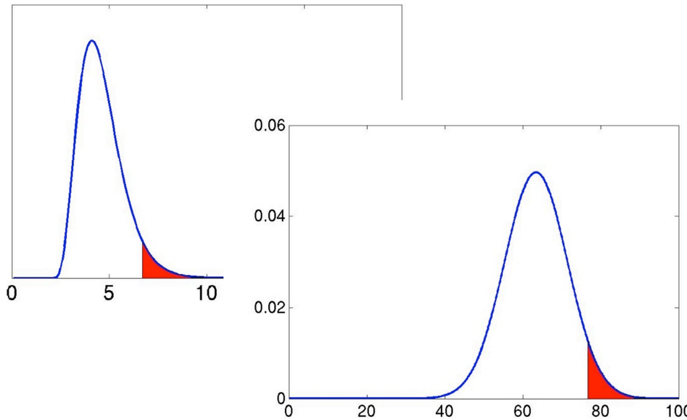
- Null-hypothesis and Null-distribution 零假设和零分布
- Multiple comparisons and Family-wise error 多重比较和族错误率
- Different ways of being surprised 惊奇的不同方式
  - Voxel-wise inference (Maximum z) 体素推断（最大z）
  - Cluster-wise inference (Maximum size) 簇水平推断（最大尺寸）
- Parametric vs non-parametric tests 参数vs非参检验
- Enhanced clusters 增强的簇
- FDR - False Discovery Rate FDR-错误发现率

Parametric vs non-parametric
参数对非参



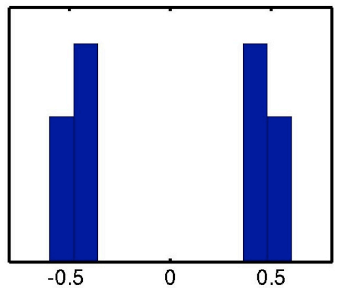
Provided that **e** ~ $N(0,\sigma^2)$

- As we described earlier, one of the great things about for example the t-test is that we know the null-distribution
  之前所说，t检验最大的一点优点是我们知道零分布。



- But most distributions are not that simple
  但大多数分布不是如此简单

- And errors are not always normal-distributed
  误差并不总是正态分布

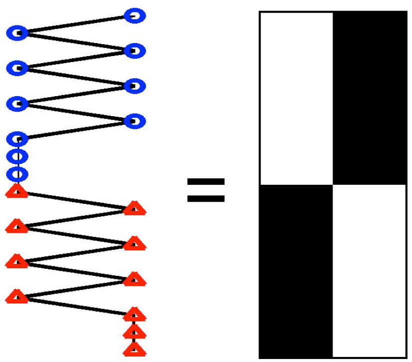


Example: VBM-style analysis
例如VBM类型的分析

- Our data is segmented grey matter maps 数据是分割好的灰质图
- A voxel is either grey matter, or not. 一个体素是或不是灰质

Group #1
(Oxford students)
组1 牛津学生

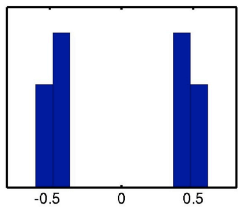
Group #2
(Train spotters)
组2 训练好的观察者



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

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

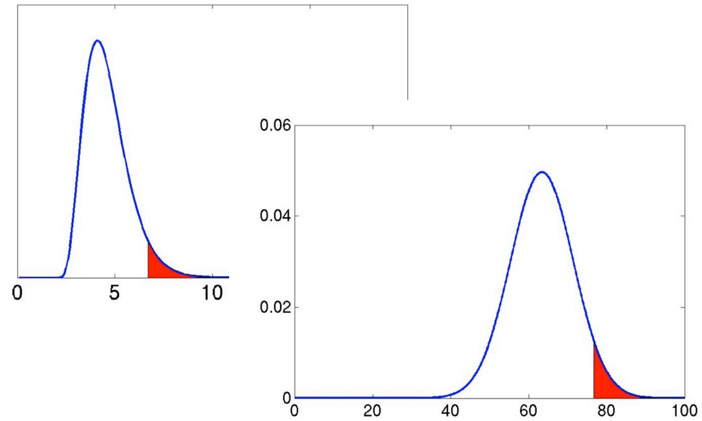
hist(e)



$\sim N$? ☹

Parametric vs non-parametric
参数对非参

- There are <u>approximations</u> to the Max-z and Max-size statistics
  Max-z和Max-size统计信息近似



- These are valid under certain sets of assumptions
  这些在某些假设下有效

- 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
  在另外一些情况下可能会有问题

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

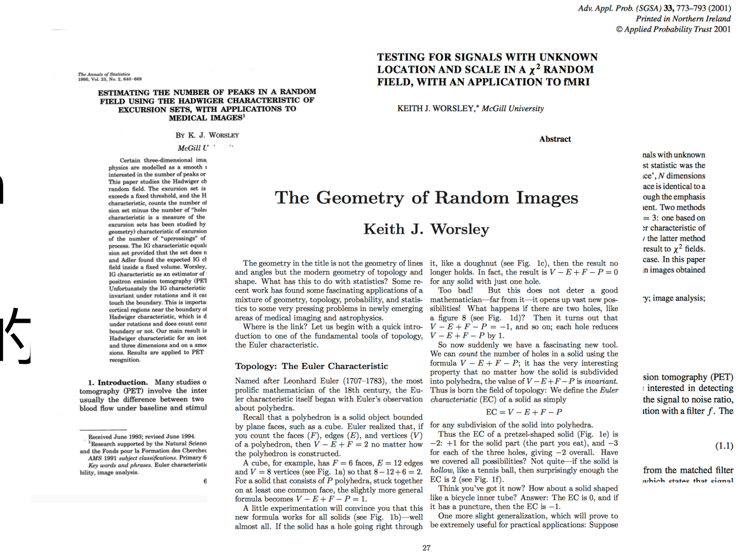
Edited by Emery N. Brown, Massachusetts General Hospital, Boston, MA, and approved May 17, 2016 (received for review February 12, 2016)

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
参数对非参

- Those approximations were based on Gaussian Random Field Theory, and was an impressive body of work
近似基于高斯随机场理论，令人印象深刻的工作


Adv. Appl. Prob. (SGSA) 33, 773–793 (2001)
Printed in Northern Ireland
© Applied Probability Trust 2001

ESTIMATING THE NUMBER OF PEAKS IN A RANDOM FIELD USING THE HADWIGER CHARACTERISTIC OF EXCURSION SETS, WITH APPLICATIONS TO MEDICAL IMAGES

By K. J. Worsley

TESTING FOR SIGNALS WITH UNKNOWN LOCATION AND SCALE IN A χ^2 RANDOM FIELD, WITH AN APPLICATION TO fMRI

KEITH J. WORSLEY, McGill University

Abstract

The Geometry of Random Images

Keith J. Worsley

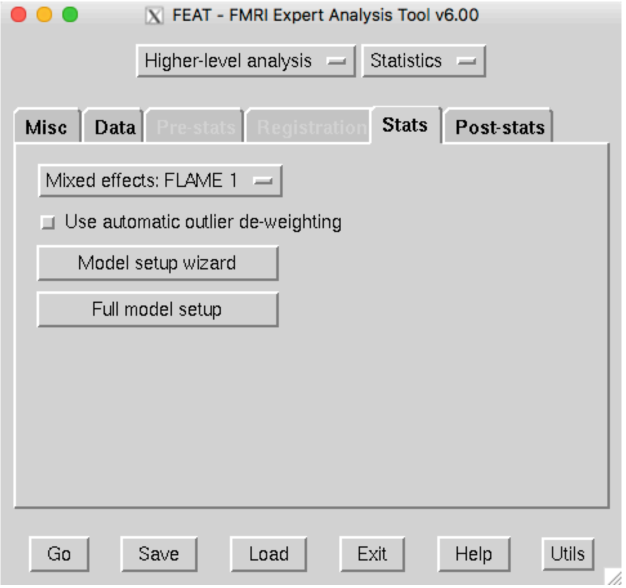
Topology: The Euler Characteristic

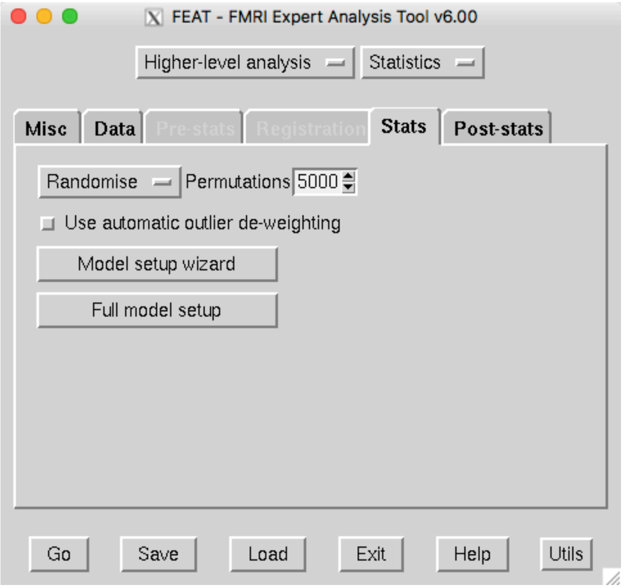
- They served us fantastically well at a time when we had little choice
在我们别无选择的时候，他们为我们提供了出色的服务

- But the future is non-parametric
但未来属于非参检验!


Parametric vs non-parametric

参数对非参





The Red (randomise) Baron

红色巨头（randomise工具）

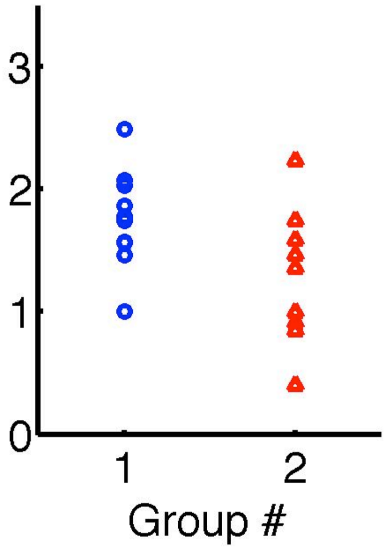
FLAME going down in flames

FLAME在硝烟中阵亡

A simple permutation test 简单的置换检验

- We can permute the data itself to create a distribution that we can use to test our statistic.
  我们可以对数据本身进行置换，以创建可用于测试统计信息的分布。
  - \+ Makes very few assumptions about the data 对数据做很少的假设
  - \+ Works for any test statistic 适应于任何统计

We have performed an experiment
我们做了个实验



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

$$t = 2.27$$

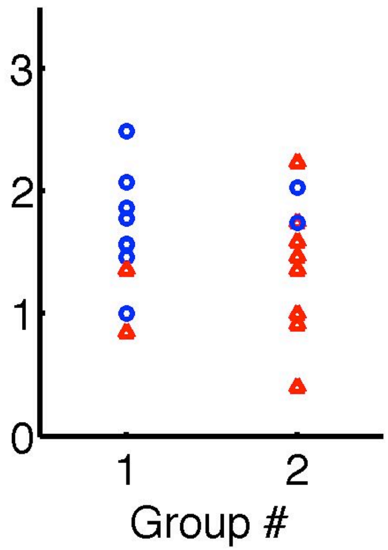
If the null-hypothesis is true, there is no difference between the groups. That means we should be able to "re-label" the individual points without changing anything.
如果零假设为真，则组间没有差异。 这意味着我们应该能够"重新标记"各个点，而不会造成任何改变。

A simple permutation test 简单的置换检验

- We can permute the data itself to create a distribution that we can use to test our statistic.
  我们可以对数据本身进行置换，以创建可用于测试统计信息的分布。

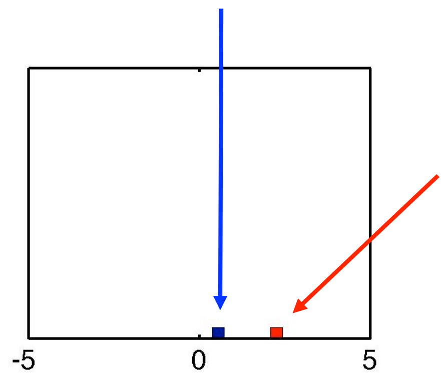
  + Makes very few assumptions about the data 对数据做很少的假设
  + Works for any test statistic 适应于任何统计

One re-labelling
第一次重标记



t-value after re-labelling
重标记后t值

$$t = 0.67$$



Original labelling
原始标记

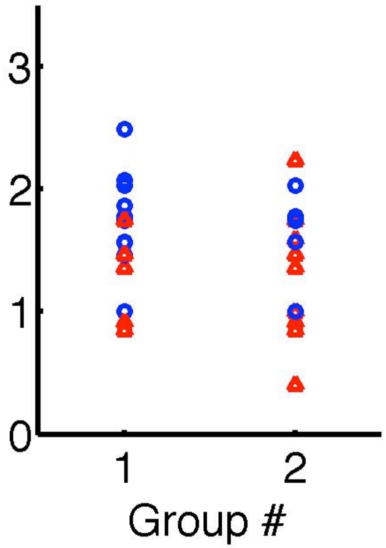
Let’s start collecting them开始收集他们

A simple permutation test 简单的置换检验

- We can permute the data itself to create a distribution that we can use to test our statistic.
  我们可以对数据本身进行置换，以创建可用于测试统计信息的分布。

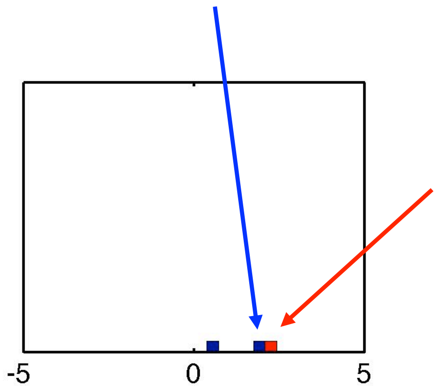
  + Makes very few assumptions about the data 对数据做很少的假设
  + Works for any test statistic 适应于任何统计

Second re-labelling
第二次重标记



t-value after re-labelling重
标记后的t值

$$t = 1.97$$



Original labelling
原始标记

And another one另一个

A simple permutation test 简单的置换检验

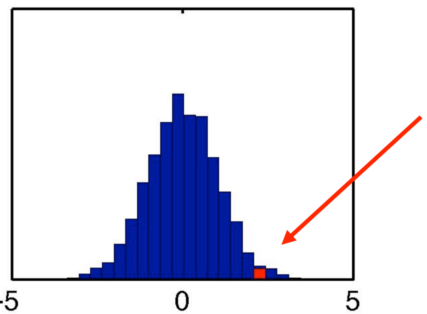
- We can permute the data itself to create a distribution that we can use to test our statistic.
  我们可以对数据本身进行置换，以创建可用于测试统计信息的分布。

  + Makes very few assumptions about the data 对数据做很少的假设

  + Works for any test statistic 适应于任何统计

Of the 5000 re-labellings, only 90 had a t-value > 2.27 (the original labelling).
在5000个重新标记中，只有90个的t值> 2.27（原始标记）。

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}
即如果各组之间没有差异，则只有1.8%（90/5000）的机会获得> 2.27的值。C.F. t18的p（x 2.27）= 1.79%



Original labelling
原始标记

5000 re-labellings. Phew!
5000次重标记，啧！

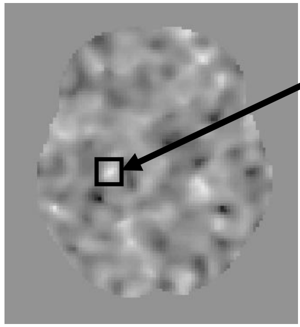
And we can use this for any statistic

我们可以对任何统计使用这个

This is what we got 这是我们得到的

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

我们比较了两组中每组5名受试者的疼痛刺激激活情况。



Very intriguing activation. $t_8 = 4.65$非常奇妙的激活

Prof. ran to write to Science.

But, did she jump the gun?

教授跑去给Science投稿了，但是他操之过急了吗？

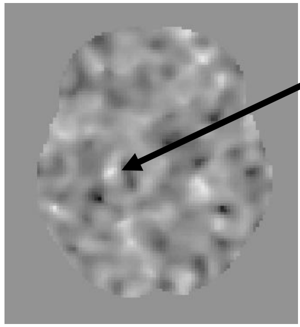
And we can use this for any statistic

我们可以对任何统计使用这个

This is what we got 这是我们得到的

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

我们比较了两组中每组5名受试者的疼痛刺激激活情况。



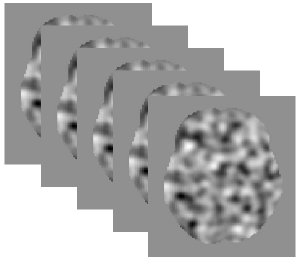
Very intriguing activation. $t_8 = 4.65$非常奇妙的激活

Prof. ran to write to Science.

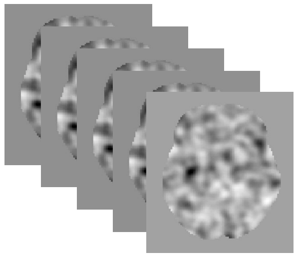
But, did she jump the gun?

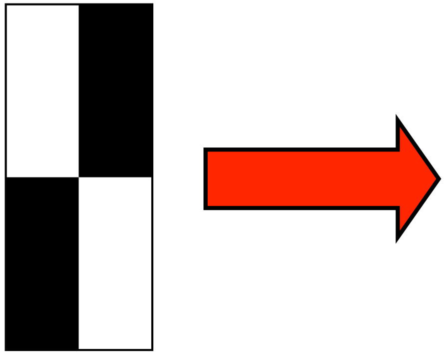
教授跑去给Science投稿了，但是他操之过急了吗?

Group 1



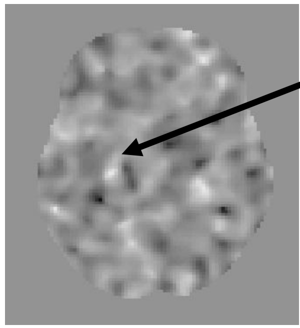
Group 2





2nd level model

第二水平模型



max(t)=4.65

Our group difference map

我们差异图

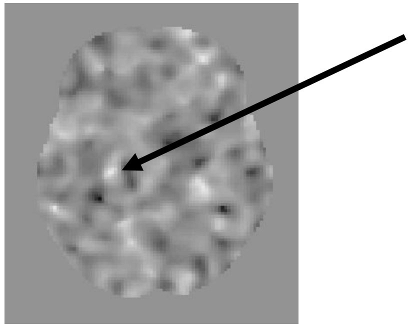
And we can use this for any statistic

我们可以对任何统计使用这个

This is what we got 这是我们得到的

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

我们比较了两组中每组5名受试者的疼痛刺激激活情况。

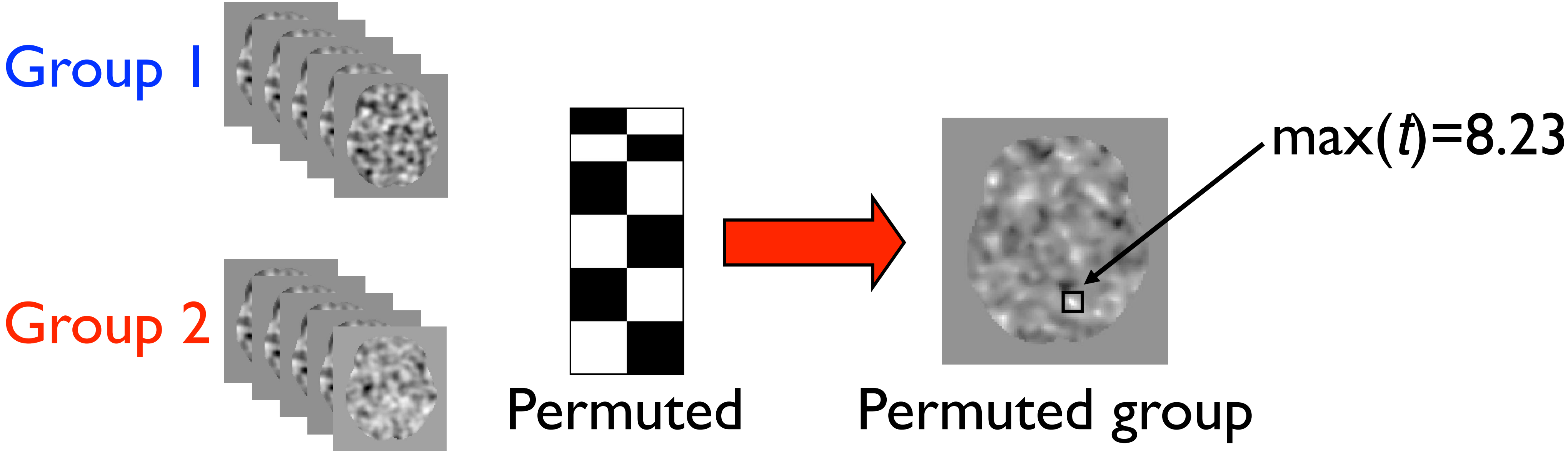


Very intriguing activation. $t_8 = 4.65$非常奇妙的激活

Prof. ran to write to Science.

But, did she jump the gun?

教授跑去给Science投稿了，但是他操之过急了吗?



Permuted model

置换模型

Permuted group difference map

置换组 差异图

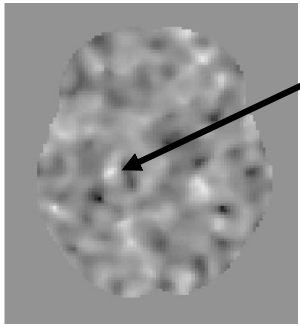
And we can use this for any statistic

我们可以对任何统计使用这个

This is what we got 这是我们得到的

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

我们比较了两组中每组5名受试者的疼痛刺激激激活情况。



Very intriguing activation. t_8 = 4.65非常奇妙的激活

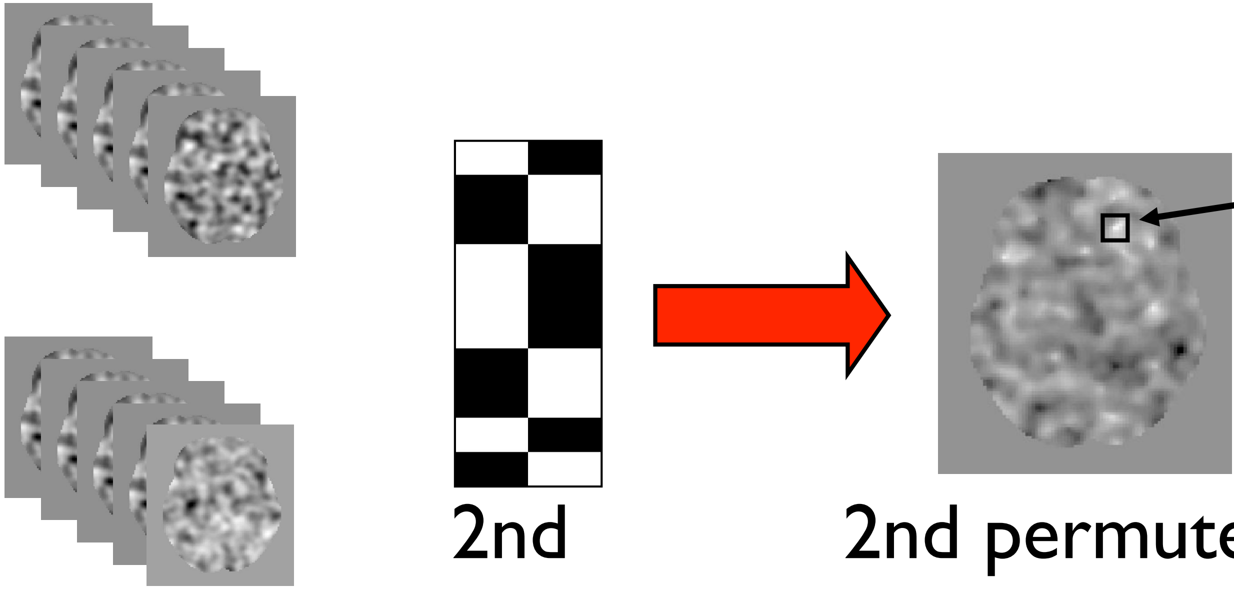
Prof. ran to write to Science.

But, did she jump the gun?

教授跑去给Science投稿了，但是他操之过急了吗?

Group 1

Group 2



max(t)=5.43

2nd Permutation

第二次置换

2nd permuted map

第二次置换图

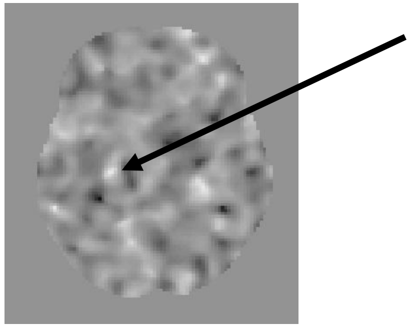
And we can use this for any statistic

我们可以对任何统计使用这个

This is what we got 这是我们得到的

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

我们比较了两组中每组5名受试者的疼痛刺激激活情况。



Very intriguing activation. t_8 = 4.65非常奇妙的激活

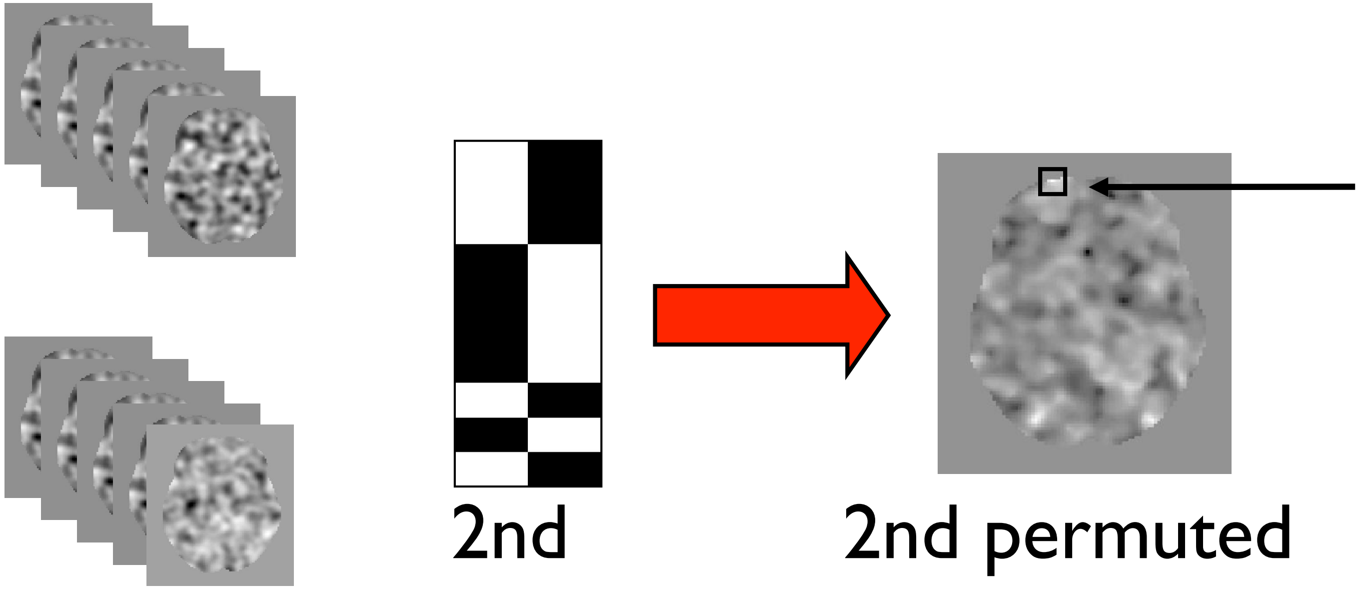
Prof. ran to write to Science.

But, did she jump the gun?

教授跑去给Science投稿了，但是他操之过急了吗?

Group 1

Group 2



max(t)=5.84

2nd Permutation

第三次置换

2nd permuted map

第三次置换图

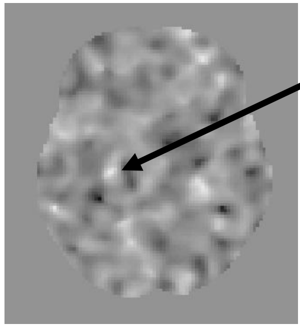
And we can use this for any statistic

我们可以对任何统计使用这个

This is what we got 这是我们得到的

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

我们比较了两组中每组5名受试者的疼痛刺激激活情况。



Very intriguing activation. t_8 = 4.65非常奇妙的激活

Prof. ran to write to Science.

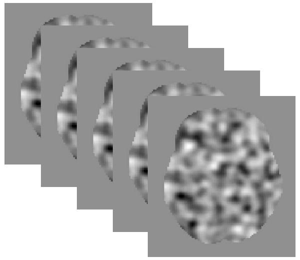
But, did she jump the gun?

教授跑去给Science投稿了，但是他操之过急了吗?

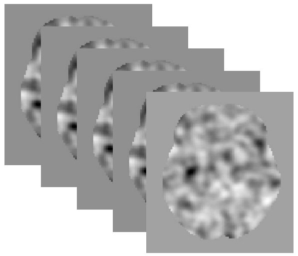
Original labelling

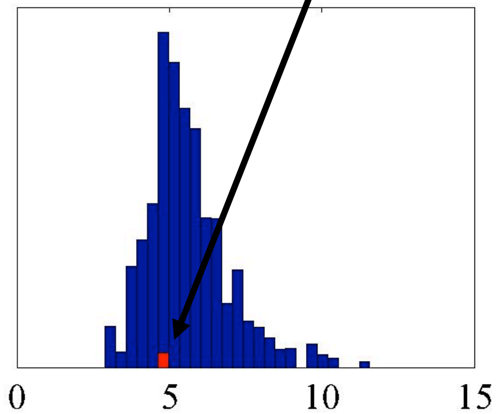
原始标记

Group 1



Group 2





5000 permutations

5000次置换

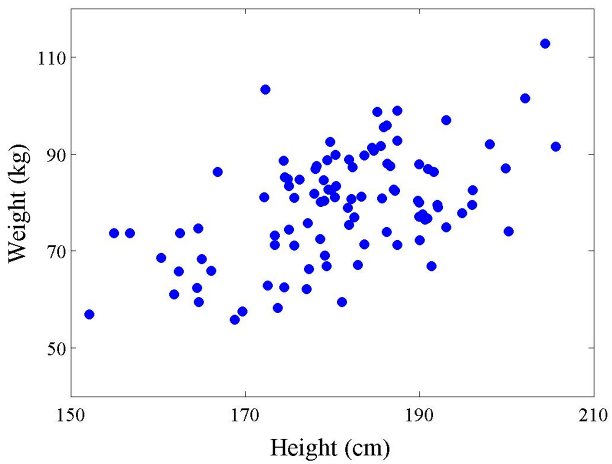
3925 permutations yielded higher max(t)-value than original labelling. We cannot reject the null-hypothesis.

3925次置换导致的最大t值比原始标记高。我不能拒绝零假设

But beware the “exchangeability”

但要小心“可交换性”

- When we swap the labels of two data-points we need to make sure that they are “exchangeable”
  当我们置换两个数据点的标记时，要确保他们可交换
- I will start to explain “exchangeability” through a case that is not
  我将通过一个不可交换的例子开始解释可交换性
- But first we need to learn about covariance matrices
  但首先我们得了解协方差矩阵



Height and weight of a random sample of Swedish men

瑞典男子随机样本的身高和体重

Covariance matrices

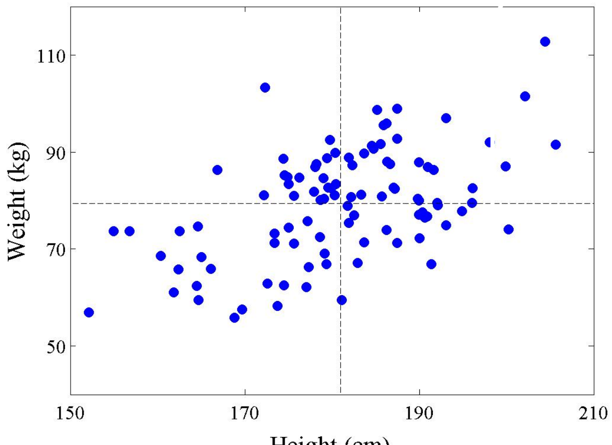
协方差矩阵

- When we swap the labels of two data-points we need to make sure that they are “exchangeable”
  当我们置换两个数据点的标记时，要确保他们可交换

- I will start to explain “exchangeability” through a case that is not
  我将通过一个不可交换的例子开始解释可交换性

- But first we need to learn about covariance matrices
  但首先我们得了解协方差矩阵

Mean height ≈181 cm　平均身高181cm



Mean weight ≈79.4 kg
平均体重79.4kg

Characterised by two means
用两个平均值表征

Covariance matrices

协方差矩阵

- When we swap the labels of two data-points we need to make sure that they are "exchangeable"

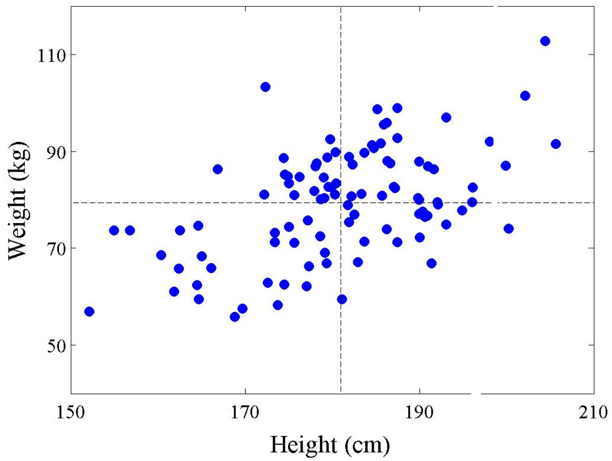
  当我们置换两个数据点的标记时，要确保他们可交换

- I will start to explain "exchangeability" through a case that is not

  我将通过一个不可交换的例子开始解释可交换性

- But first we need to learn about covariance matrices

  但首先我们得了解协方差矩阵



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

And a covariance - matrix

协方差矩阵

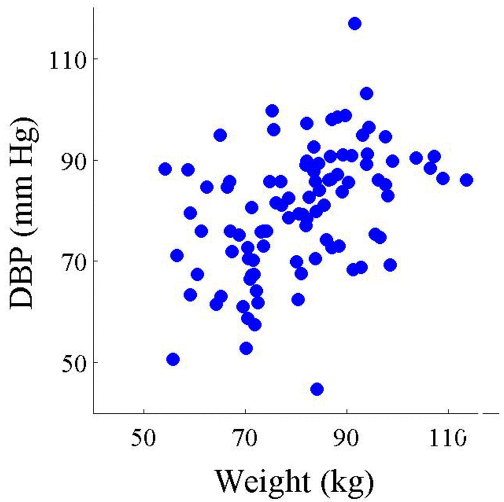
Covariance matrices

协方差矩阵

- When we swap the labels of two data-points we need to make sure that they are “exchangeable”
当我们置换两个数据点的标记时，要确保他们可交换

- I will start to explain “exchangeability” through a case that is not
我将通过一个不可交换的例子开始解释可交换性

- But first we need to learn about covariance matrices
但首先我们得了解协方差矩阵



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

Covariance matrices

协方差矩阵

- When we swap the labels of two data-points we need to make sure that they are "exchangeable"

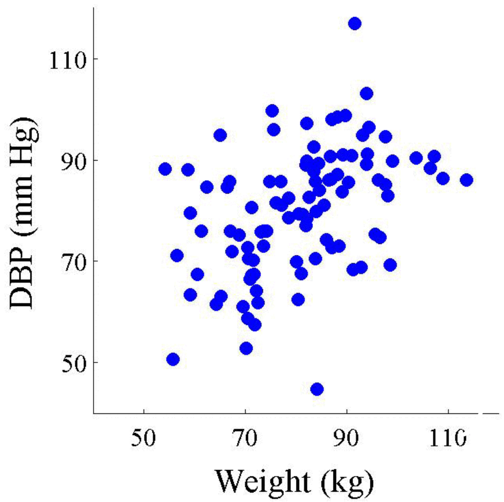
  当我们置换两个数据点的标记时，要确保他们可交换

- I will start to explain "exchangeability" through a case that is not

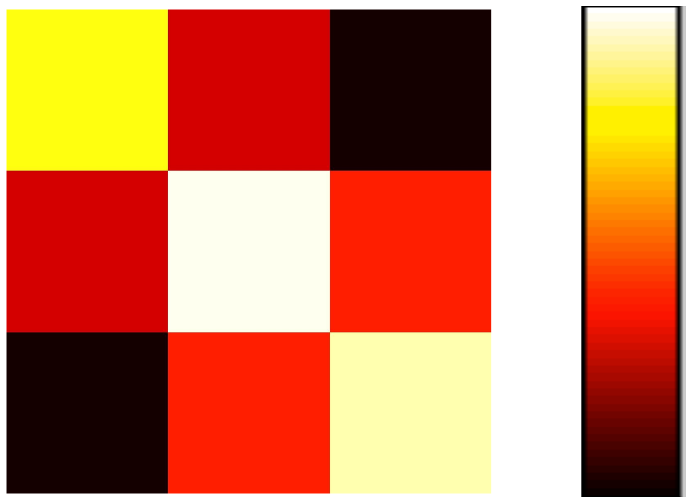
  我将通过一个不可交换的例子开始解释可交换性

- But first we need to learn about covariance matrices

  但首先我们得了解协方差矩阵



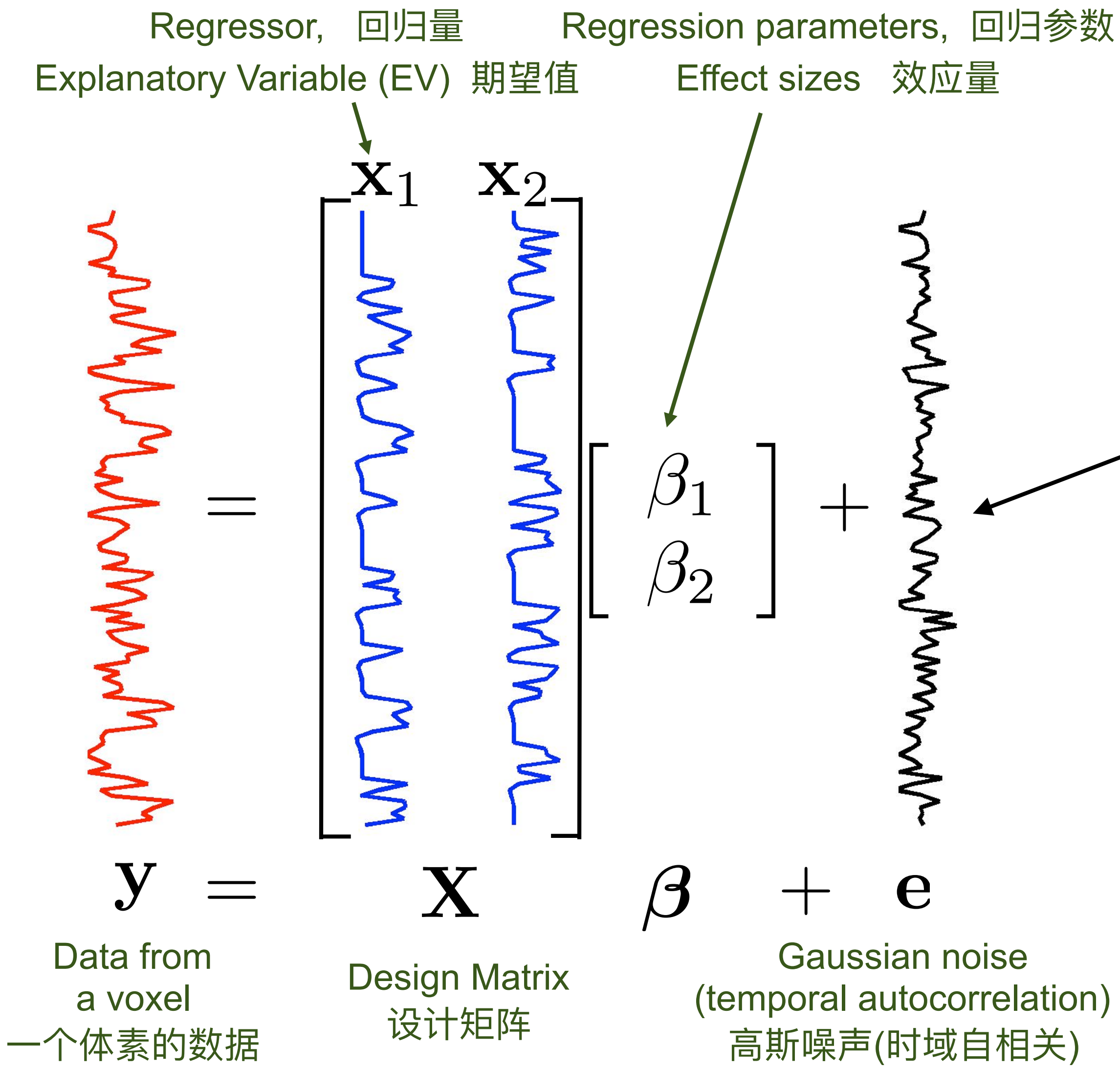
$$\Sigma =$$



1st level fMRI data is not exchangeable

第一阶的功能磁共振成像数据不可交换

- You may, or may not, have seen this slide in the 1st level GLM talk. 你或许在第一阶的GLM讨论中见过这个幻灯片



This time we will look more closely at this part

这次我们会仔细看这部分

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

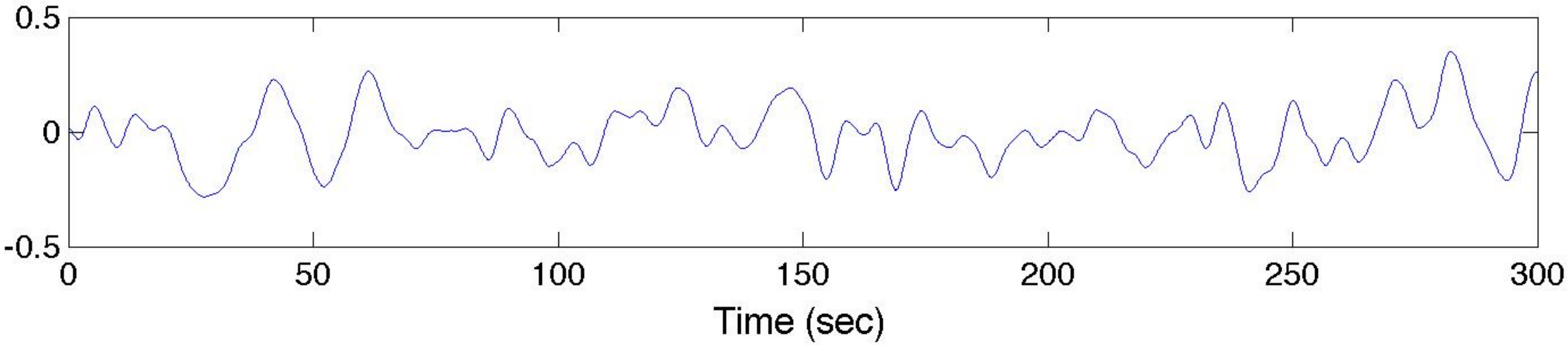
Our old friend "the covariance matrix"

我们的老朋友"协方差矩阵"

1st level fMRI data is not exchangeable

第一阶的功能磁共振成像数据不可交换

- One important component of noise in fMRI consists of physiological/neuronal events convolved by the HRF
  fmri中一个重要的噪声组分是由生理/神经时间与HRF的卷积造成的。

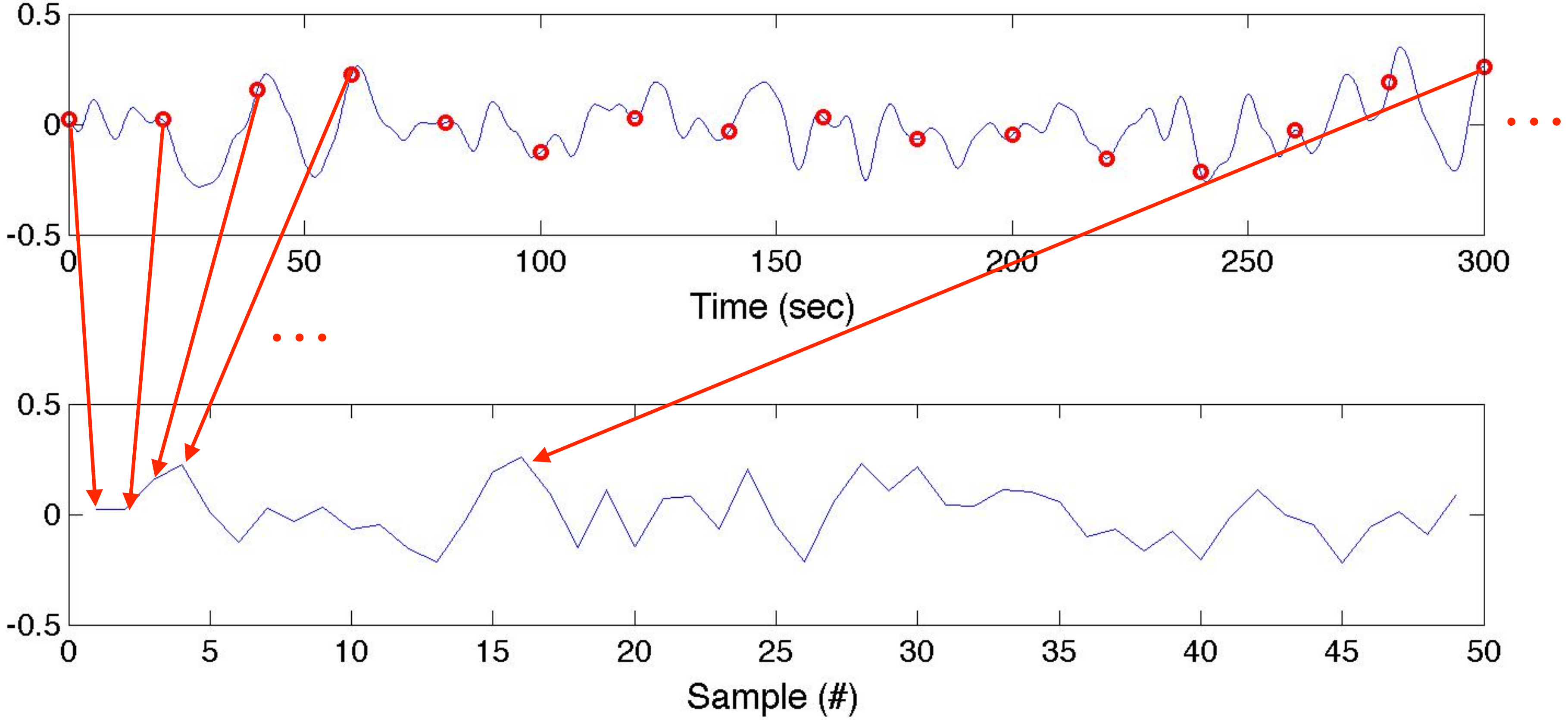


1st level fMRI data is not exchangeable

第一阶的功能磁共振成像数据不可交换

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

  fmri中一个重要的噪声组分是由生理/神经时间与HRF的卷积造成的。



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

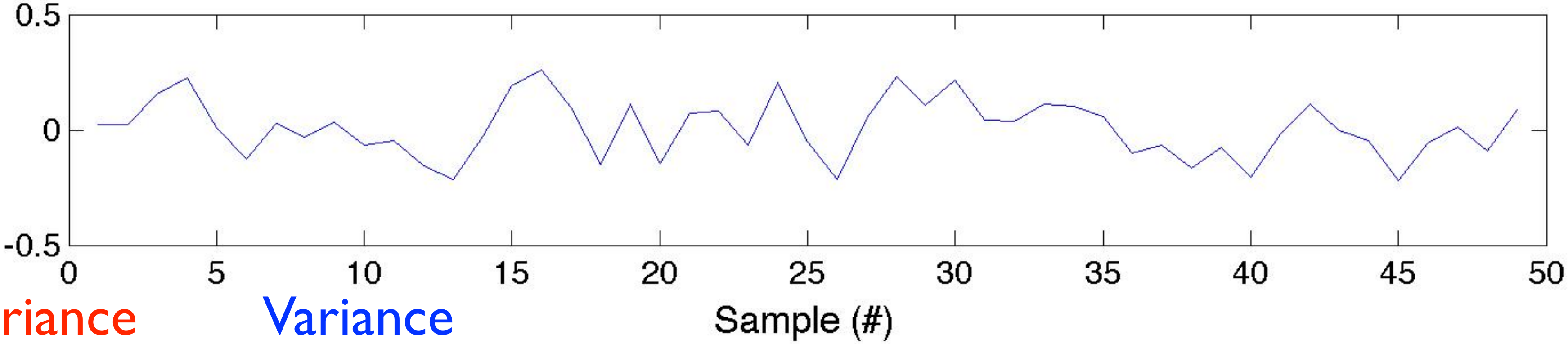
如果我们每个20s采样，他看起来不再“平滑”

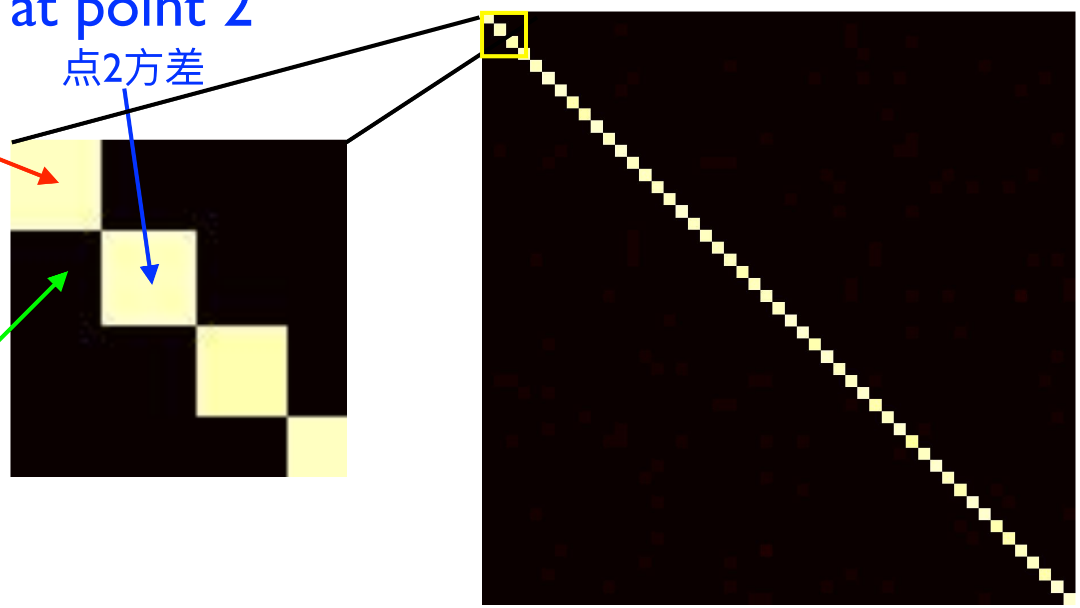
1st level fMRI data is not exchangeable

第一阶的功能磁共振成像数据不可交换

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

fmri中一个重要的噪声组分是由生理/神经时间与HRF的卷积造成的。





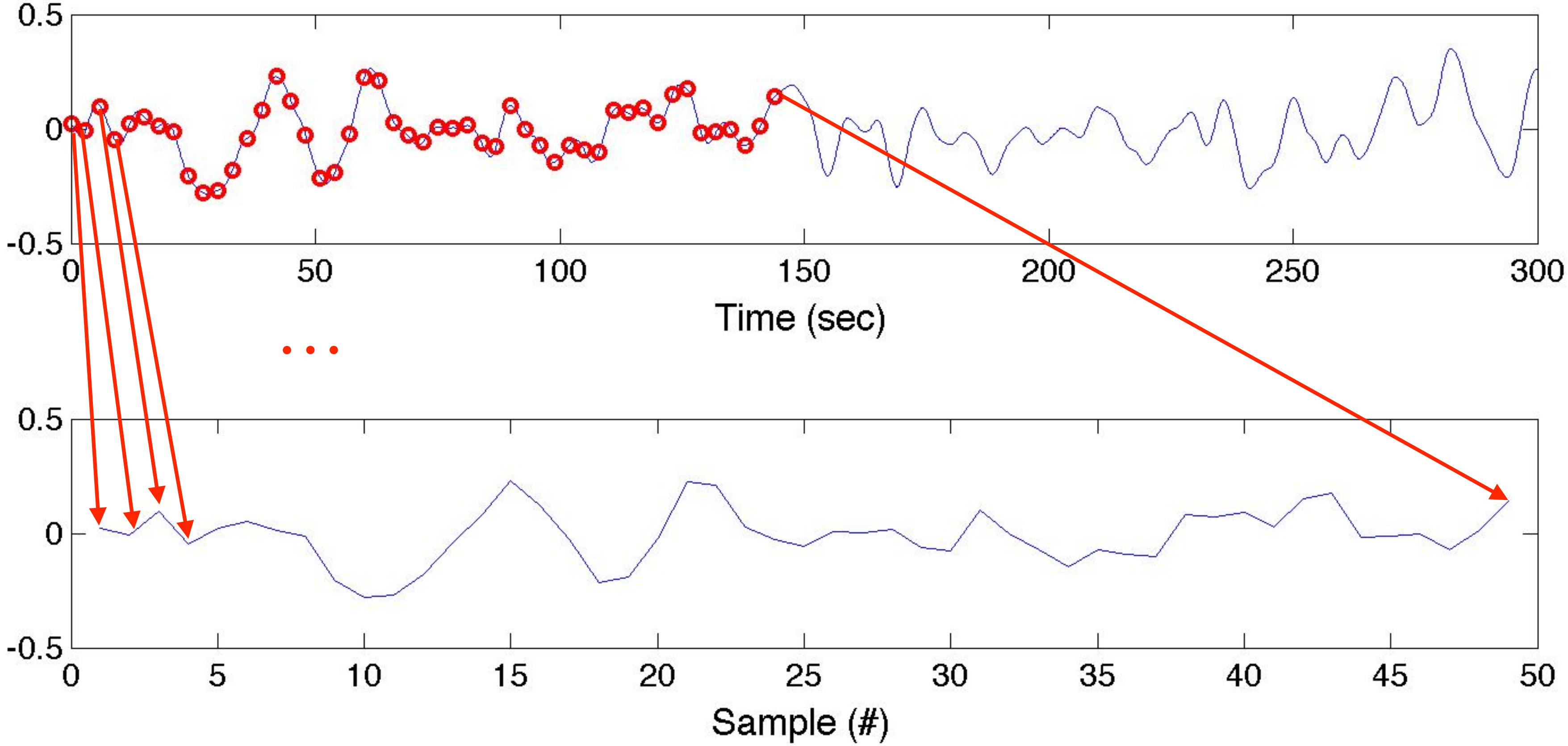
$$\mathbf{e} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$$

1st level fMRI data is not exchangeable

第一阶的功能磁共振成像数据不可交换

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

  fmri中一个重要的噪声组分是由生理/神经时间与HRF的卷积造成的。



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

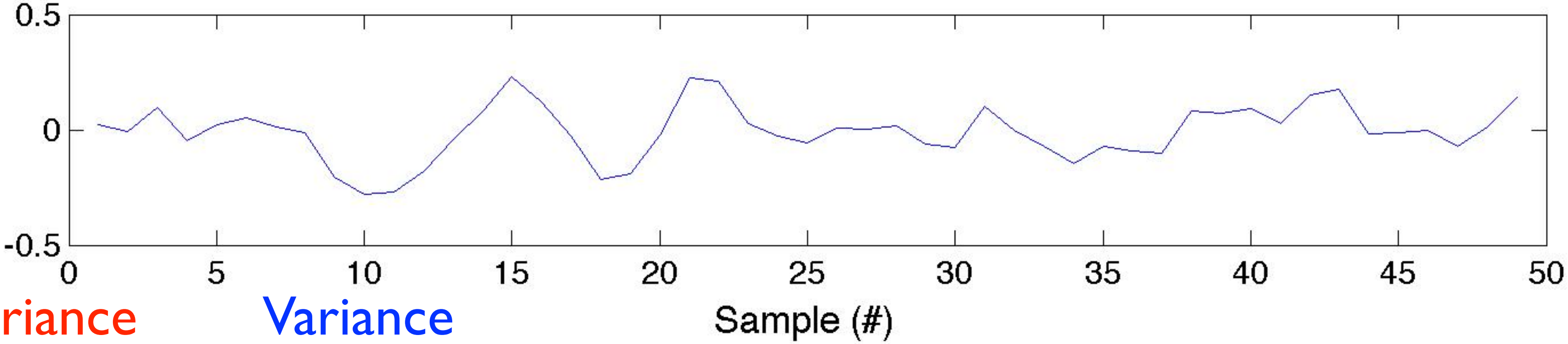
但这个TR不现实，3s咋样?

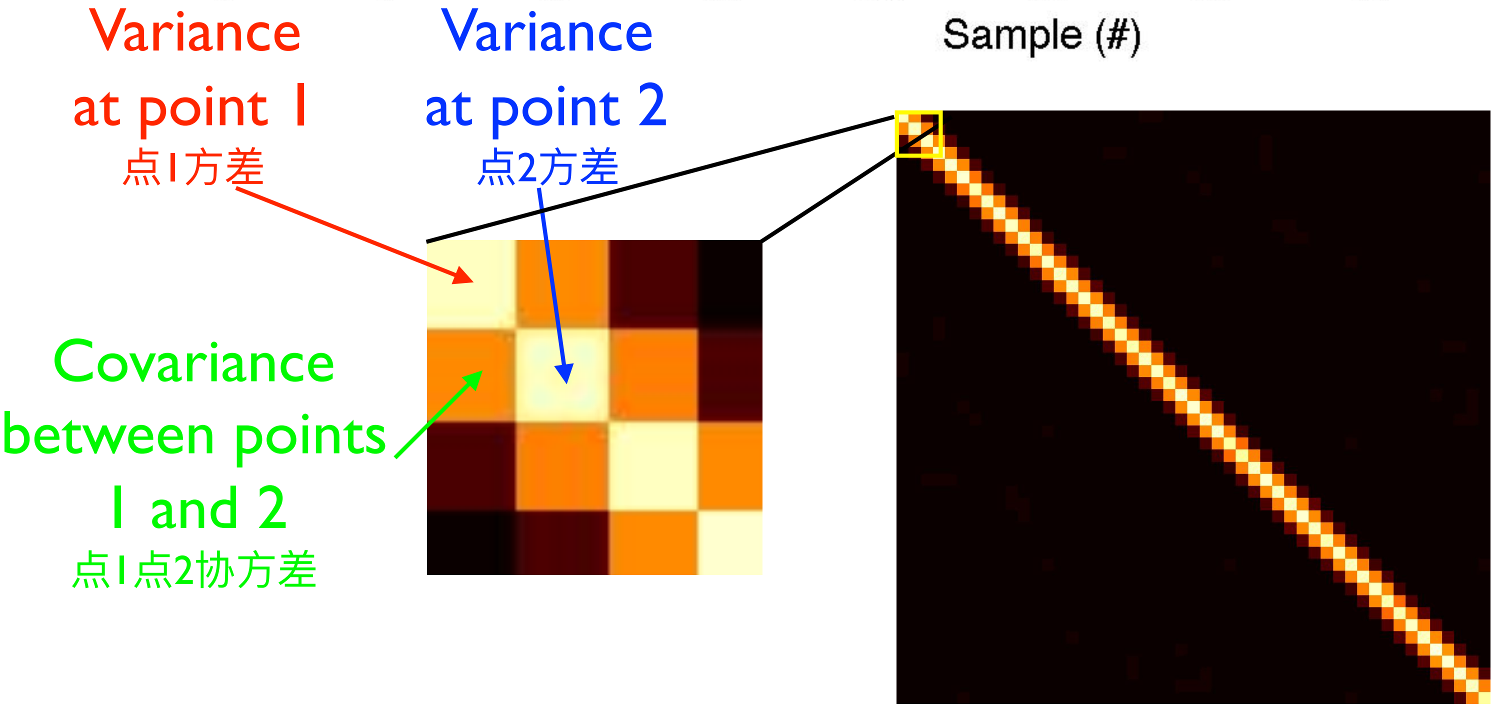
1st level fMRI data is not exchangeable

第一阶的功能磁共振成像数据不可交换

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

  fmri中一个重要的噪声组分是由生理/神经时间与HRF的卷积造成的。

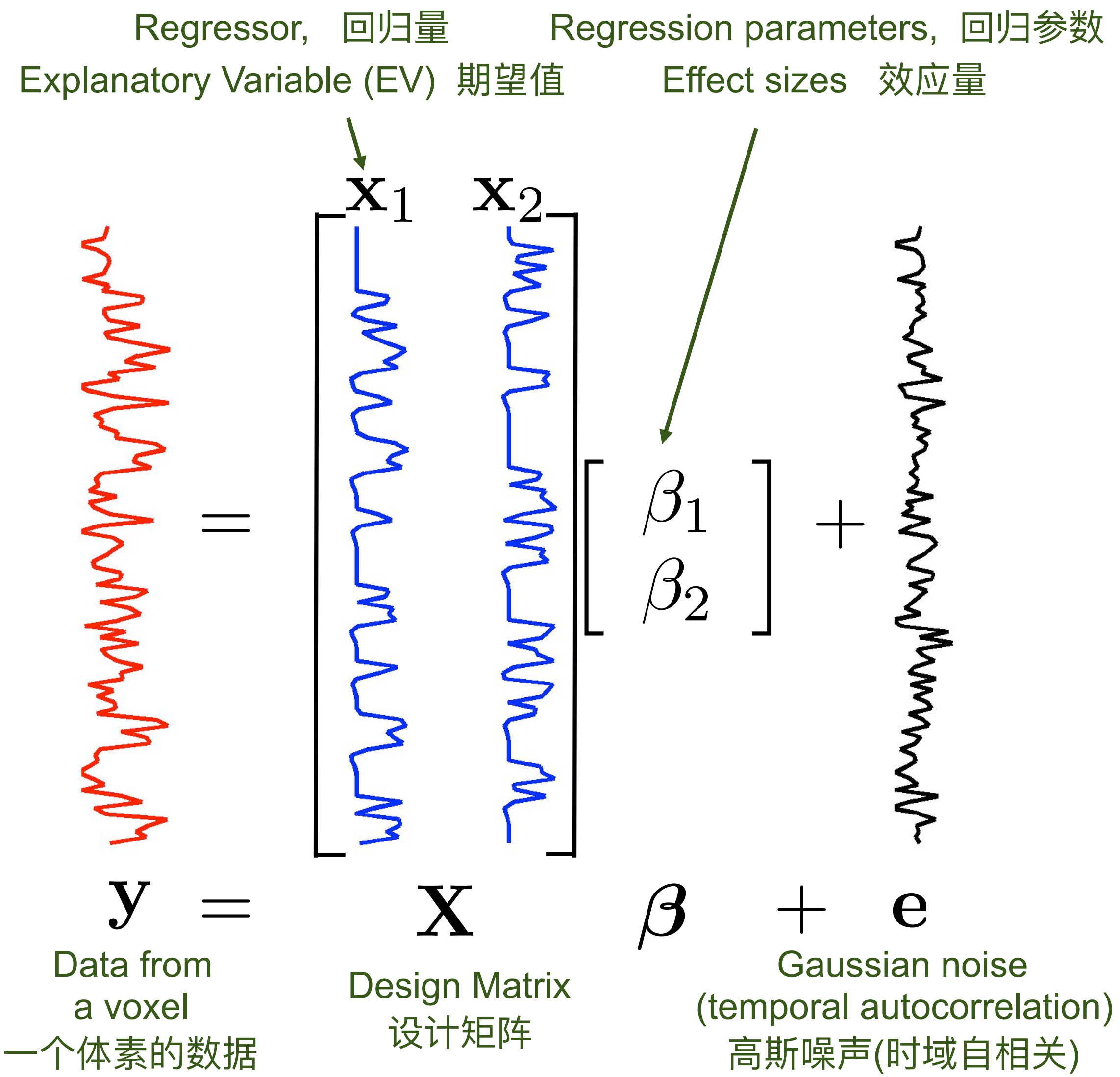




1st level fMRI data is not exchangeable

第一阶的功能磁共振成像数据不可交换

- Let us now return to our model again 我们再返回模型



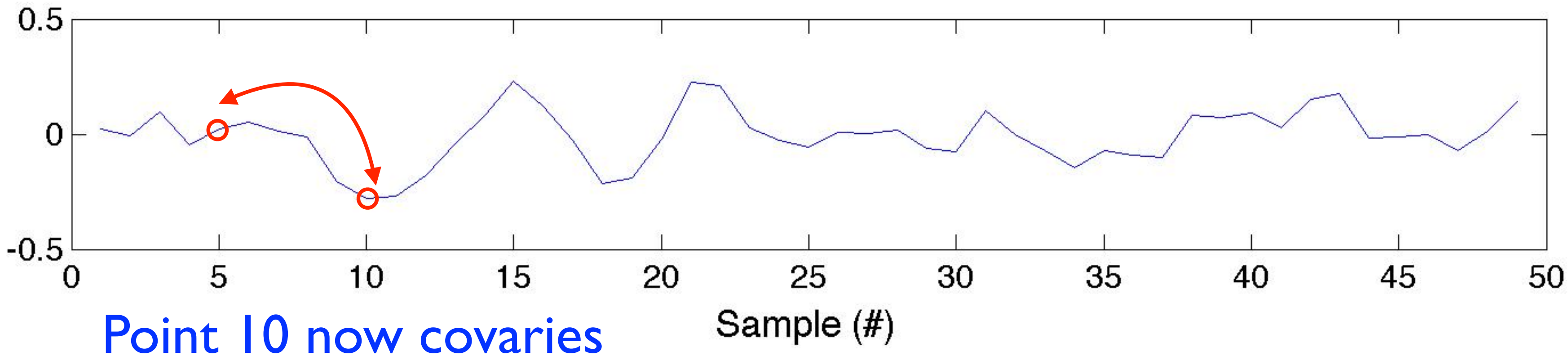
- The model consists of our regressors **X** and the noise model
  模型由回归量X和噪声组成
- All permutations must result in “equivalent models”
  所有的置换一定会造成“等效模型”
- Let us now see what happens if we swap two data-points(points 5 and 10)
  让我们看看如果交换两个数据点会发生什么。

1st level fMRI data is not exchangeable

第一阶的功能磁共振成像数据不可交换

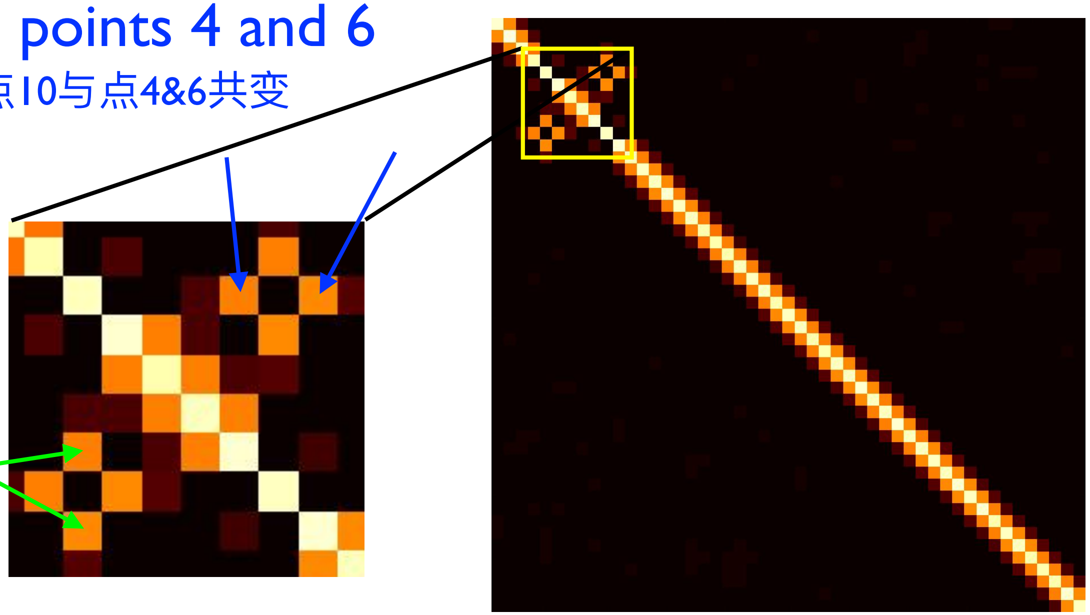
- One important component of noise in fMRI consists of physiological/neuronal events convolved by the HRF

  fmri中一个重要的噪声组分是由生理/神经时间与HRF的卷积造成的。



Point 10 now covaries with points 4 and 6

点10与点4&6共变



"Point 5" now covaries with points 9 and 11

点5和点9&11共变

And the models are no longer equivalent

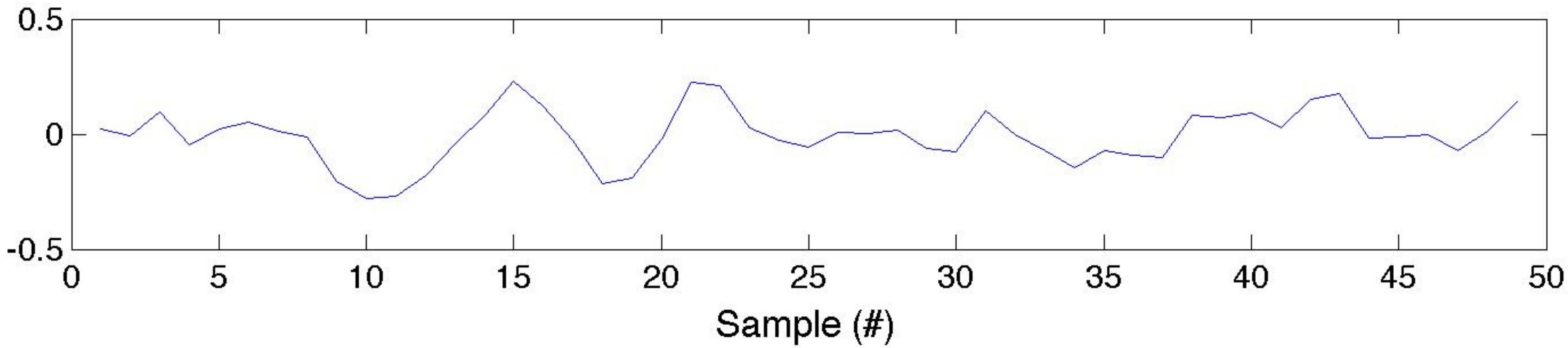
模型不再等价

1st level fMRI data is not exchangeable

第一阶的功能磁共振成像数据不可交换

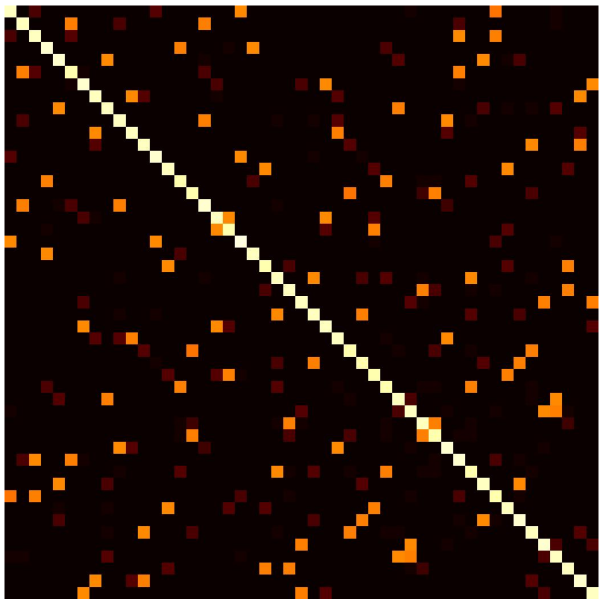
- One important component of noise in fMRI consists of physiological/neuronal events convolved by the HRF

fmri中一个重要的噪声组分是由生理/神经时间与HRF的卷积造成的。



And for a random permutation …

对于随机的置换…



And the models are no longer equivalent

模型也不再等价

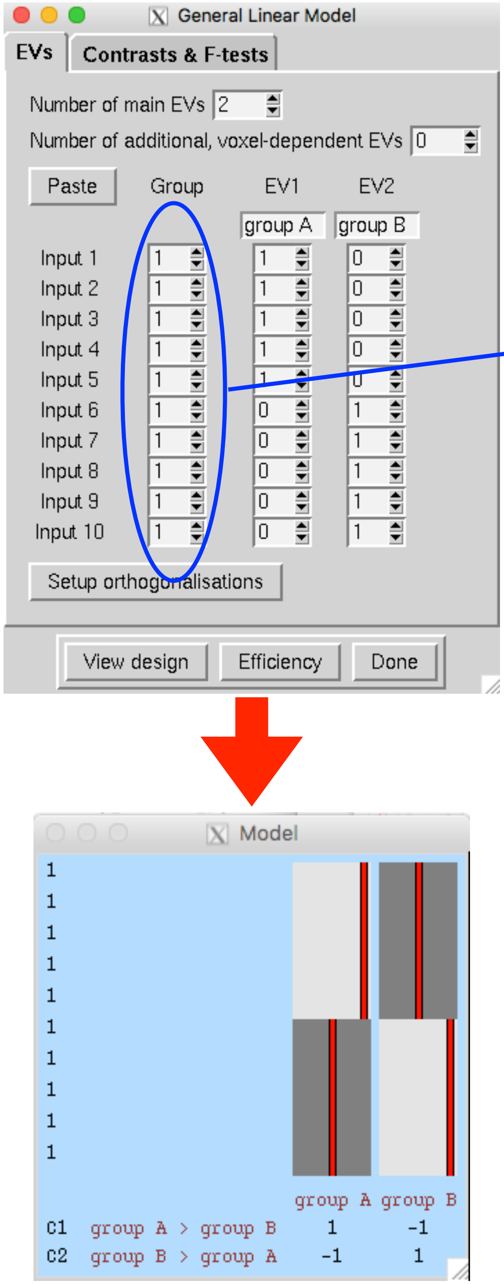
Back to exchangeability

回到可交换性

- Data-points are not "exchangeable" if swapping them means that the noise covariance-matrix ends up looking differently. 如果交换数据点造成了噪声协方差矩阵的差异，他们就不可交换。
- Formally "The joint distribution of the data must be unchanged by the permutations under the null-hypothesis". 在零假设下进行置换，数据的联合分布一定不变。
- If the noise covariance-matrix has non-zero off-diagonal elements (covariances) you need to beware.
  如果噪声协方差矩阵有非零的非对角线元素(协方差)，你要小心。
- You typically never estimate or see the covariance-matrix. You need to "imagine it" and determine from that if there is a problem.
  通常你不能估计或看到协方差矩阵，需要你去想象他并判断是否有问题。

Examples of exchangeability: 可交换例子：

Two groups unpaired 两组非配对



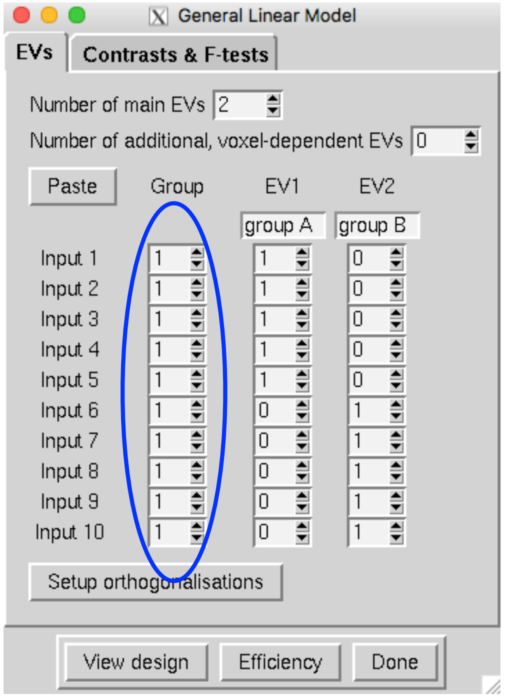
This is the "exchangeability group". Here all scans are in the same group, which means any scan can be exchanged for any other.

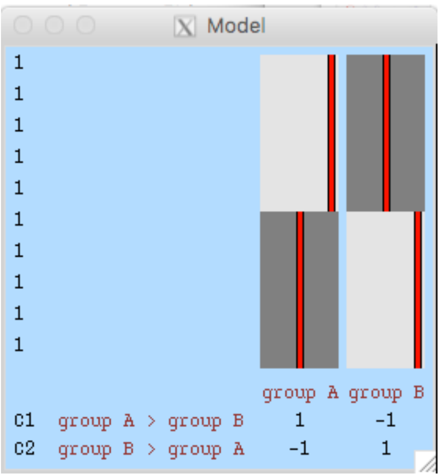
这是可交换组，所有扫面在同一组里，意味着任一扫描可以随意置换

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

注意：进行FLAME/GRFT时，"组"标记用于完全不同的目的。

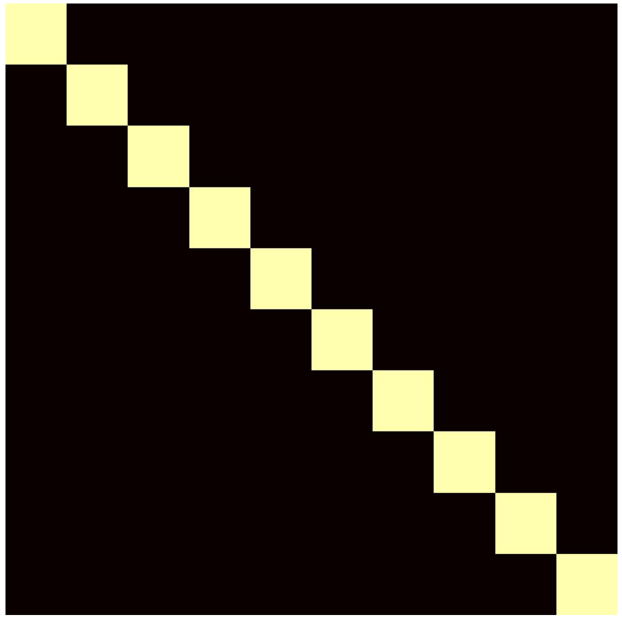
Examples of exchangeability: 可交换例子：
Two groups unpaired 两组非配对





Assumed covariance matrix

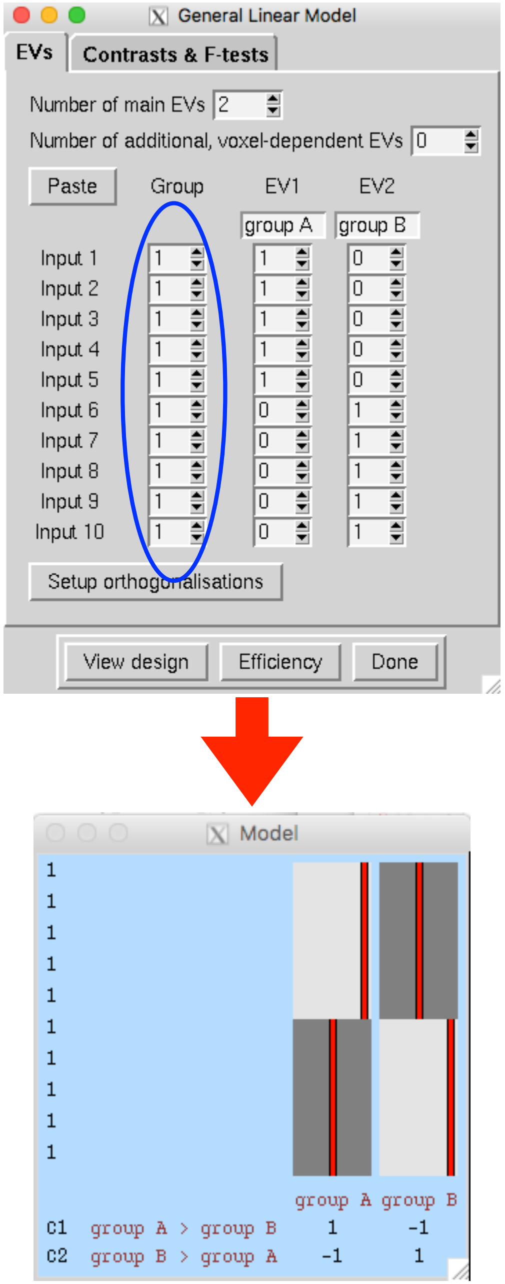
假设的协方差矩阵



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

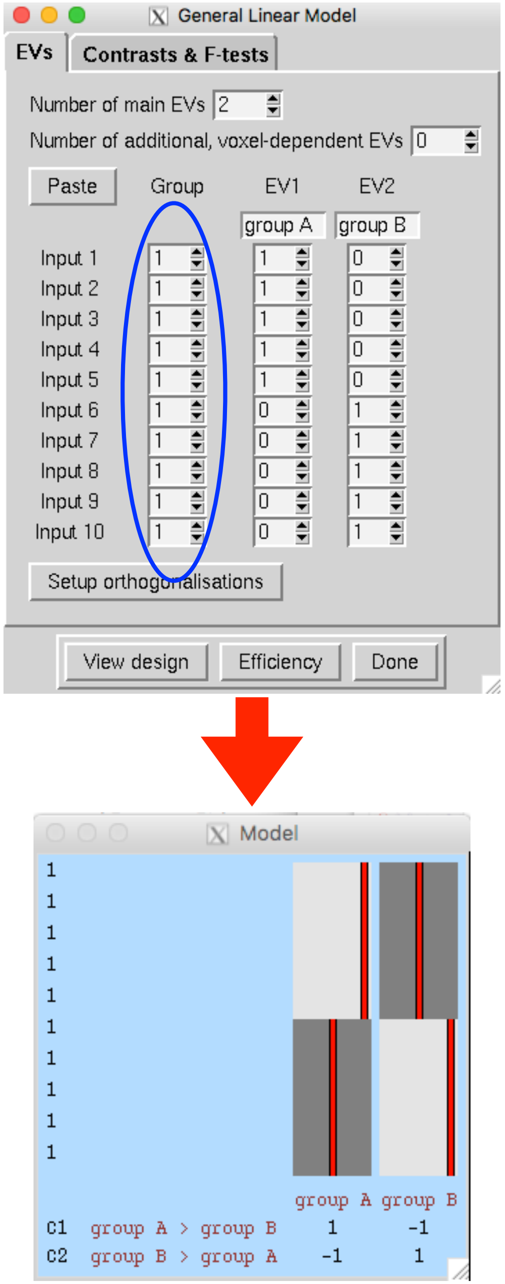
这隐含假设是来自所有被试的数据都具有相同不确定性并且都是独立的。

Examples of exchangeability: 可交换例子：
Two groups unpaired 两组非配对



Original 原始	Perm 1 置换1	Perm 2 置换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	

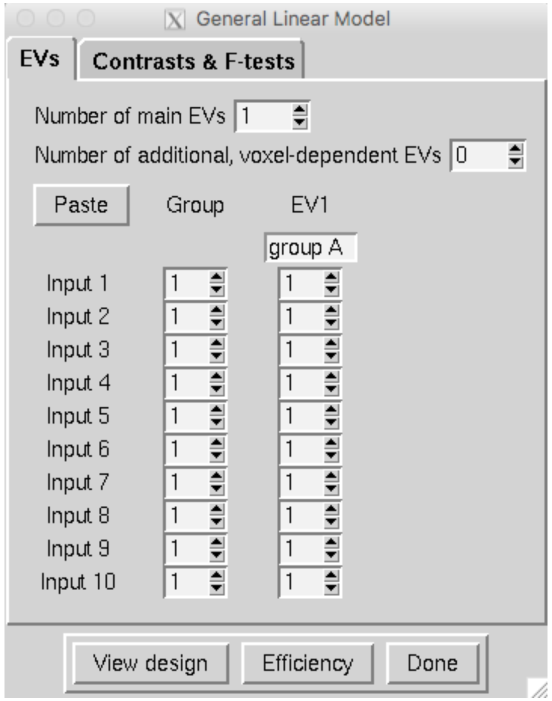
Examples of exchangeability: 可交换例子：
Two groups unpaired 两组非配对

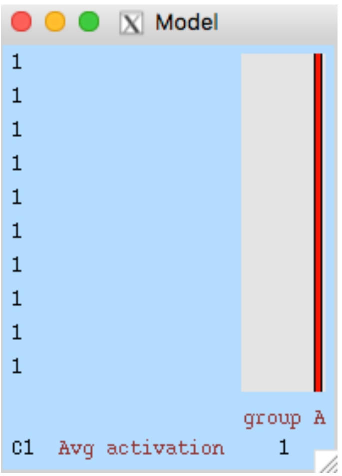


Original 原始	Perm 1 置换1	Perm 2 置换2	~~Perm 3~~ ~~置换3~~
1	6	6	1
2	3	1	7
3	7	7	6
4	8	4	9
5	5	9	4
6	1	5	2
7	2	8	8
8	4	3	5
9	9	10	3
10	10	2	10

N.B. Equivalent 等效的

Examples of exchangeability: 可交换例子：
Single group average 单组平均





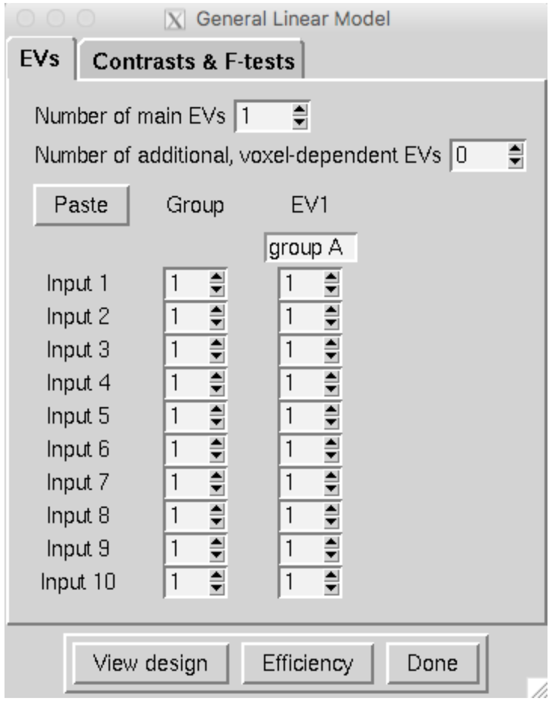
Here we model a single mean and want to know if that is different from zero

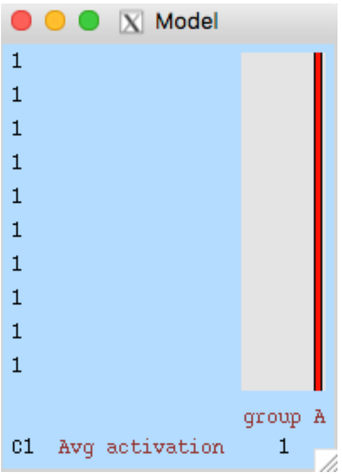
我们对单个平均值进行建模，看是否不同于零。

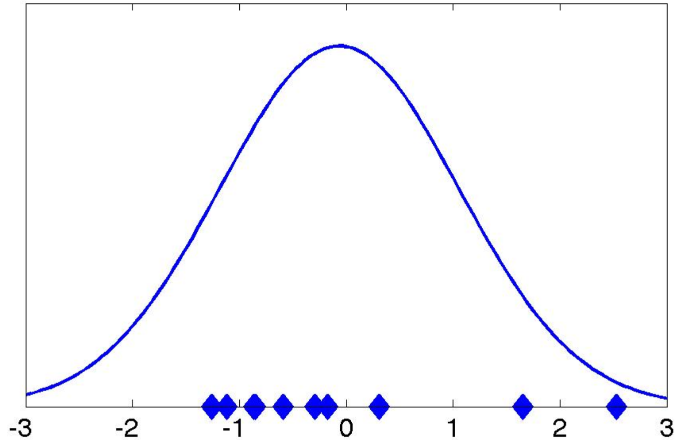
But there isn't really anything to permute, or is there?

但实际并没有置换任何东西，事实是这样吗?

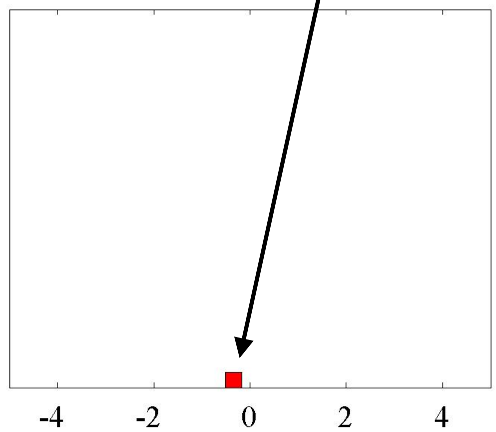
Examples of exchangeability: 可交换例子：
Single group average 单组平均



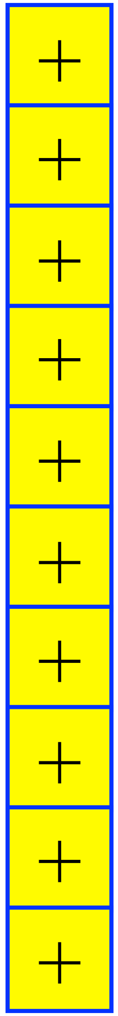




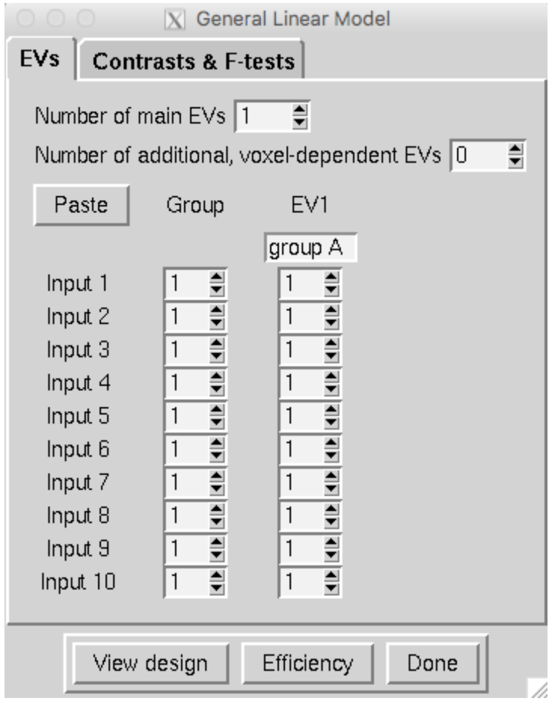
$$t = -0.17$$

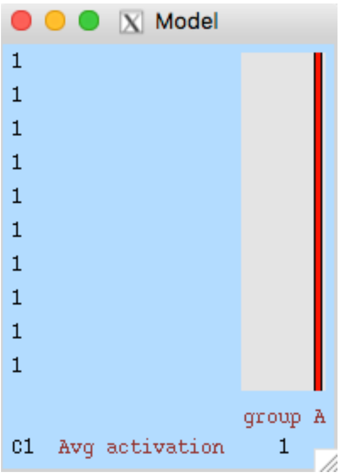


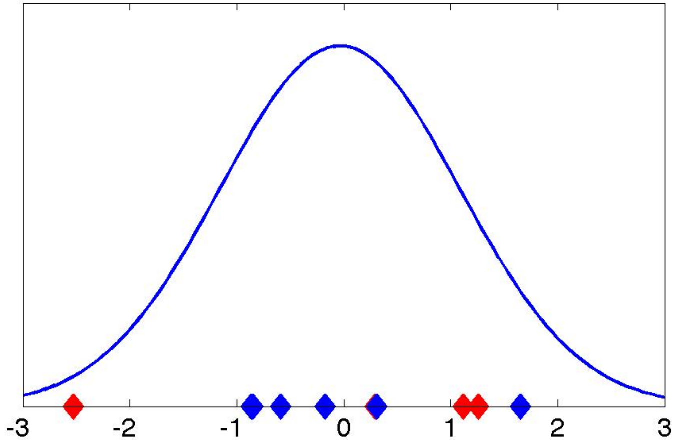
Original 原始



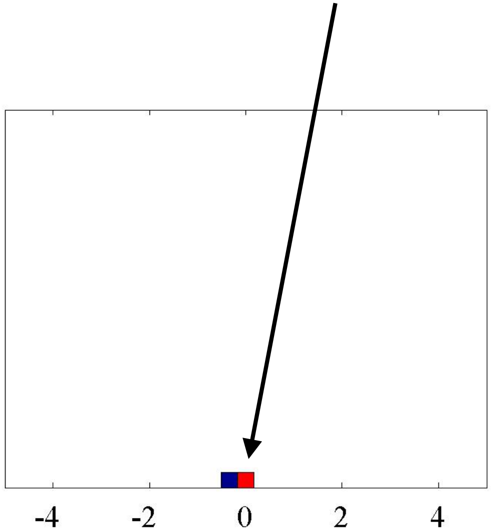
Examples of exchangeability: 可交换例子：
Single group average 单组平均





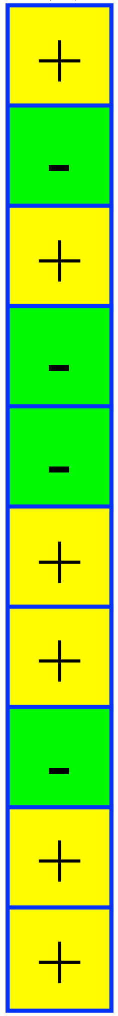


$$t = -0.09$$

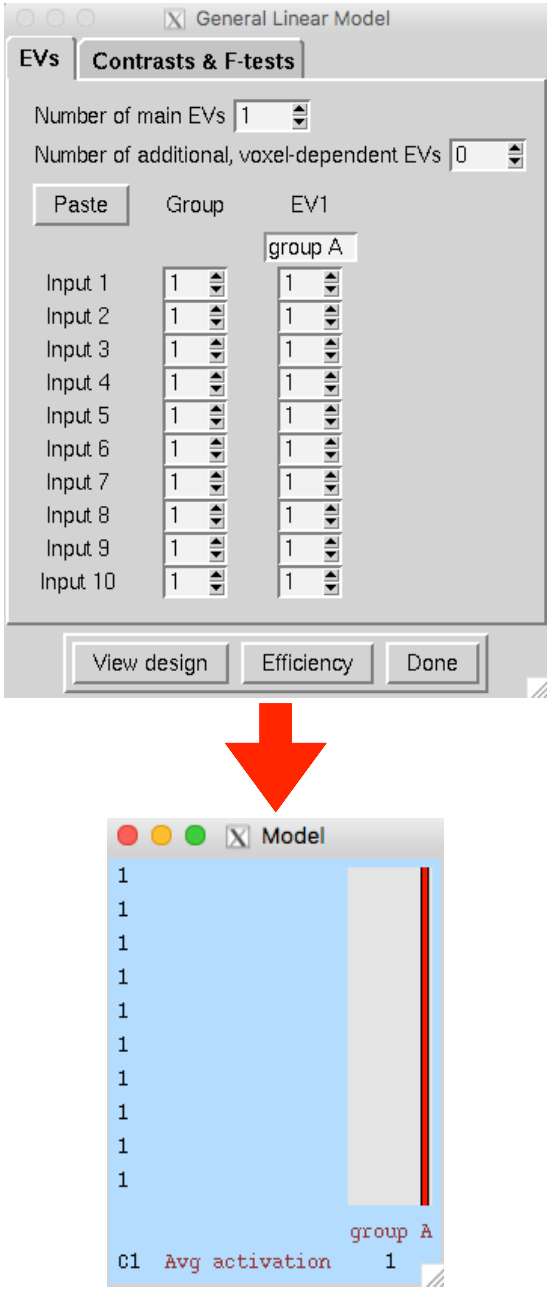


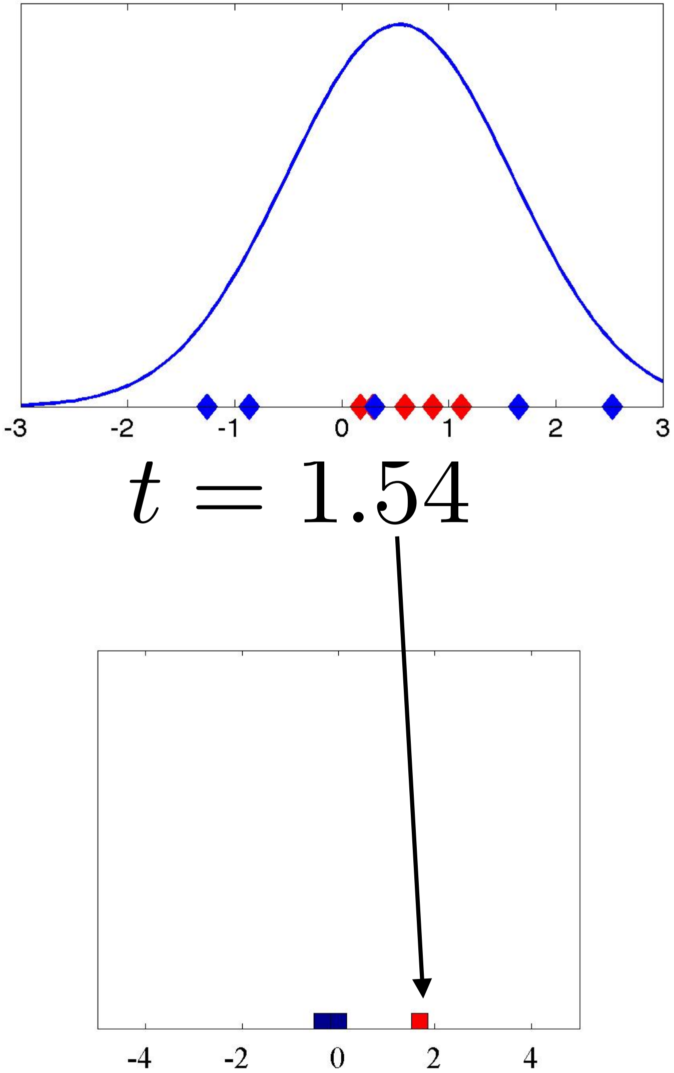
First flip

第一次置换

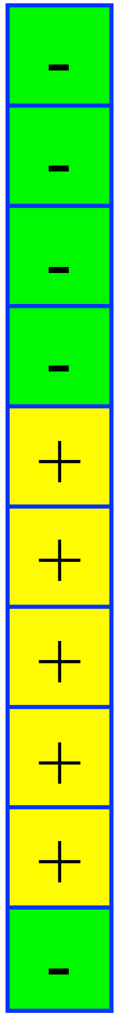


Examples of exchangeability: 可交换例子:
Single group average 单组平均

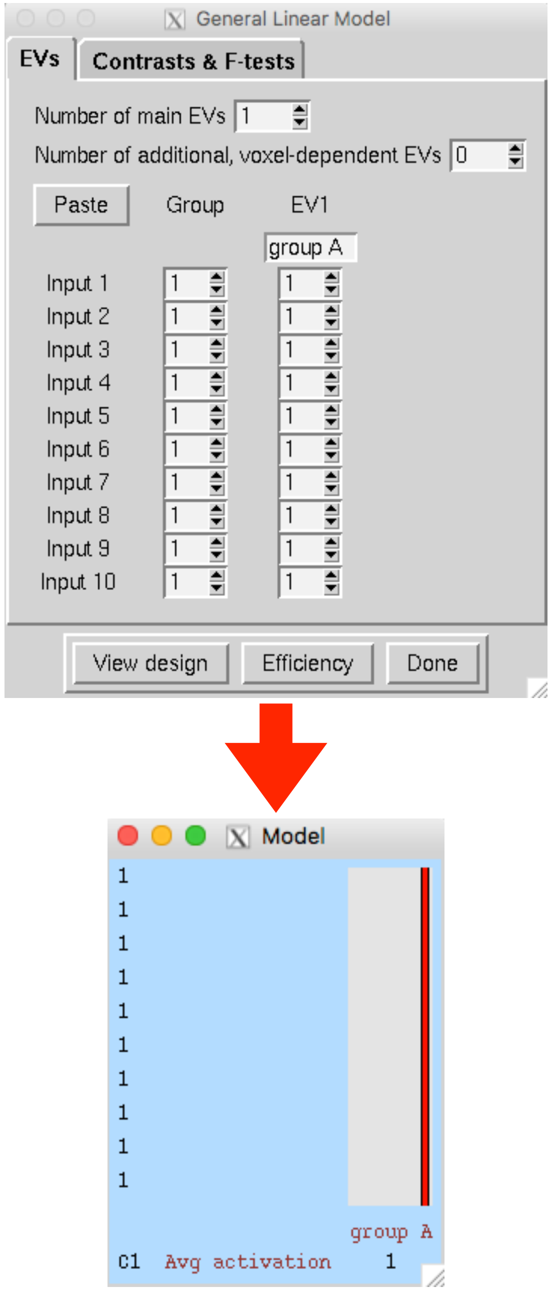




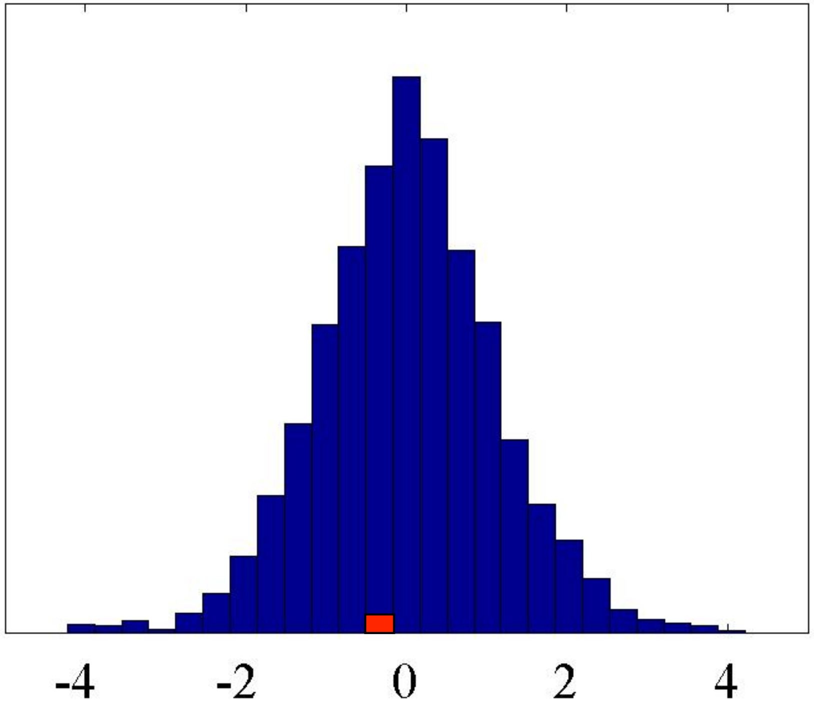
Second flip
第二次置换

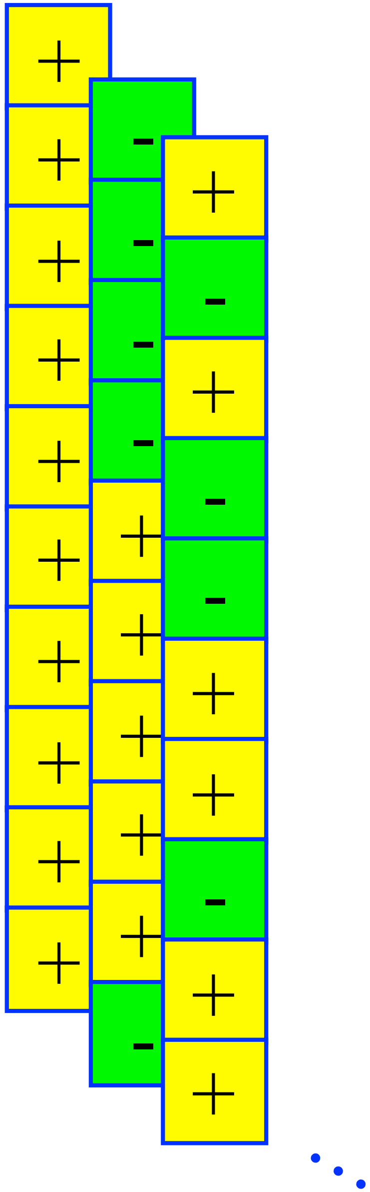


Examples of exchangeability: 可交换例子：
Single group average 单组平均

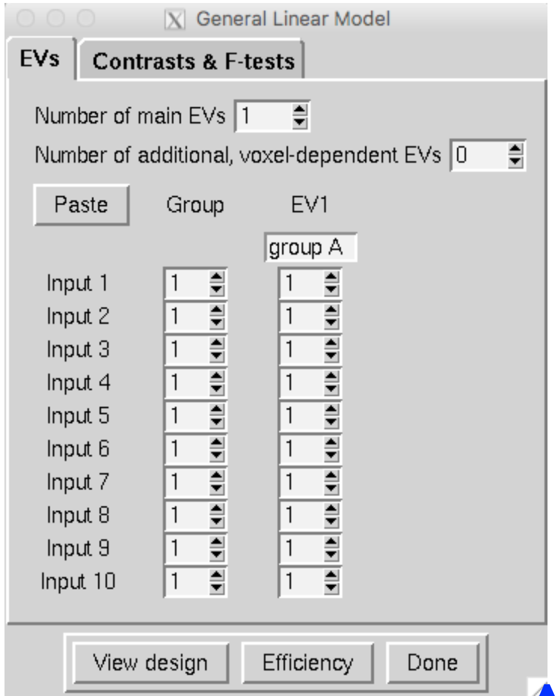


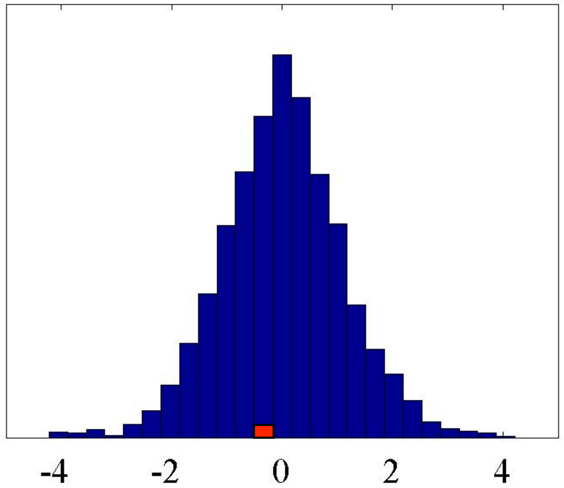
Etc … 等等…

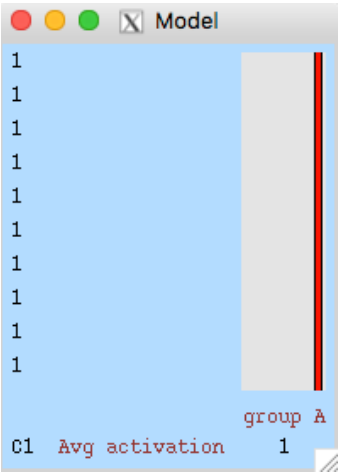




Examples of exchangeability: 可交换例子：
Single group average 单组平均



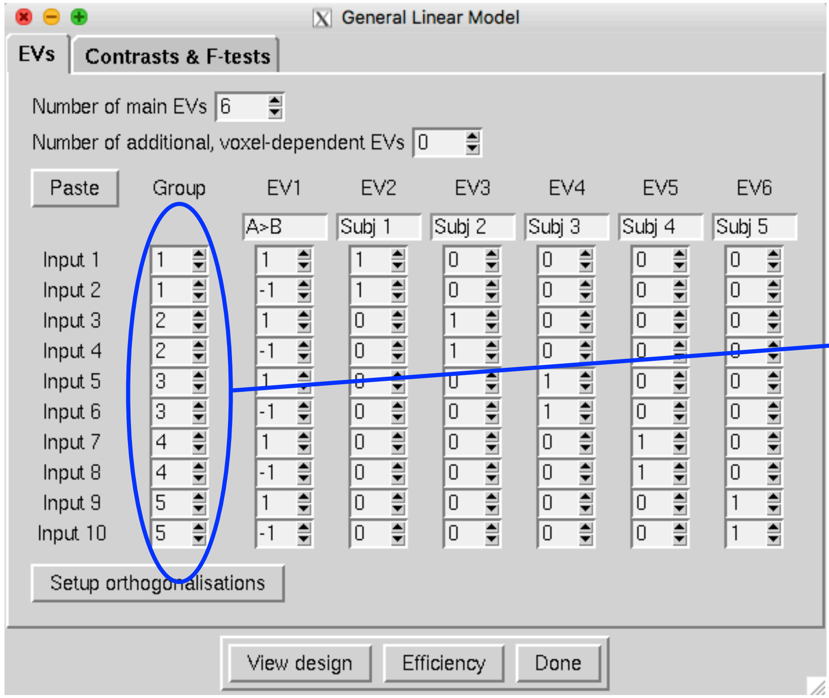




And the assumptions are: 假设是

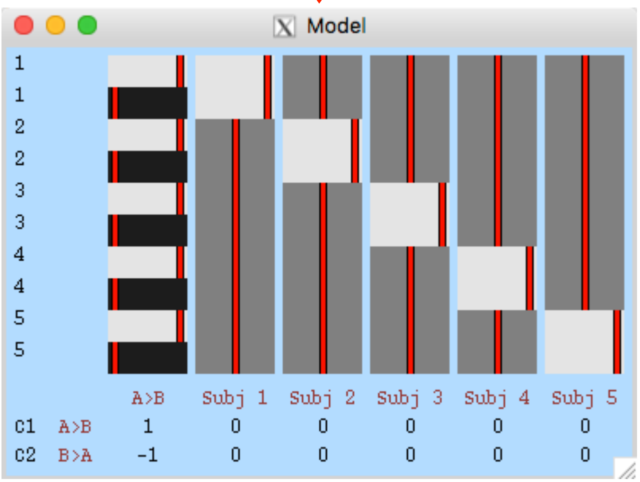
- Symmetric errors 对称误差
- Errors independent 误差独立
- Subjects drawn from a single population
  来自单一人群的被试

Examples of exchangeability: 可交换例子：
Two groups paired 两组配对



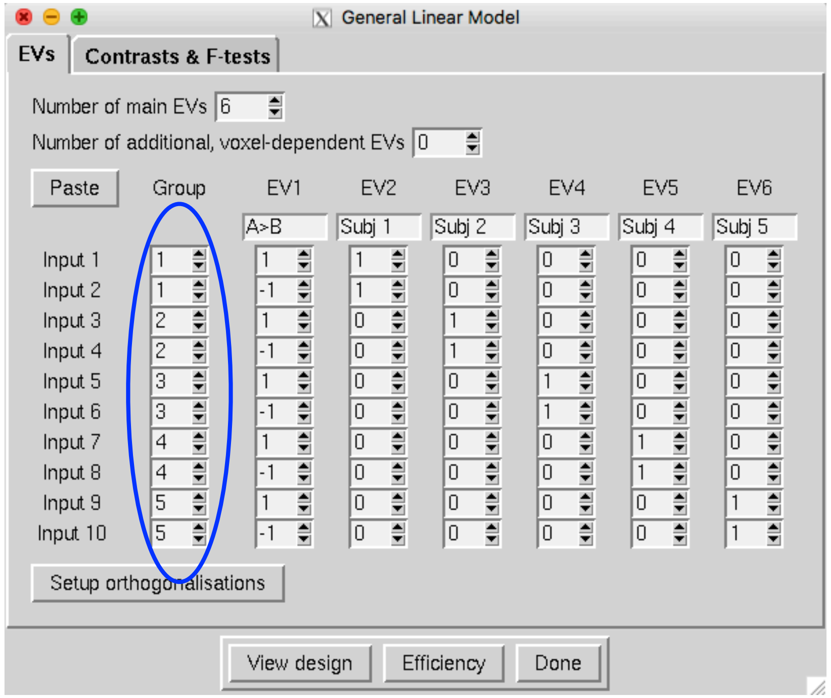
Here we can only exchange scans within each subject. I.e. Input 1 for Input 2, Input 3 for Input 4 etc

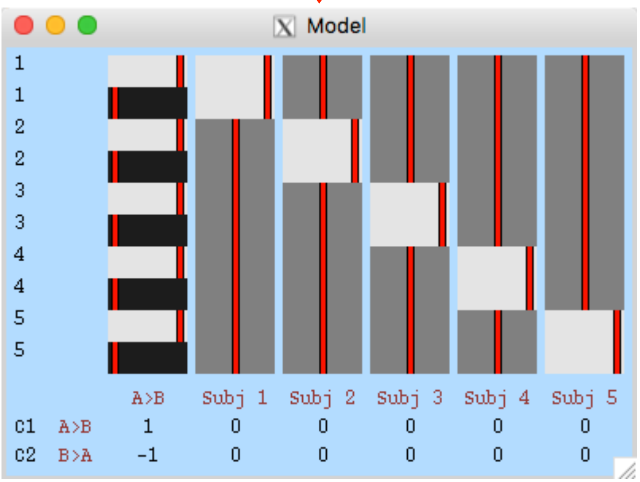
这里我们只能交换被试内的扫描，也就是1对2，3对4



Examples of exchangeability: 可交换例子：

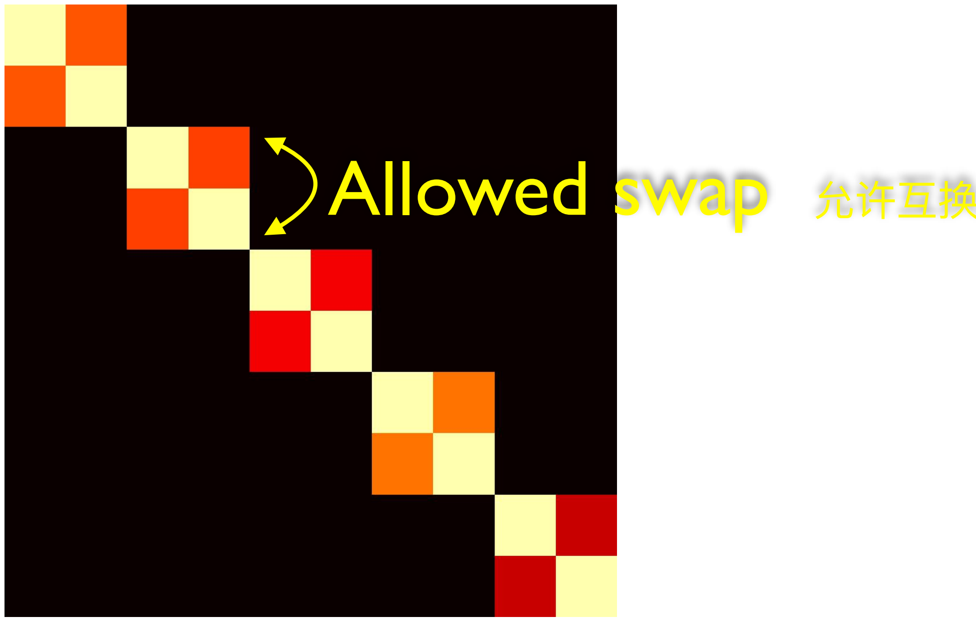
Two groups paired 两组配对





Assumed covariance matrix

假设的协方差矩阵

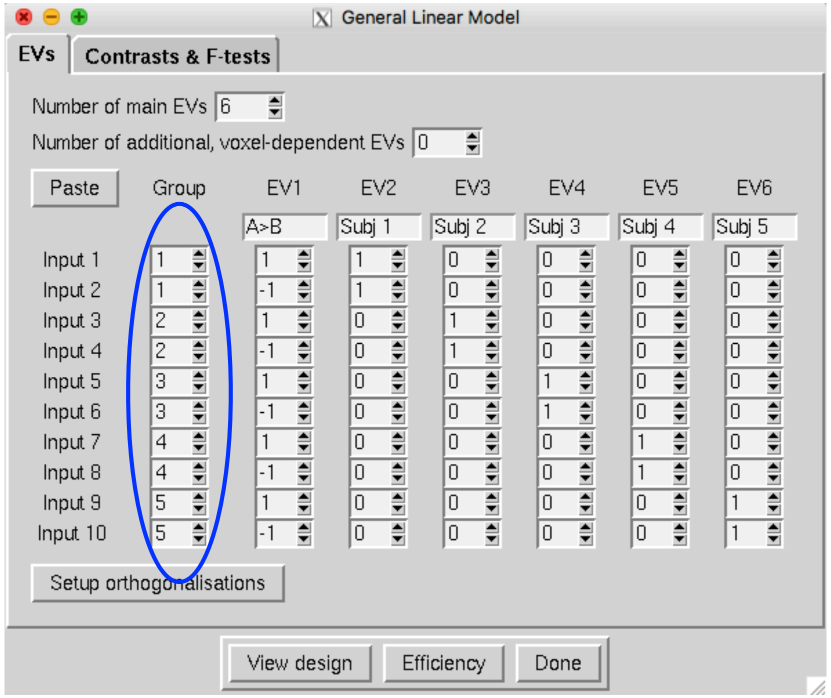


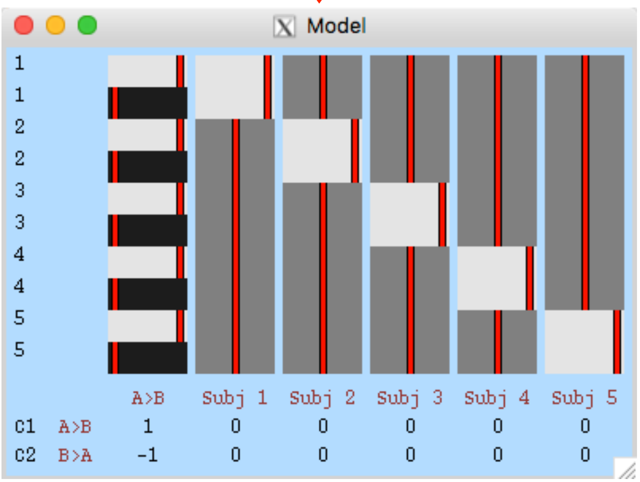
The implicit assumption here is that data from all subjects have the same uncertainty and that there is no dependence between subjects

隐含假设是所有被试的数据具有相同不确定性，并且被试间没有依赖性。

Examples of exchangeability: 可交换例子：

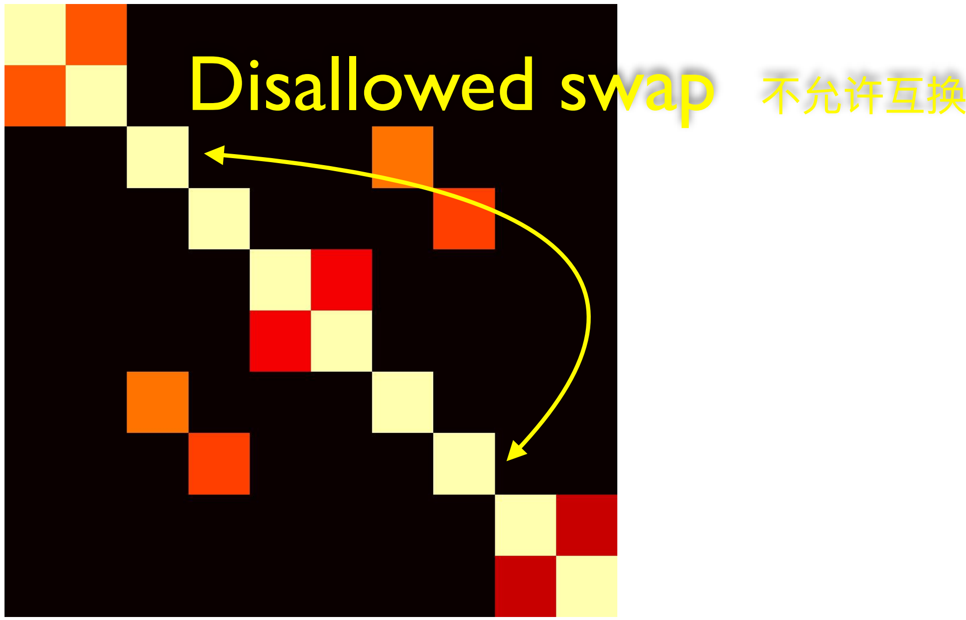
Two groups paired 两组配对





Assumed covariance matrix

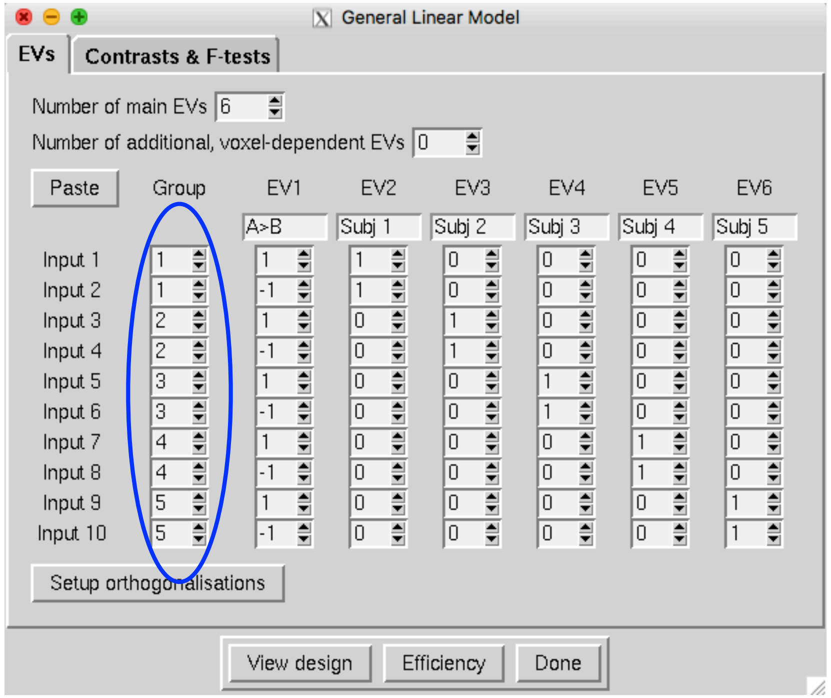
假设的协方差矩阵

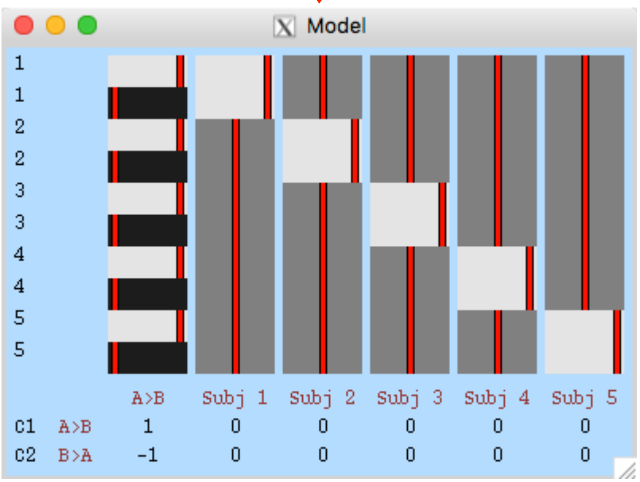


The implicit assumption here is that data from all subjects have the same uncertainty and that there is no dependence between subjects

隐含假设是所有被试的数据具有相同不确定性，并且被试间没有依赖性。

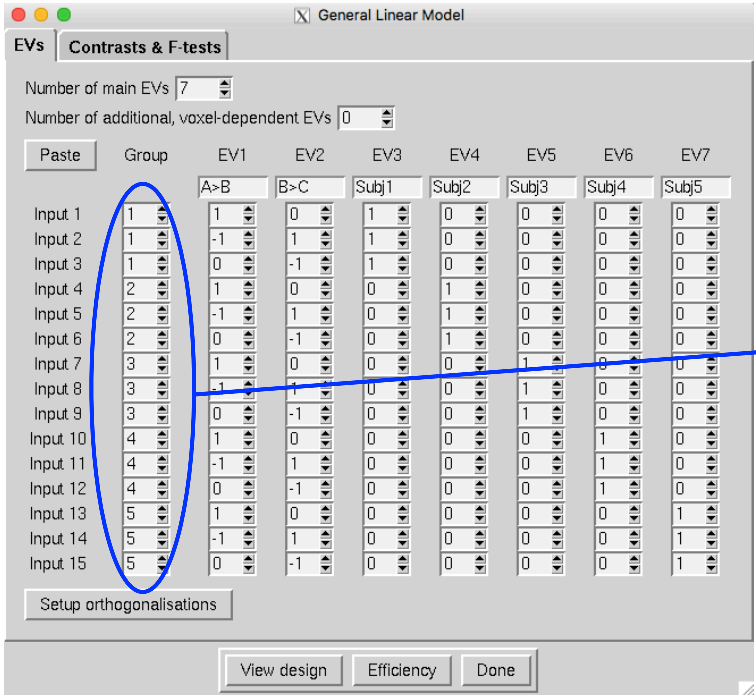
Examples of exchangeability: 可交换例子：
Two groups paired 两组配对





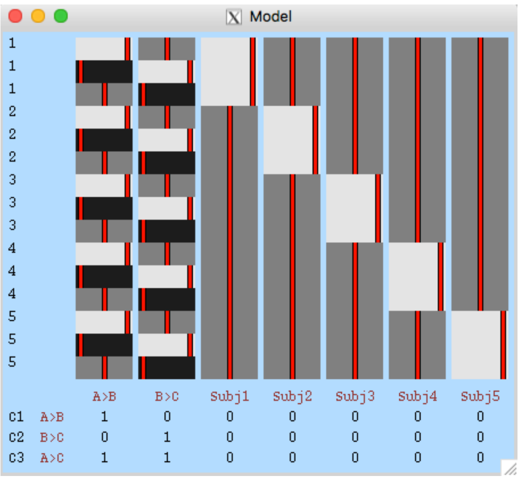
Original 原始	Perm 1 置换1	Perm 2 置换2	…
1	2	1	
2	1	2	
3	3	3	
4	4	4	
5	5	6	
6	6	5	
7	8	8	
8	7	7	
9	9	9	
10	10	10	

Examples of exchangeability: 可交换例子：
blocked ANOVA 组块方差分析

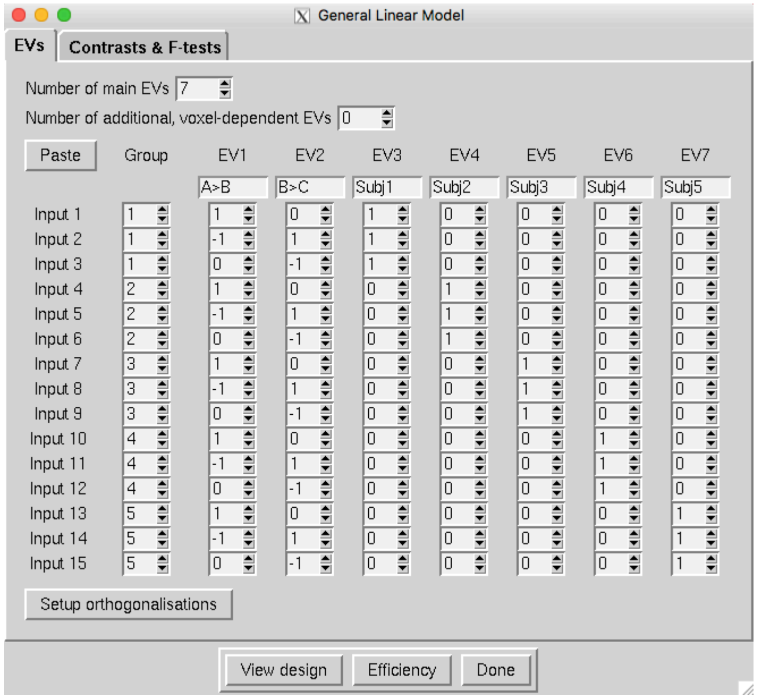


Same as previous: We can only swap labels within each subject

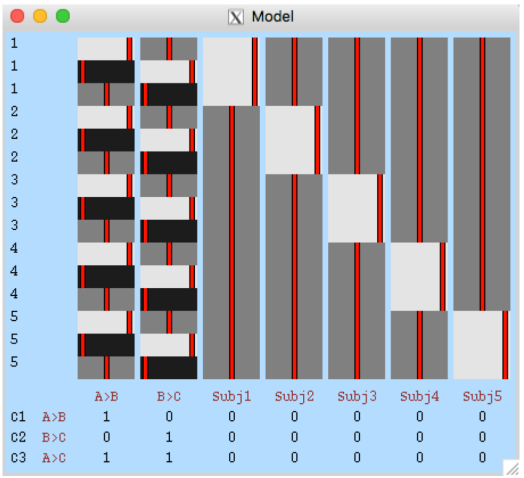
和之前一样，我们只能在被试内交换标记



Examples of exchangeability: 可交换例子：
blocked ANOVA 组块方差分析



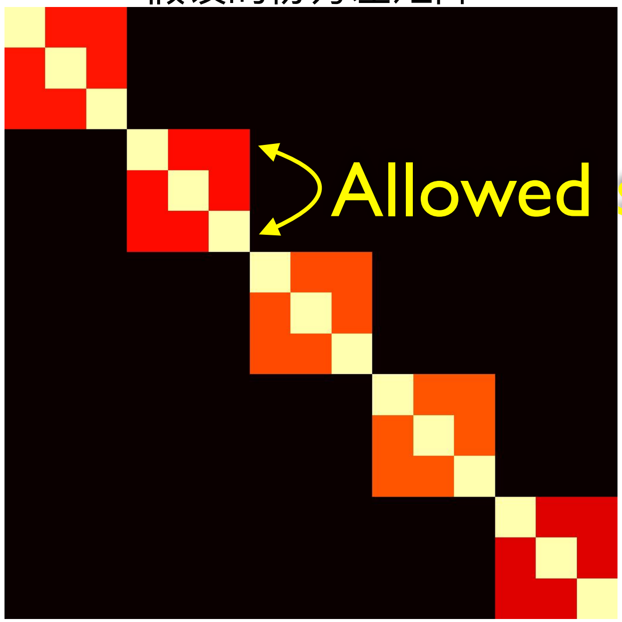
	Group	EV1	EV2	EV3	EV4	EV5	EV6	EV7
		A>B	B>C	Subj1	Subj2	Subj3	Subj4	Subj5
Input 1	1	1	0	1	0	0	0	0
Input 2	1	-1	1	1	0	0	0	0
Input 3	1	0	-1	1	0	0	0	0
Input 4	2	1	0	0	1	0	0	0
Input 5	2	-1	1	0	1	0	0	0
Input 6	2	0	-1	0	1	0	0	0
Input 7	3	1	0	0	0	1	0	0
Input 8	3	-1	1	0	0	1	0	0
Input 9	3	0	-1	0	0	1	0	0
Input 10	4	1	0	0	0	0	1	0
Input 11	4	-1	1	0	0	0	1	0
Input 12	4	0	-1	0	0	0	1	0
Input 13	5	1	0	0	0	0	0	1
Input 14	5	-1	1	0	0	0	0	1
Input 15	5	0	-1	0	0	0	0	1



		A>B	B>C	Subj1	Subj2	Subj3	Subj4	Subj5
C1	A>B	1	0	0	0	0	0	0
C2	B>C	0	1	0	0	0	0	0
C3	A>C	1	1	0	0	0	0	0

Assumed covariance matrix

假设的协方差矩阵

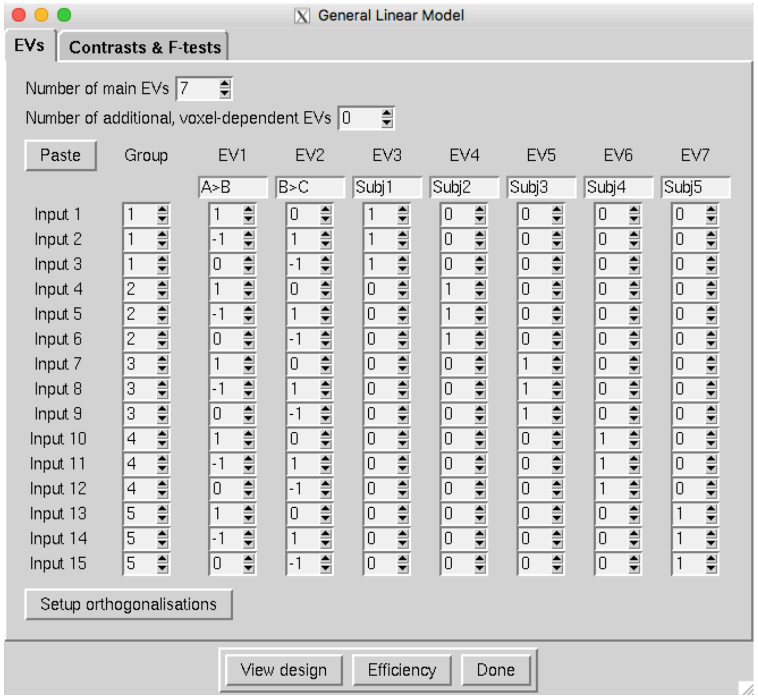


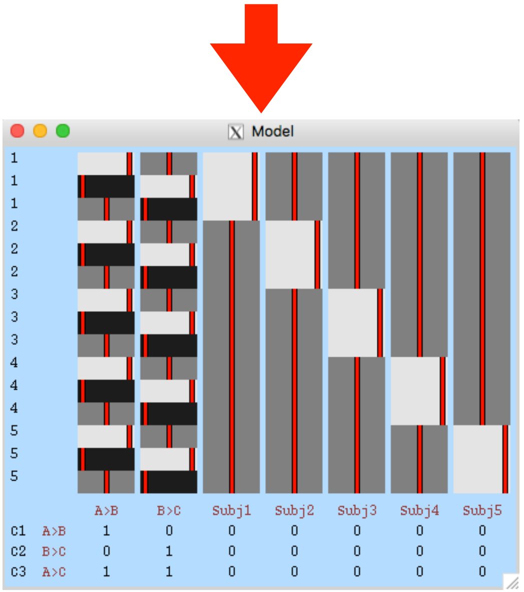
Allowed swap 允许交换

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

假设：所有被试来自同一人群，被试间无依赖性，并且被试内"复合对称"

Examples of exchangeability: 可交换例子：
blocked ANOVA 组块方差分析





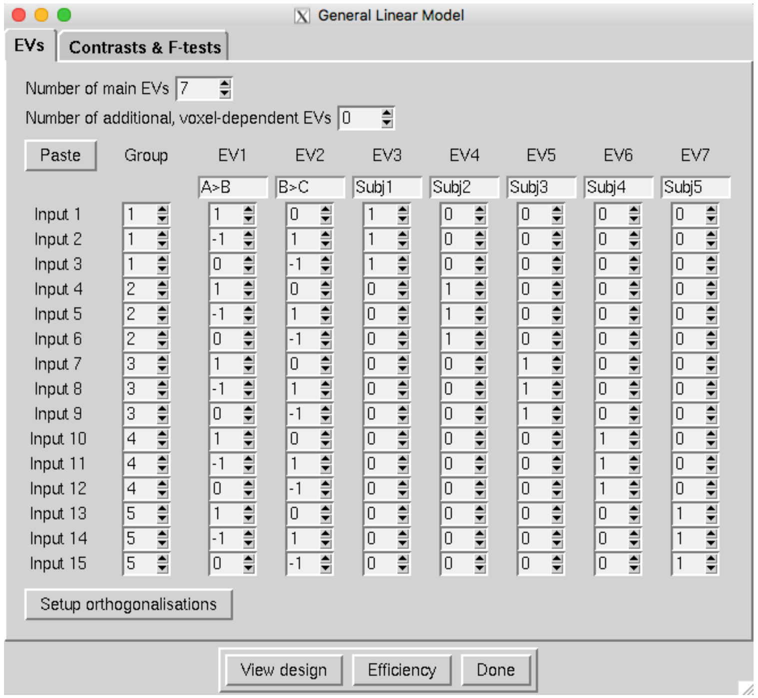
Assumed covariance matrix

假设的协方差矩阵

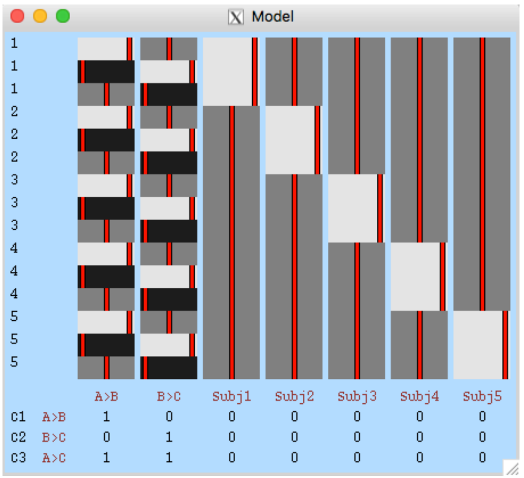
Assumptions: All subjects from the same "population", no dependence between subjects and "compound symmetry" within subjects

假设：所有被试来自同一人群，被试间无依赖性，并且被试内"复合对称"

Examples of exchangeability: 可交换例子：
blocked ANOVA 组块方差分析



	Group	EV1	EV2	EV3	EV4	EV5	EV6	EV7
		A>B	B>C	Subj1	Subj2	Subj3	Subj4	Subj5
Input 1	1	1	0	1	0	0	0	0
Input 2	1	-1	1	1	0	0	0	0
Input 3	1	0	-1	1	0	0	0	0
Input 4	2	1	0	0	1	0	0	0
Input 5	2	-1	1	0	1	0	0	0
Input 6	2	0	-1	0	1	0	0	0
Input 7	3	1	0	0	0	1	0	0
Input 8	3	-1	1	0	0	1	0	0
Input 9	3	0	-1	0	0	1	0	0
Input 10	4	1	0	0	0	0	1	0
Input 11	4	-1	1	0	0	0	1	0
Input 12	4	0	-1	0	0	0	1	0
Input 13	5	1	0	0	0	0	0	1
Input 14	5	-1	1	0	0	0	0	1
Input 15	5	0	-1	0	0	0	0	1



Assumed covariance matrix

假设的协方差矩阵

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

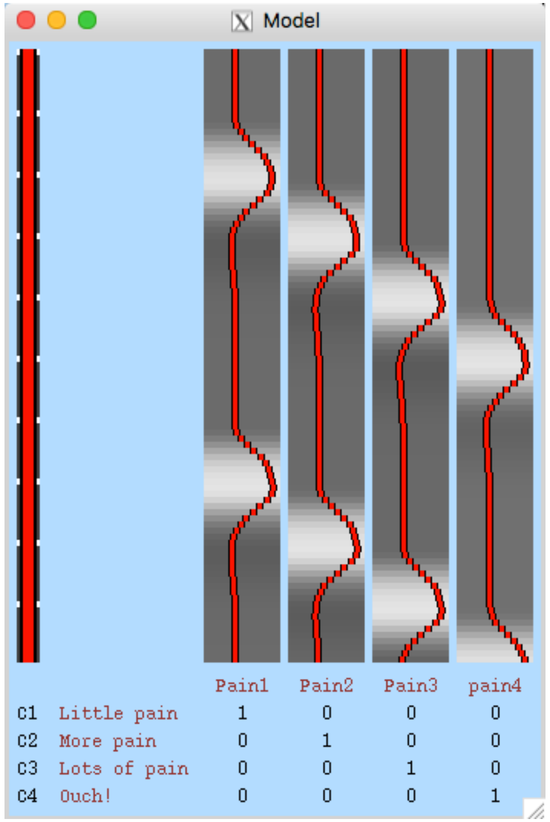
假设：所有被试来自同一人群，被试间无依赖性，并且被试内"复合对称"

My advice: Keep it simple!

我的建议：保持简单

Each subject scanned like this

每个被试扫描看起来这样



Taking 4 contrasts to 2nd level

第二水平有4个对比

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

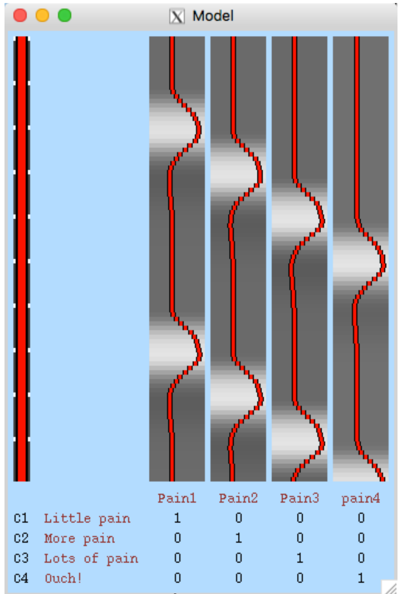
我们想找到对疼痛线性响应的区域。

My advice: Keep it simple!

我的建议：保持简单

Each subject scanned like this

每个被试扫描看起来这样

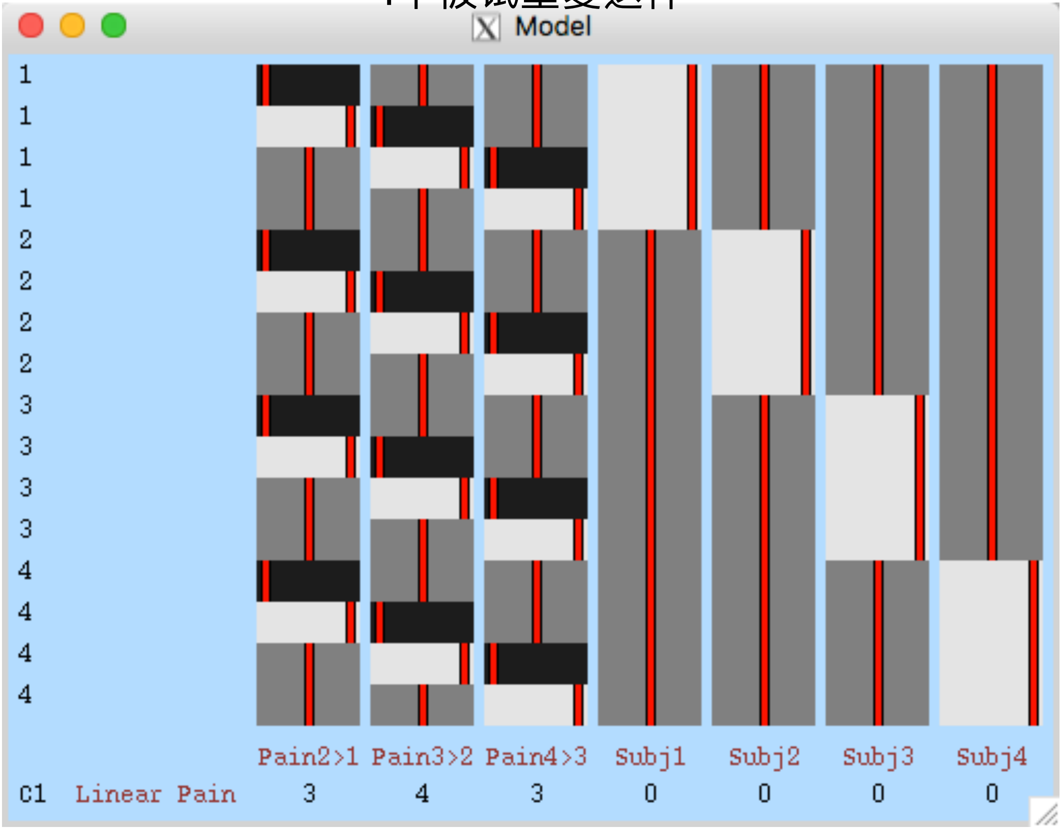


Taking 4 contrasts to 2nd level

第二水平有4个对比

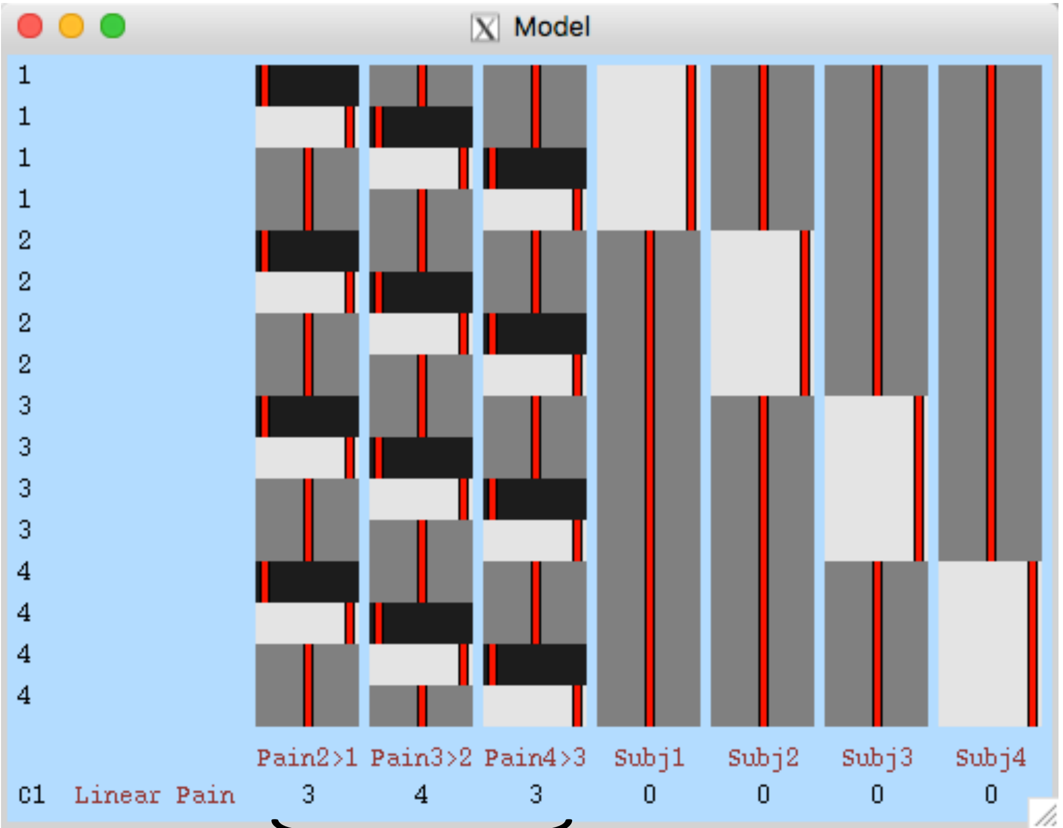
Repeating this for four subjects

4个被试重复这样



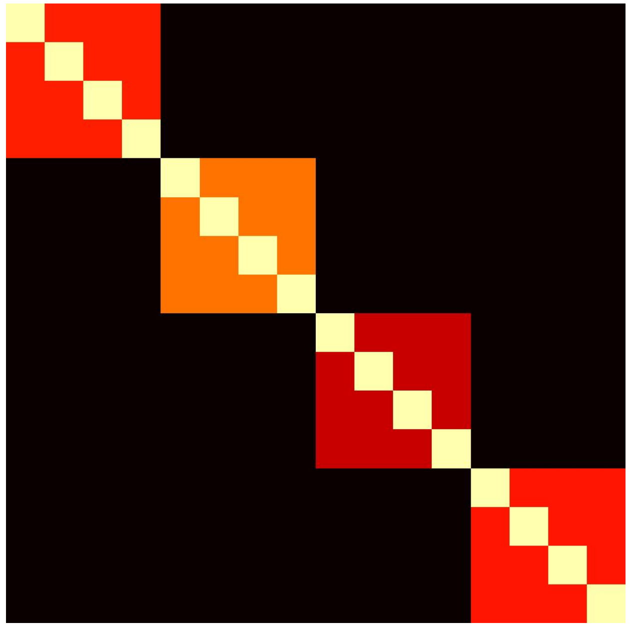
My advice: Keep it simple!

我的建议：保持简单



And figure out this contrast

推理得到这样设置对比



You have to assume this covariance matrix

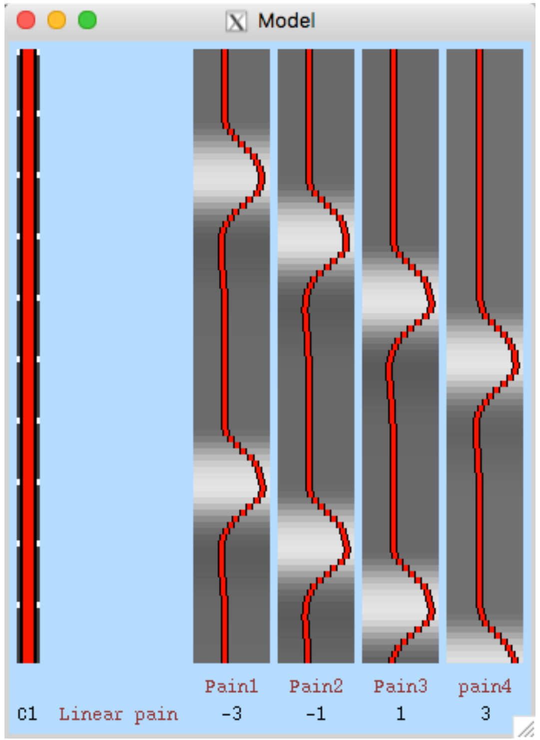
必须假设这个协方差矩阵

Why put yourself through all that pain?

为什么要忍受所有痛苦?

My advice: Keep it simple!

我的建议：保持简单

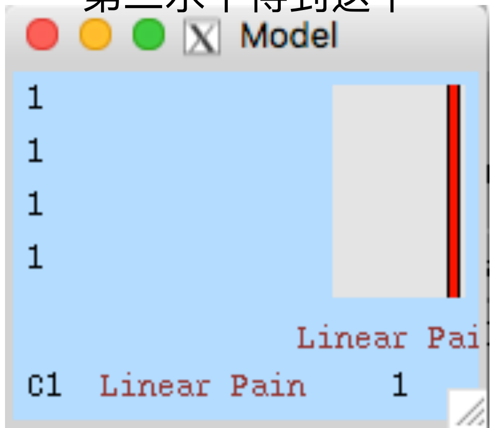


When you can take a single contrast from the first level

当你可以跟第一水平做简单对比时

And get this at the second level

第二水平得到这个



Assuming only symmetric errors

仅假设对称误差

Much nicer, no?

更赞，不是吗？

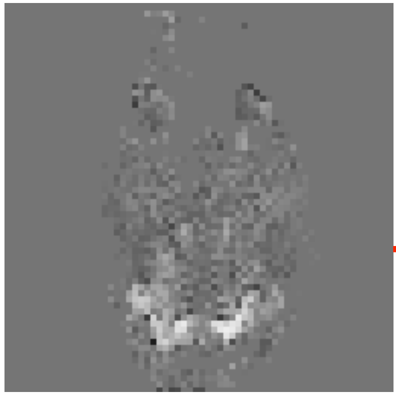
Outline大纲

- Null-hypothesis and Null-distribution 零假设和零分布
- Multiple comparisons and Family-wise error 多重比较和族错误率
- Different ways of being surprised 惊奇的不同方式
  - Voxel-wise inference (Maximum z) 体素推断（最大z）
  - Cluster-wise inference (Maximum size) 簇水平推断（最大尺寸）
- Parametric vs non-parametric tests 参数vs非参检验
- Enhanced clusters 增强的簇
- FDR - False Discovery Rate FDR-错误发现率

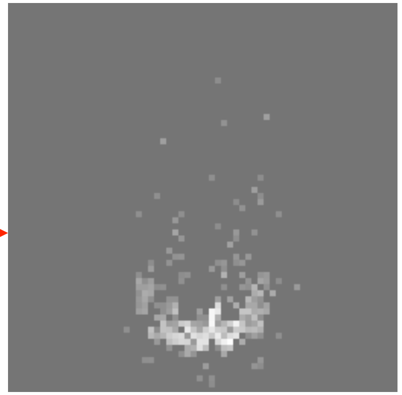
Clustering cookbook 簇指导

Instead of resel-based correction, we can do clustering:
除了基于resel的校正外，我么还能做簇：

z stat image z值图



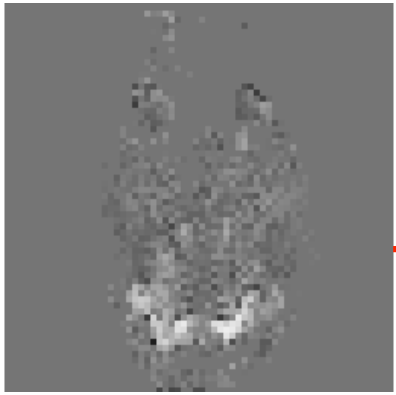
Threshold at (arbitrary!) z level
在z水平(任意)的阈值



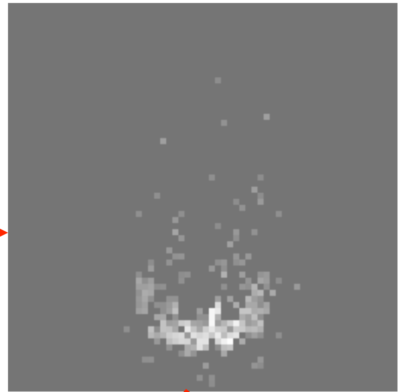
Clustering cookbook 簇指导

Instead of resel-based correction, we can do clustering:
除了基于resel的校正外，我么还能做簇：

z stat image z值图



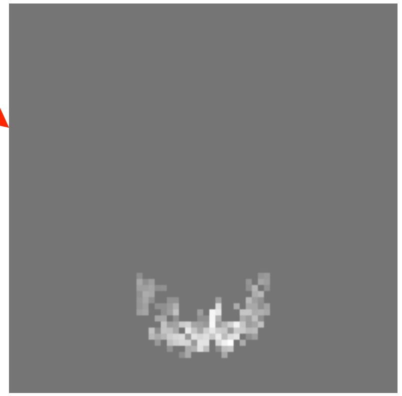
Threshold at (arbitrary!) z level
在z水平(任意)的阈值



Form clusters from surviving voxels.
从尚存体素形成簇

Calculate the size threshold $u(R,z)$.
计算大小的阈值$u(R,z)$

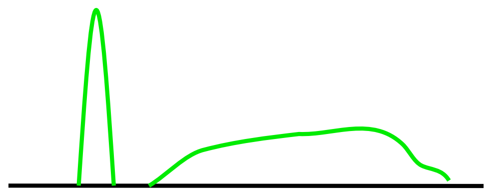
Any cluster larger than u "survives" and we reject the null-hypothesis for that.
存在大于u的簇，我们拒绝零假设



How do we choose the (arbitrary!) z-threshold?

我们是怎样(随意)选择z阈值的?

This is arbitrary and a trade-off 这是一个随意和权衡

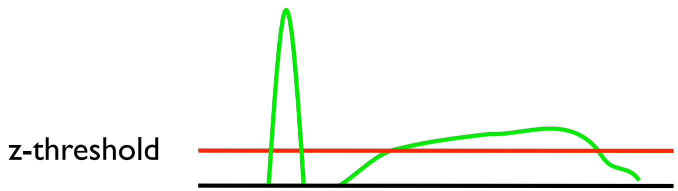


How do we choose the (arbitrary!) z-threshold?

我们是怎样(随意)选择z阈值的?

This is arbitrary and a trade-off 这是一个随意和权衡

1. **Low threshold** - can violate RFT assumptions, but can detect clusters with large spatial extent and low z
低阈值-可能违反RFT假设，但可以检测到具有较大空间范围和较低z的簇

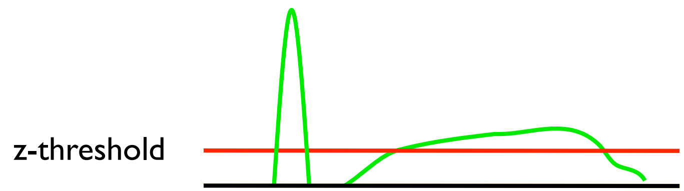


How do we choose the (arbitrary!) z-threshold?

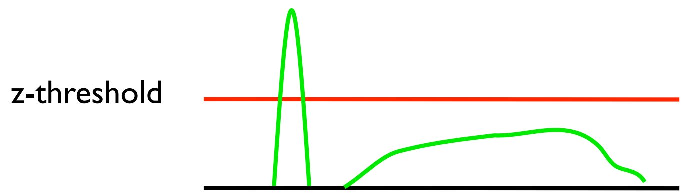
我们是怎样(随意)选择z阈值的?

This is arbitrary and a trade-off 这是一个随意和权衡

1. **Low threshold** - can violate RFT assumptions, but can detect clusters with large spatial extent and low z
低阈值-可能违反RFT假设，但可以检测到具有较大空间范围和较低z的簇



2. **High threshold** - gives more power to clusters with small spatial extent and high z
高阈值-为空间范围较小且z较高的簇提供更多功能

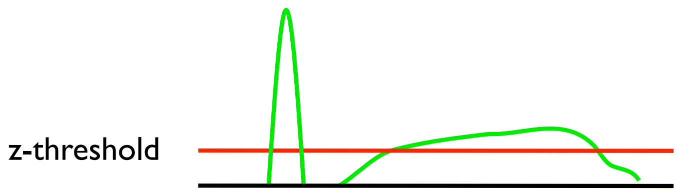


How do we choose the (arbitrary!) z-threshold?

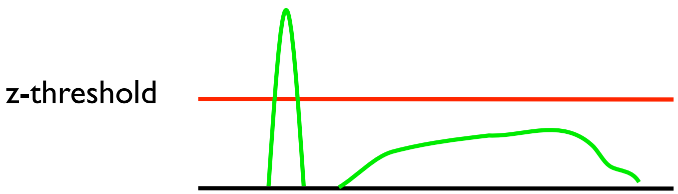
我们是怎样(随意)选择z阈值的?

This is arbitrary and a trade-off 这是一个随意和权衡

1. **Low threshold** - can violate RFT assumptions, but can detect clusters with large spatial extent and low z
低阈值-可能违反RFT假设，但可以检测到具有较大空间范围和较低z的簇



2. **High threshold** - gives more power to clusters with small spatial extent and high z
高阈值-为空间范围较小且z较高的簇提供更多功能



Tends to be more sensitive than voxel-wise corrected testing
倾向于比体素校正测试更敏感

Results depend on extent of spatial smoothing in pre-processing
结果取决于预处理中空间平滑的程度

TFCE

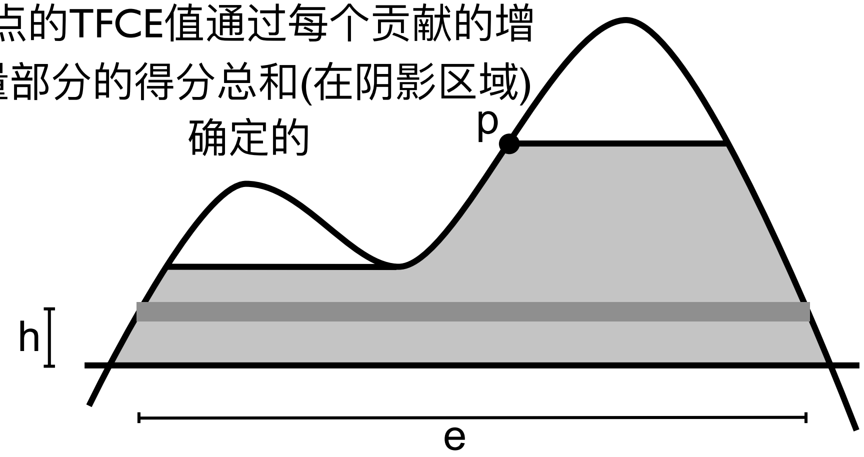
Threshold-Free Cluster Enhancement 无阈值簇增强

[Smith & Nichols, NeuroImage 2009]

- Cluster thresholding: 簇阈值
  - popular because it’s sensitive, due to its use of spatial extent 因为在空间范围内的使用导致的很敏感，所以流行
  - but the pre-smoothing extent is arbitrary 但是预平滑程度任意
  - and so is the cluster-forming threshold 簇形成阈值也如此

    ➡ unstable and arbitrary 不稳定和随意
- TFCE
  - integrates cluster “scores” over all possible thresholds 整合所有可能阈值的簇“得分
  - output at each voxel is measure of local cluster-like support 在每个体素输出局部簇支持的测量
  - similar sensitivity to optimal cluster-thresholding, but stable and non-arbitrary 与最佳簇阈值敏感性相同，但稳定且不随意

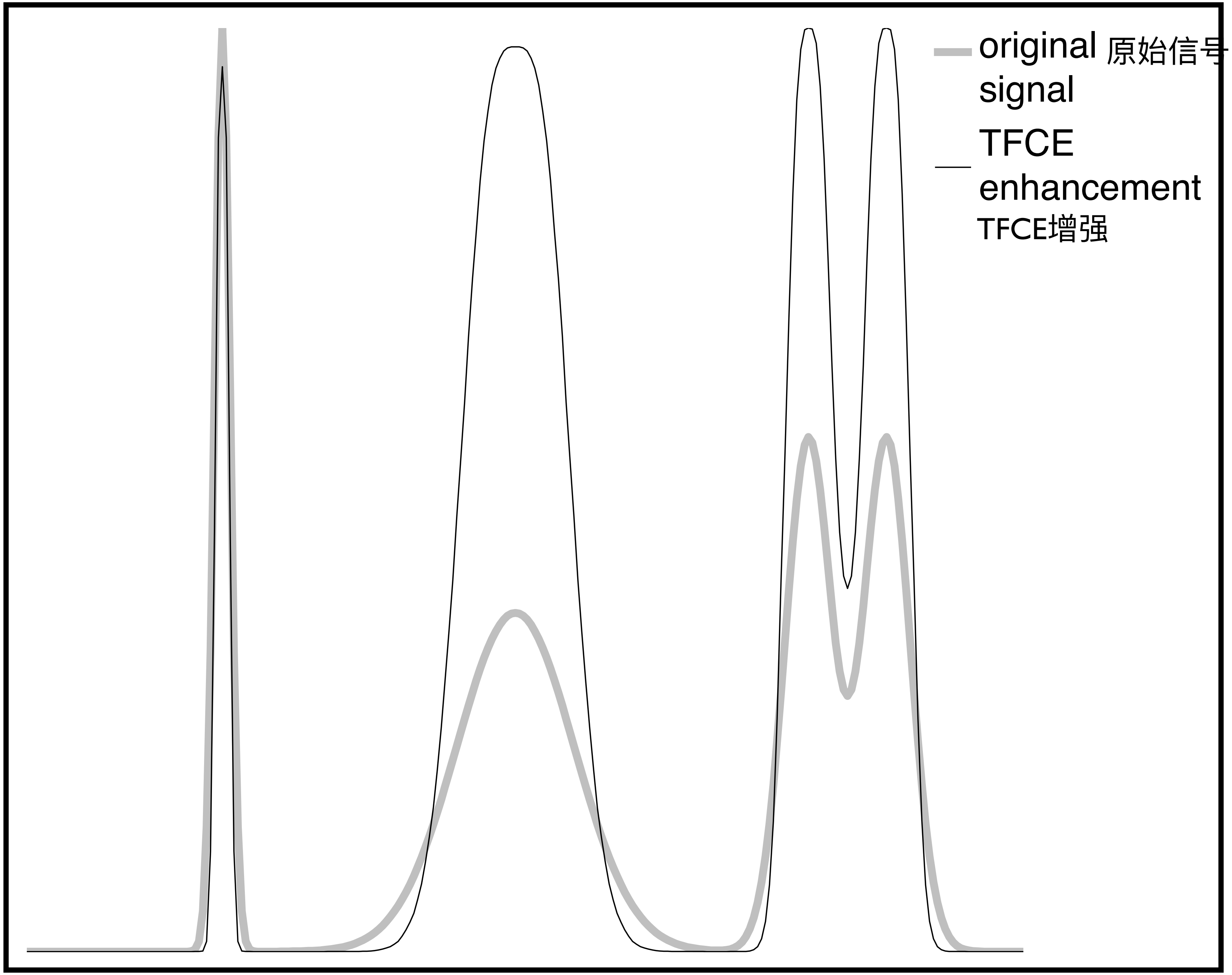
p点的TFCE值通过每个贡献的增量部分的得分总和(在阴影区域)确定的



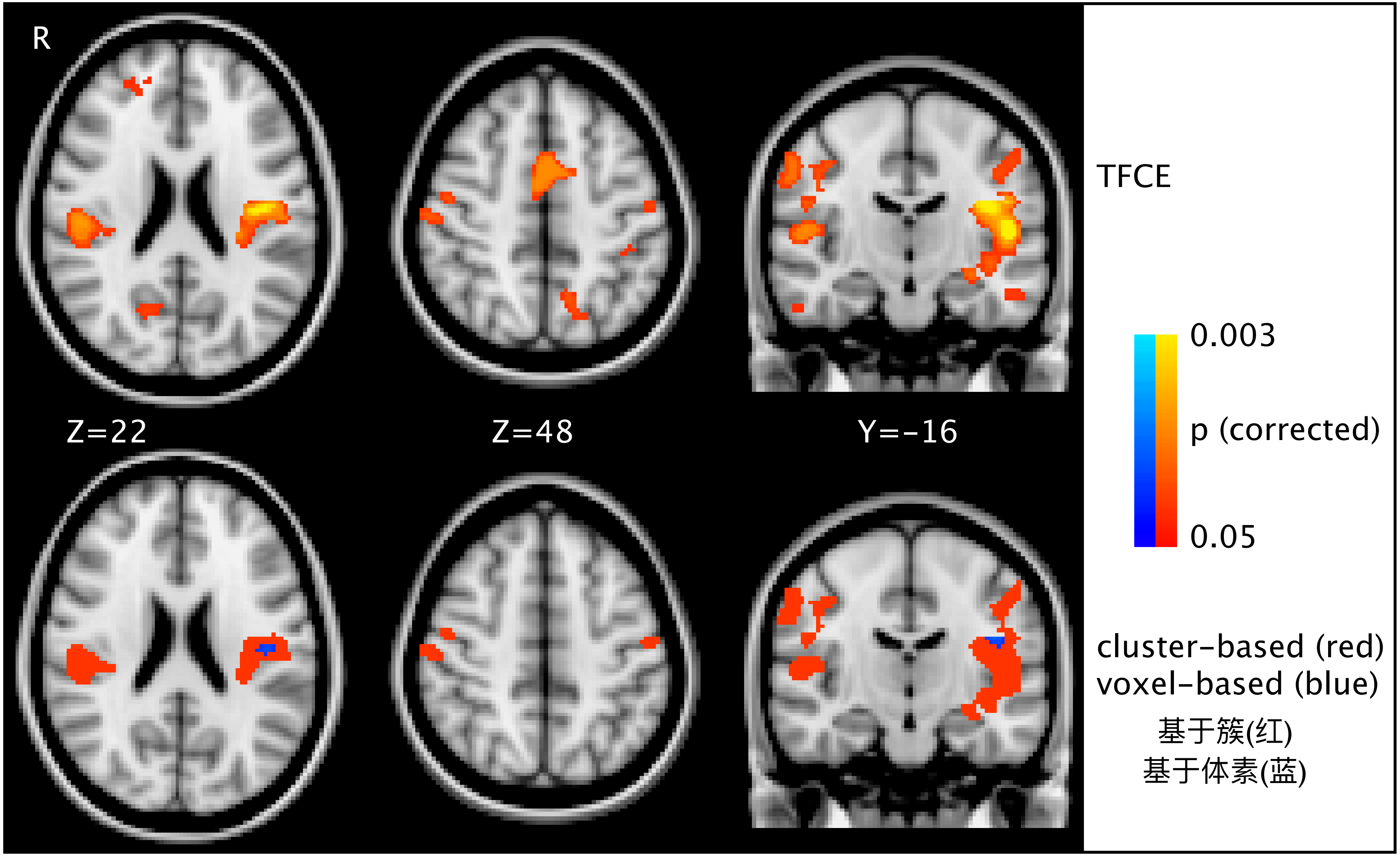
The TFCE value at point p is given by the sum, over the shaded area, of the score from each contributing incremental section:

$$\mathrm{TFCE}(p) = \sum_h e(h)^E \cdot h^H$$

Qualitative example
定性例子



TFCE for FSL-VBM

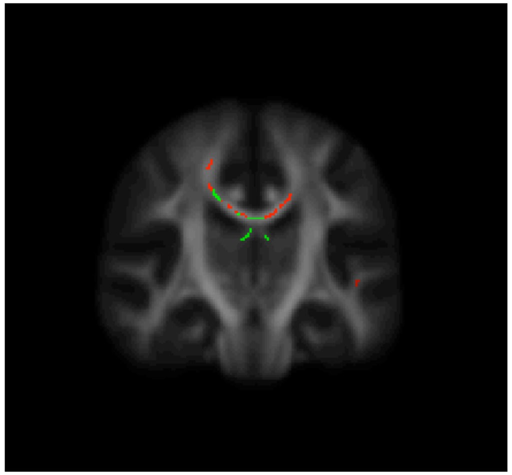


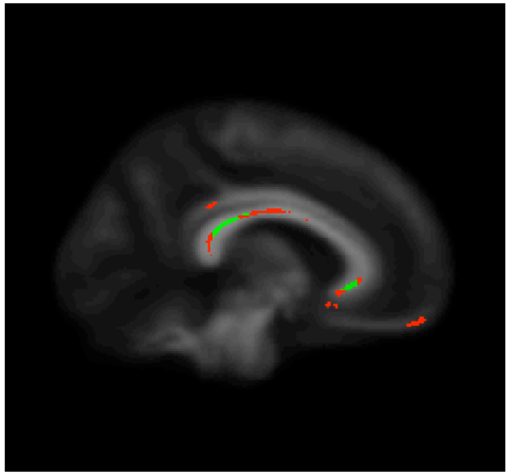
TFCE for TBSS

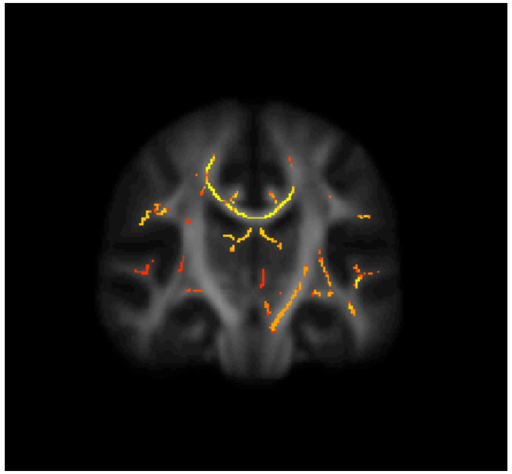
controls > schizophrenics 对照 > 精分患者

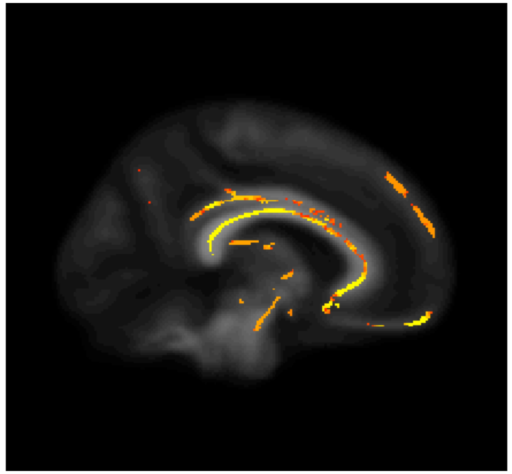
p<0.05 corrected for multiple comparisons across space, using randomise

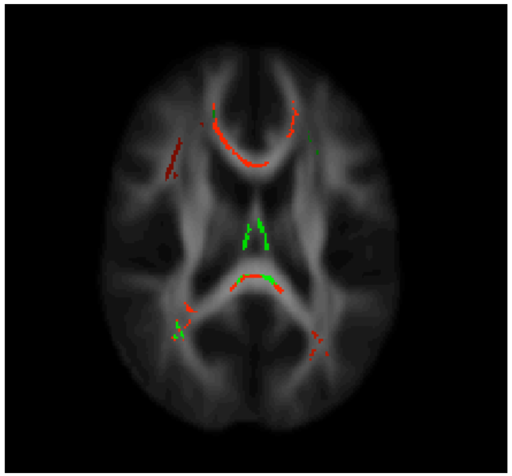
跨空间的多重比较校正，使用randomise工具











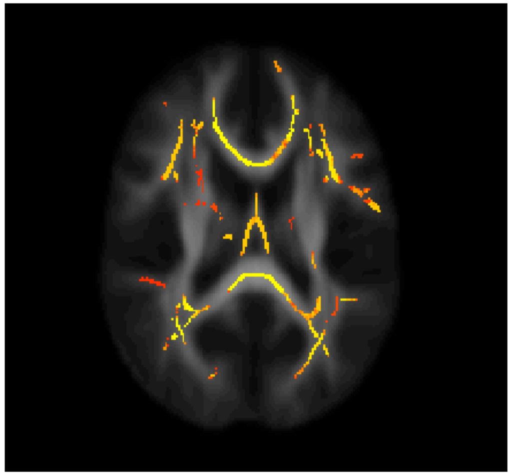
cluster-based:

基于簇

cluster-forming threshold =

2 or 3

簇形成阈值=2或3



TFCE

Outline大纲

- Null-hypothesis and Null-distribution 零假设和零分布
- Multiple comparisons and Family-wise error 多重比较和族错误率
- Different ways of being surprised 惊奇的不同方式
  - Voxel-wise inference (Maximum z) 体素推断（最大z）
  - Cluster-wise inference (Maximum size) 簇水平推断（最大尺寸）
- Parametric vs non-parametric tests 参数vs非参检验
- Enhanced clusters 增强的簇
- FDR - False Discovery Rate FDR-错误发现率

False Discovery Rate
错误发现率

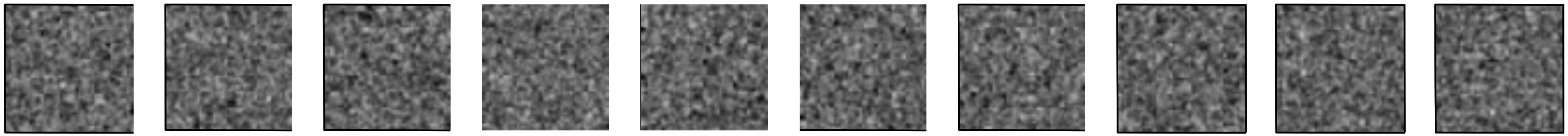
- FDR: False Discovery Rate 错误发现率
  A “new” way to look at inference. 一种新的推断方式
- Uncorrected (for multiple-comparisons): 未校正(多重比较)
  - Is equivalent to saying: “I am happy to nearly always say something silly about my experiments”.
    等于说：“我很乐意针对我全部实验总结一些辣鸡结论。”
- Family-Wise Error (FWE): 总体误差
  - Is equivalent to saying: “I am happy to say something silly about 5% of my experiments”.
    相当于说：“我乐意对约5%的实验结果总结一些辣鸡结论”。
- False Discovery Rate 错误发现率
  - Is equivalent to saying: “I am happy if 5% of what I say about each experiment is silly”.
    等于说：“如果我的实验总结只有5%很辣鸡，我爽翻。”

False Discovery Rate
错误发现率

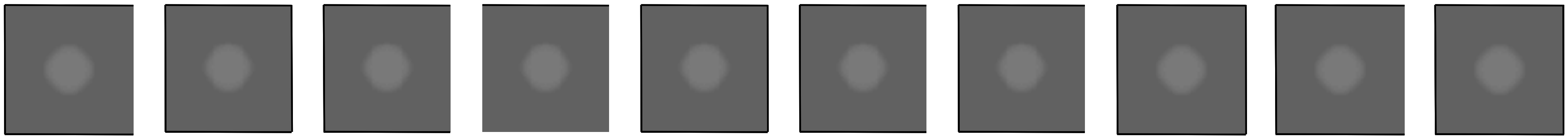
- FDR: False Discovery Rate 错误发现率
  A “new” way to look at inference. 一种新的推断方式
- Uncorrected (for multiple-comparisons): 未校正(多重比较)
  - Is equivalent to saying: “I am happy to nearly always say something silly about my experiments”.
    等于说：“我很乐意针对我全部实验总结一些辣鸡结论。”
  - On average, **5% of all voxels** are false positives
    平均而言，所有体素中有5%是假阳性
- Family-Wise Error (FWE): 总体误差
  - Is equivalent to saying: “I am happy to say something silly about 5% of my experiments”.
    相当于说：“我乐意对约5%的实验结果总结一些辣鸡结论”。
  - On average, **5% of all experiments** have one or more false positive voxels 平均而言，所有实验中有5%具有一个或多个假阳性体素
- False Discovery Rate 错误发现率
  - Is equivalent to saying: “I am happy if 5% of what I say about each experiment is silly”.
    等于说：“如果我的实验总结只有5%很辣鸡，我爽翻。”
  - On average, **5% of significant voxels** are false positives 平均而言，5%的显著体素是假阳性。

Little imaging demonstration.
影像演示

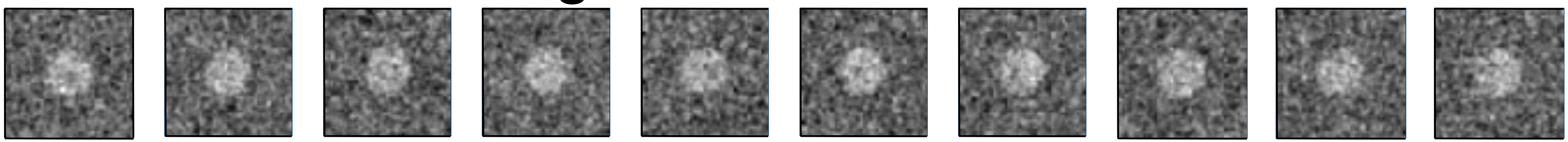
Noise 噪音



Signal 信号

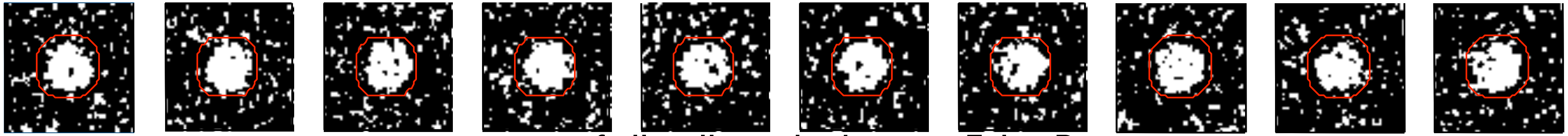


Signal+Noise 信号+噪音



uncorrected voxelwise control of FP rate at 10%

假阳性率在10%的未校正体素控制

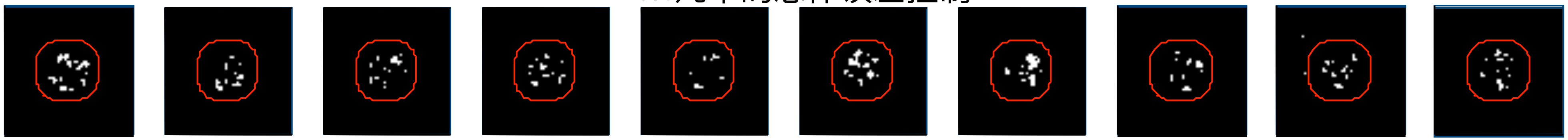


percentage of all null pixels that are False Positives

所有假阳性零像素的百分比

control of FamilyWise Error rate at 10%

10%几率的总体误差控制



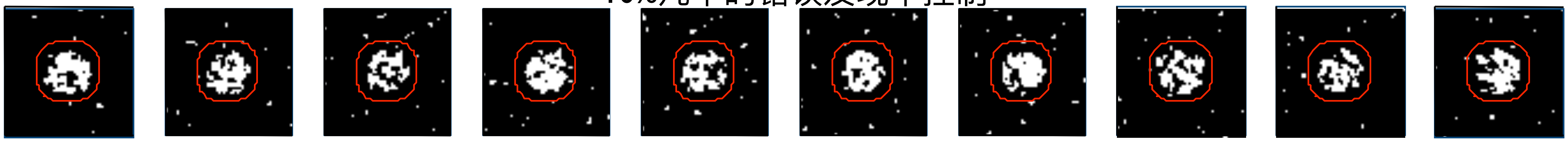
occurrence of FamilyWise Error

总体误差的发生率

FWE

control of False Discovery Rate at 10%

10%几率的错误发现率控制



percentage of activated (reported) pixels that are False Positives

报告的假阳性的激活像素百分比

FDR for dummies

傻瓜式FDR指南

- Makes assumptions about how errors are distributed (like GRT). 假设误差分布的方式
- Used to calculate a threshold. 用于计算阈值。
- Threshold such that X% of super-threshold (reported) voxels are false positives. 卡阈值使得X%超阈值(报告的)体素是假阳性。
- Threshold depends on the data. May for example be very different for [1 0] and [0 1] in the same study. 阈值取决于数据。同一个研究中的 [1 0] 和 [0 1] 可能有很大差异。