Second Level Analysis

Turku PET Centre Brain Imaging Course 2026

Rui Watanabe, Turku PET Centre

Topics

01

Overview

Second-level GLM

02

Second-level designs

03

Multiple comparisons problem

Topics

01

Overview

Second-level GLM

02

Second-level designs

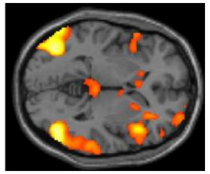
03

Multiple comparisons problem

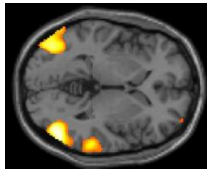
Why do we need to conduct second level analysis?

Second-level analysis allows us to test whether an effect is reliable at the group level and generalize the finding from our sample to the population.

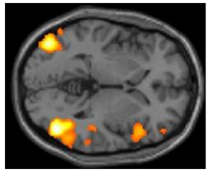
sub 1



sub 2

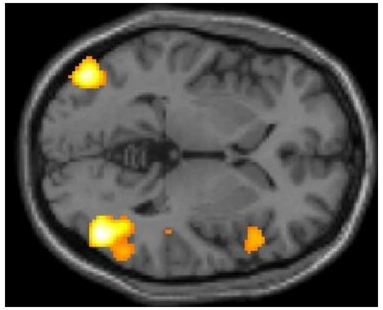


sub 3

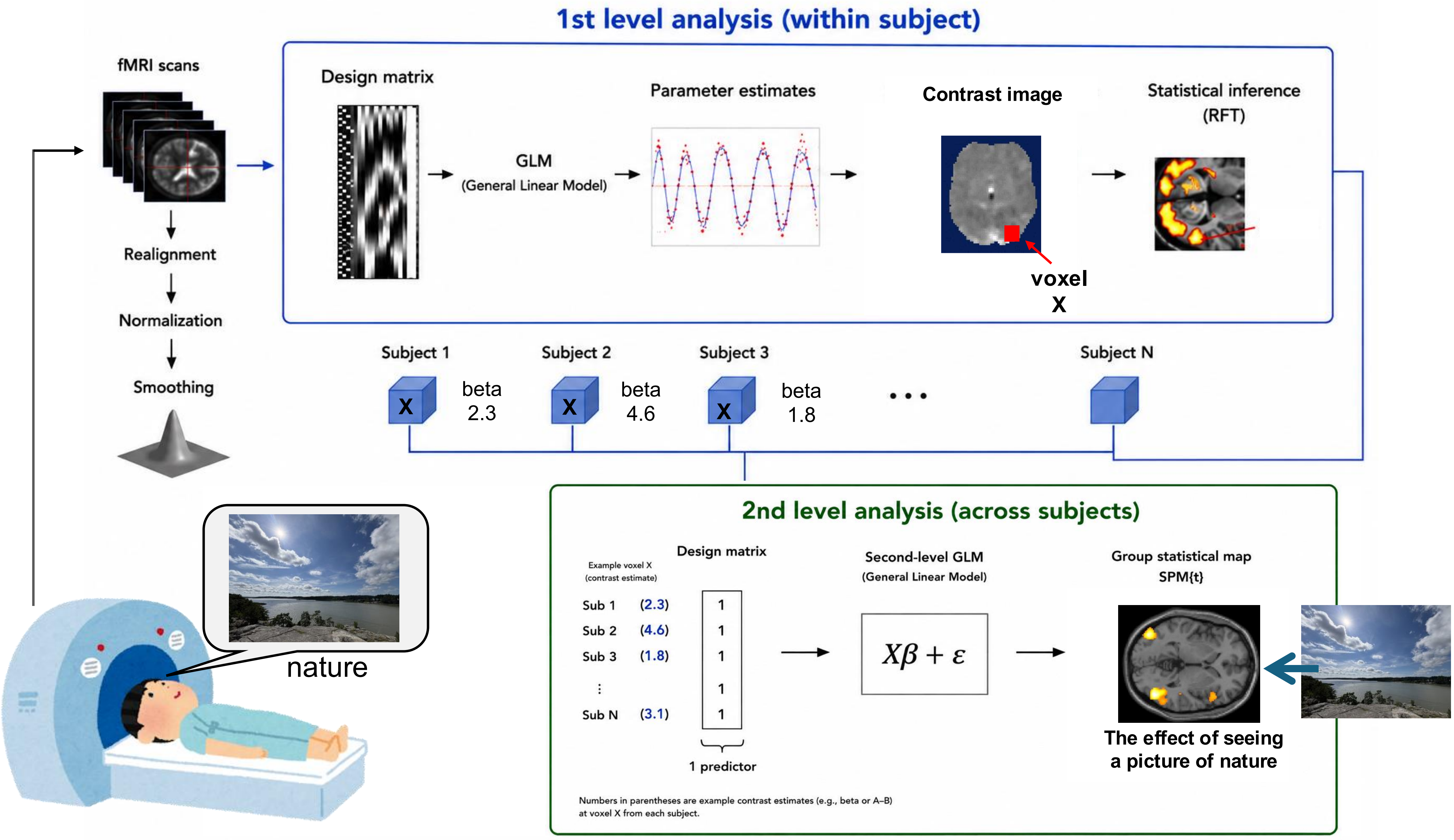


:

sub N



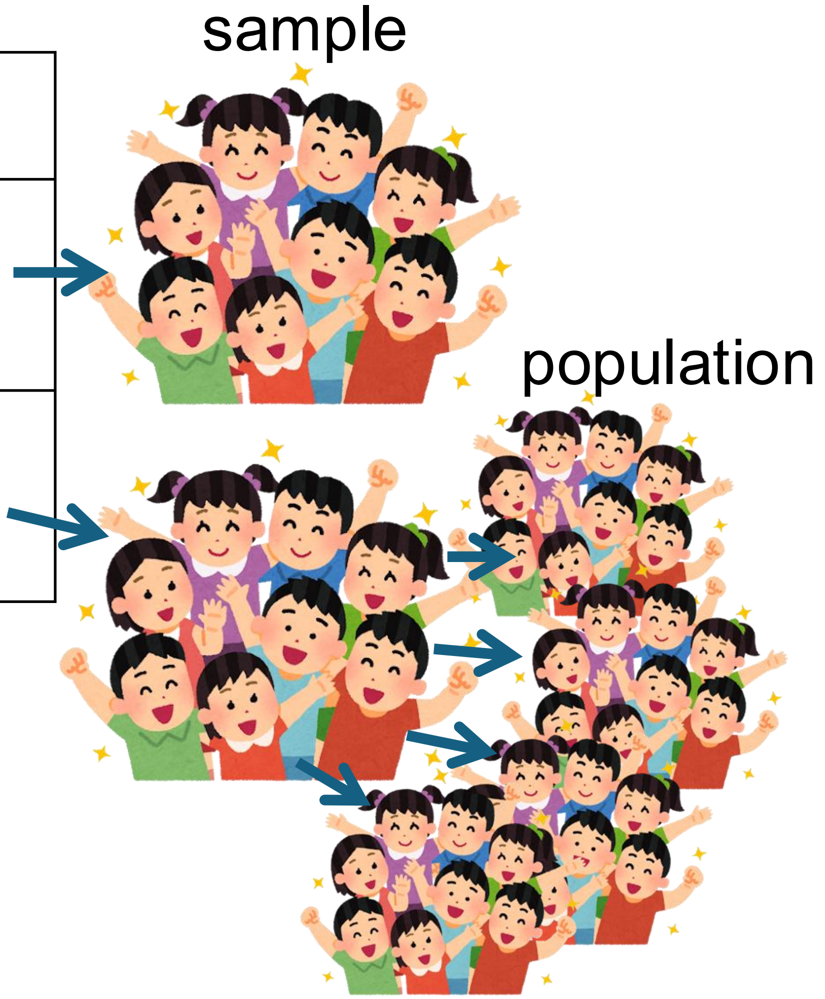
Group level



1st level analysis (within subject)
fMRI scans
Realignment
Normalization
Smoothing
Design matrix
GLM
(General Linear Model)
Parameter estimates
Contrast image
voxel
X
Statistical inference
(RFT)
Subject 1
X
beta
2.3
Subject 2
X
beta
4.6
Subject 3
X
beta
1.8
Subject N
nature
2nd level analysis (across subjects)
Example voxel X
(contrast estimate)
Design matrix
Sub 1 (2.3) 1
Sub 2 (4.6) 1
Sub 3 (1.8) 1
⋮ 1
Sub N (3.1) 1
1 predictor
Second-level GLM
(General Linear Model)
$X\beta + \varepsilon$
Group statistical map
SPM{t}
The effect of seeing
a picture of nature
Numbers in parentheses are example contrast estimates (e.g., beta or A–B)
at voxel X from each subject.

Types of second-level models

	Question	Application
Fixed Effects (FFX)	Is there an effect in these sample subjects?	Combine repeated measures within subjects
Mixed Effects (MFX)	Is there an effect in the population?	Group analysis of fMRI



FFX: $\beta_i = \beta_{common} X_{grp} + \epsilon_i$

MFX: $\beta_i = \beta_{population} X_{grp} + u_i + \epsilon_i$

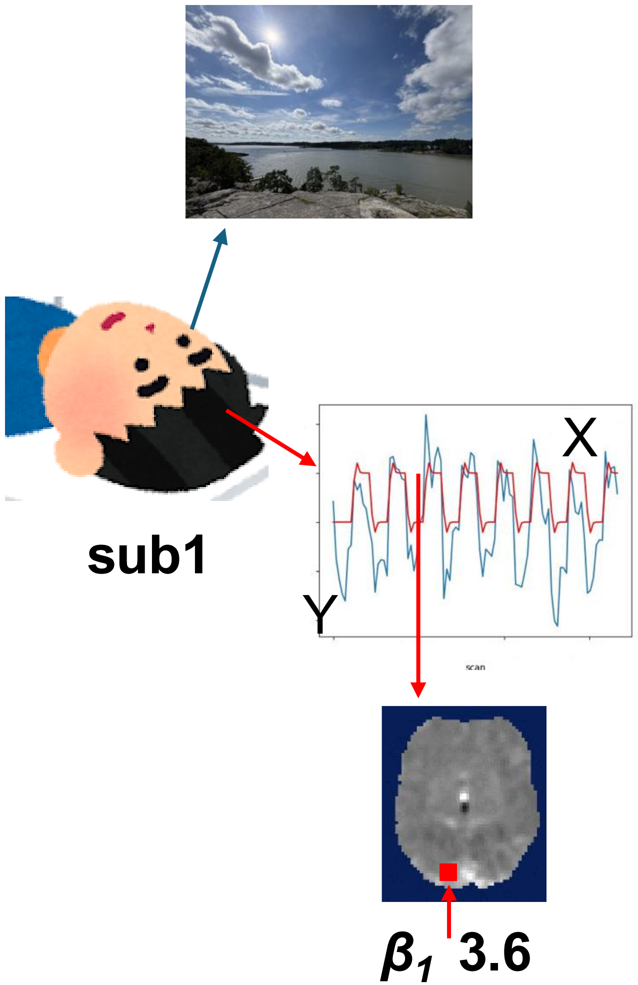
u = between-subject deviation

ϵ = within-subject error

Mixed effects model in fMRI

1st level GLM (within each subject)

$$Y_{sub} = \boldsymbol{\beta}_{sub} X_{sub} + \epsilon_{sub}$$



Y = measured BOLD time-series at each voxel

X = Predicted BOLD response based on the experimental condition

$\boldsymbol{\beta}$ = how strongly the predicted response is expressed in the measured BOLD signal

=> **representing the task-related effect.**

ϵ = unexplained variation

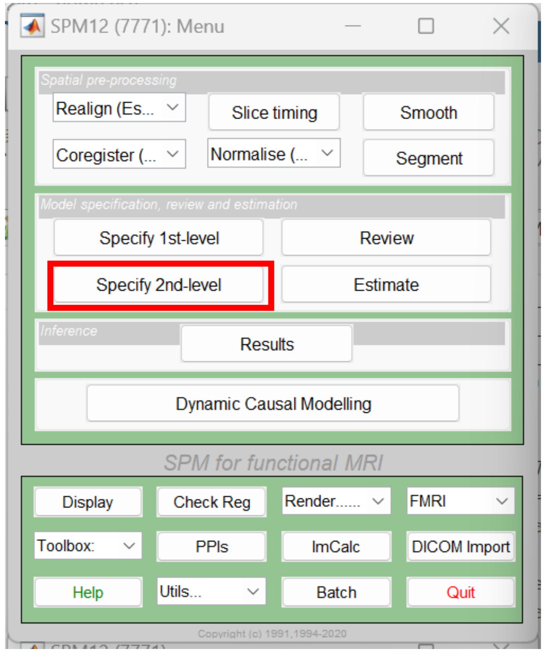
Mixed effects model in fMRI

1st level GLM (within each subject)

$$Y_{sub} = \beta_{sub} X_{sub} + \epsilon$$

2nd level GLM (between subject variance)

$$\beta_{sub} = \beta_{grp} X_{grp} + u_{grp}$$



β_{sub} = each sub's beta estimate at each voxel from 1st level analysis
$(\beta_{1,} \beta_{2,} \cdots, \beta_N)$

$\boldsymbol{\beta_{grp}}$ = Estimated group-level effect at each voxel
=> represents task-related effect at the population level

X_{grp} = group level regressor values: the group/design (1,0, age etc)

u_{grp} = second level residual (between-subject deviation)

Mixed effects model in fMRI

1st level model (within subject variance estimation)

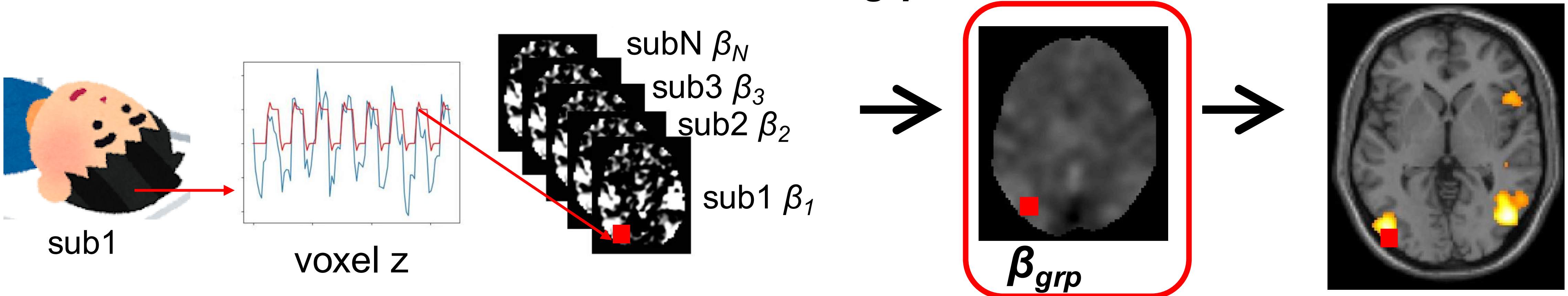
$$Y_{sub} = \beta_{sub} X_{sub} + \epsilon_{sub}$$

2nd level model (between subject variance)

$$\beta_{sub} = \boldsymbol{\beta}_{\boldsymbol{grp}} X_{grp} + u_{grp}$$

$$Y_{sub} \longrightarrow \beta_{sub} \longrightarrow \boldsymbol{\beta}_{\boldsymbol{grp}} \longrightarrow \text{SPM}\,\{t\}$$

statistical inference



Topics

01

Overview

Second-level GLM

→

02

Second-level designs

→

03

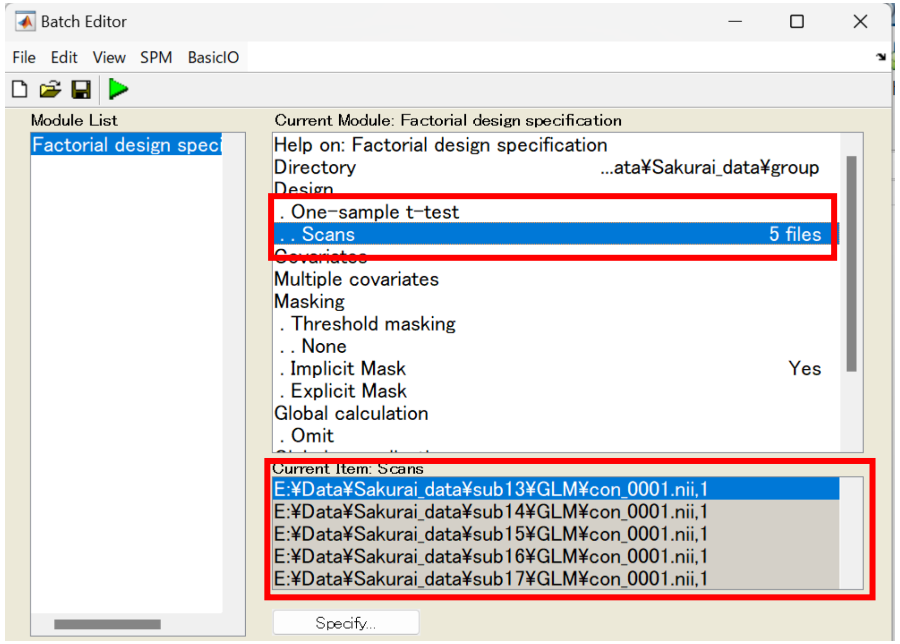
Multiple comparisons problem

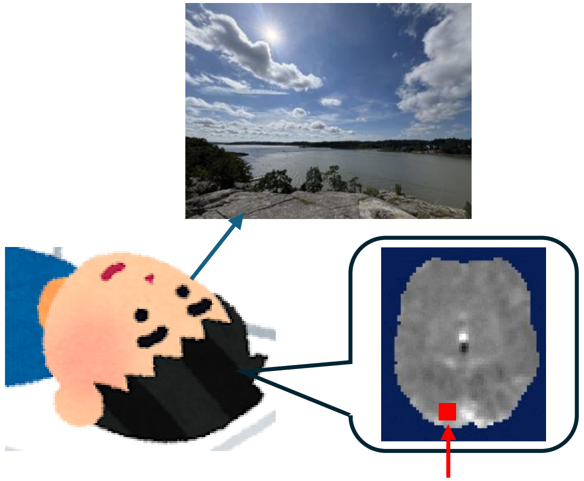
One sample T-test

Test the effect of one condition or a difference between conditions

Null hypothesis $\mathbf{H_0}$: $\beta_{group} = 0$

$$\begin{array}{c} \text{sub1} \\ \text{sub2} \\ \text{sub3} \\ \vdots \end{array} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_N \end{bmatrix} = \beta_{group} X_{group} \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + u_{grp}$$





sub1 β **3.6**

β_N
Subject-specific estimate of the condition related BOLD response

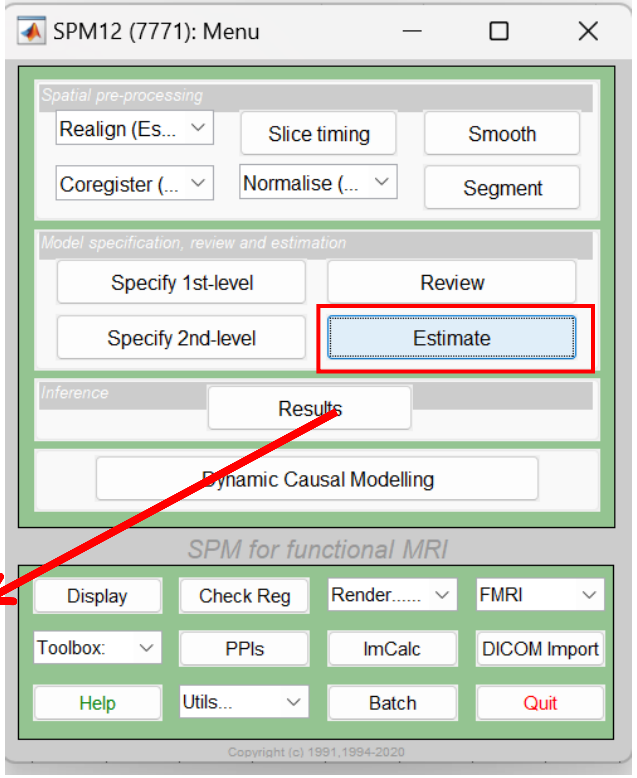
β_{group}
Mean task-related effect at the group level

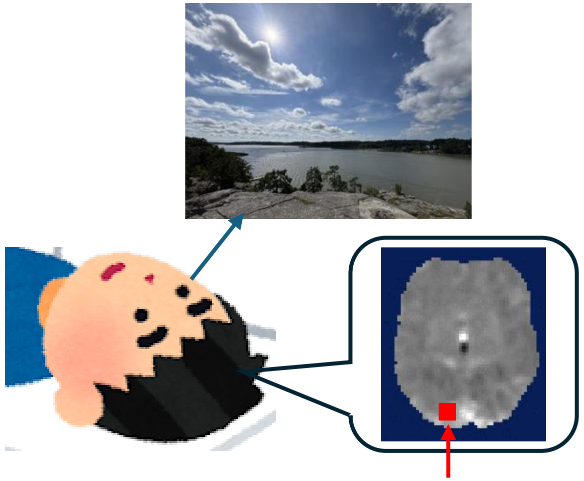
One sample T-test

Test the effect of one condition or a difference between conditions

Null hypothesis $\mathbf{H_0}$: $\beta_{group} = 0$

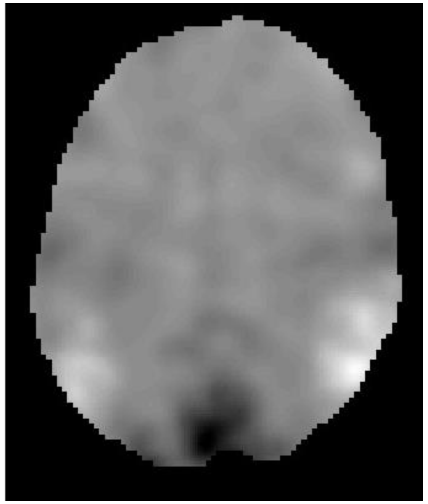
$$\begin{array}{c} \text{sub1} \\ \text{sub2} \\ \text{sub3} \\ \vdots \end{array} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_N \end{bmatrix} = \boxed{\beta_{group}} X_{group} \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + u$$





sub1 **β_1 3.6**

β_{group}

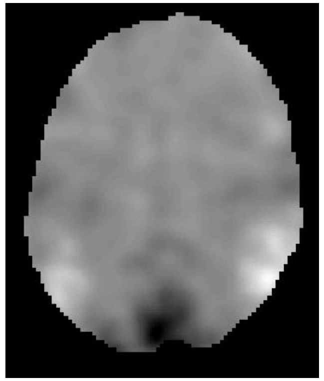


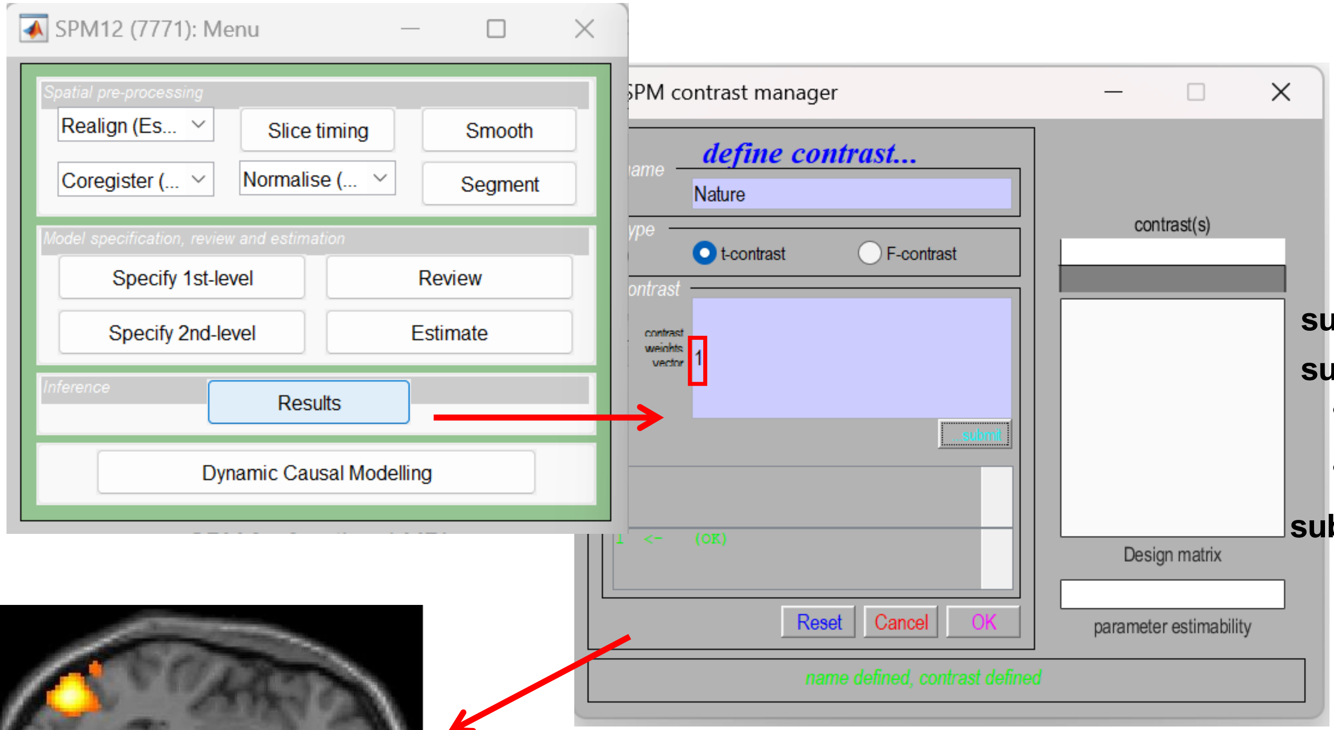
One sample T-test

Test the effect of one condition or a difference between conditions

Null hypothesis $\mathbf{H_0}$: $\beta_{group} = 0$

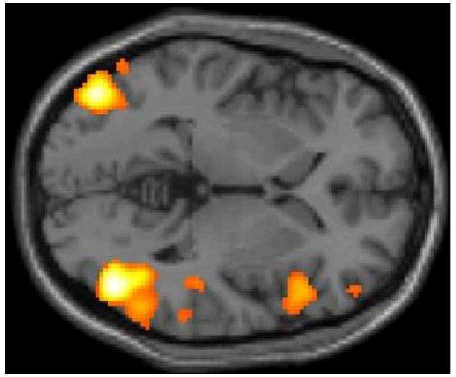
$$\begin{array}{c} \text{sub1} \\ \text{sub2} \\ \text{sub3} \\ \vdots \end{array} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_N \end{bmatrix} = \beta_{group} X_{group}$$





sub1
sub2
•
•
subN

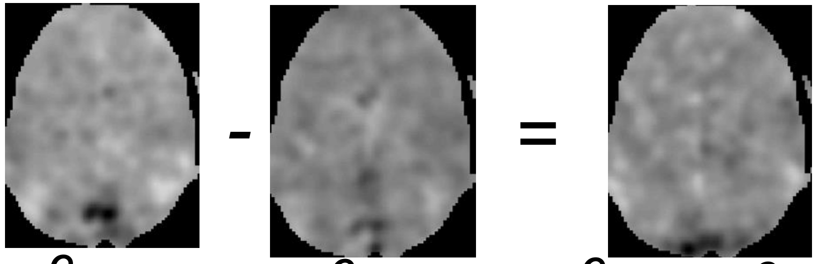
sub1



Pass the statistical threshold

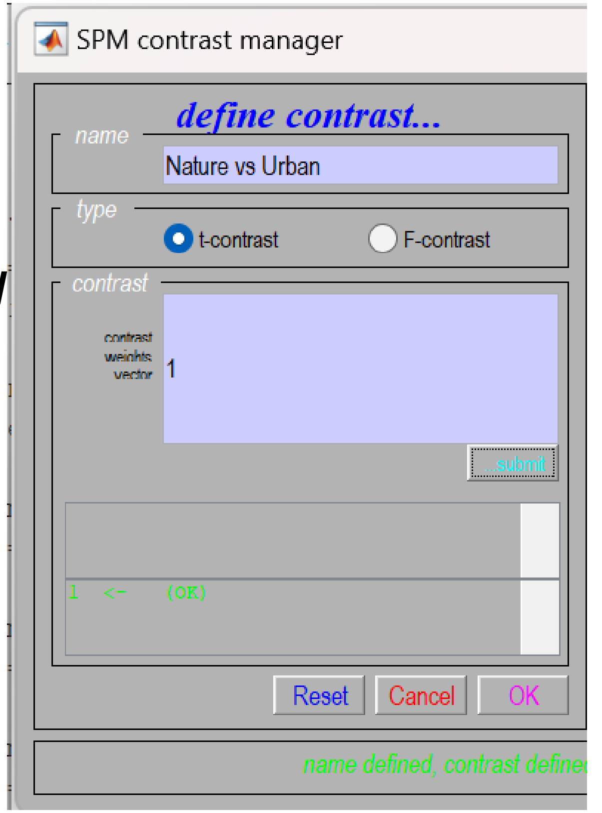
One sample T-test

Test the effect of one condition or a difference between conditions



β_{nat1} - β_{urb1} = $\beta_{nat1} - \beta_{urb1}$

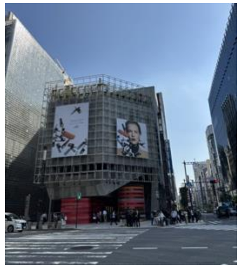
$$\begin{bmatrix} \beta_{nat1} - \beta_{urb1} \\ \beta_{nat2} - \beta_{urb2} \\ \beta_{nat3} - \beta_{urb3} \\ \vdots \\ \beta_{natN} - \beta_{urbN} \end{bmatrix} = \beta_{group} X_{group} \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + u$$



nature

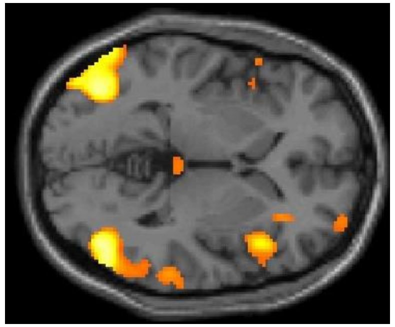
or

urban



sub1

Null hypothesis: **H_0:** $\beta_{nat} - \beta_{urb} = 0$

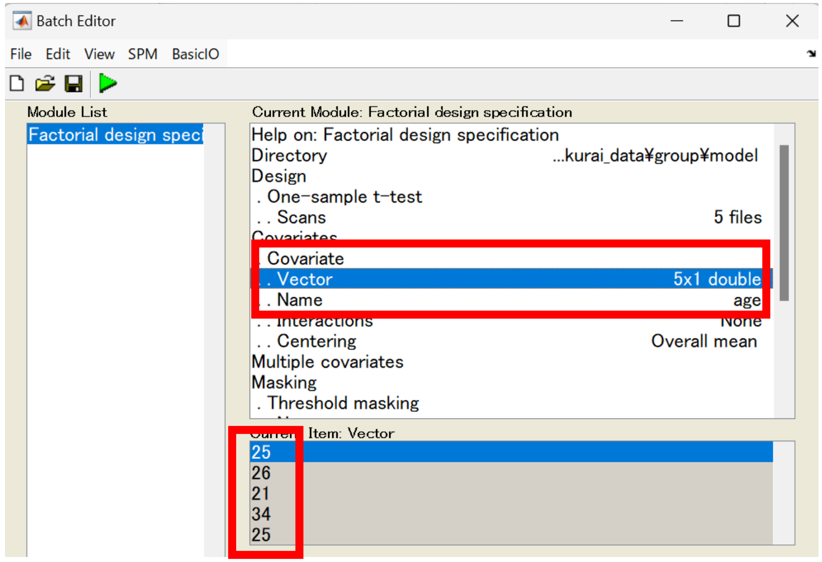


Nature > Urban

One sample T-test with covariate

Test the effect of covariate on task-related brain response

$$\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_N \end{bmatrix} = \beta_{group} X_{group} \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \boxed{\beta_{age} \begin{array}{cc} & \text{age} \\ \begin{bmatrix} X_{cov1} \\ X_{cov2} \\ X_{cov3} \\ \vdots \\ X_{covN} \end{bmatrix} & \begin{array}{c} 28 \\ 35 \\ 21 \\ \\ \\ \end{array} \end{array}} + u_N$$



β_{age}

Effect of the second-level covariate on the first-level estimates

e.g., how first level effects vary with age or behavioral scores

$X_{covariate}$

Subject-specific covariate values

e.g., age, behavioral score, questionnaire score

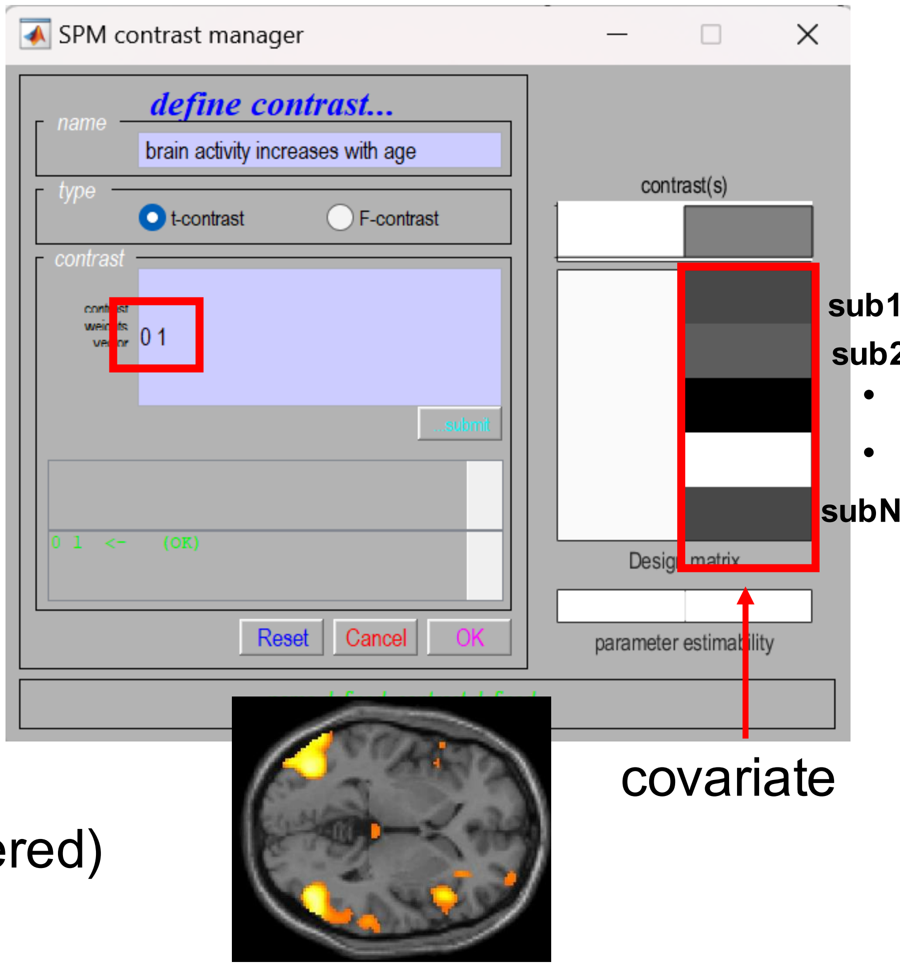
$\beta_{age} > 0$ => task-related effect increases with age

$\beta_{age} < 0$ => task-related effect decreases with age

One sample T-test with covariate

Test the effect of covariate on task-related brain response

$$\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_N \end{bmatrix} = \beta_{group} X_{group} \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \beta_{age} \begin{array}{c} \\ \begin{bmatrix} X_{cov1} \\ X_{cov2} \\ X_{cov3} \\ \vdots \\ X_{covN} \end{bmatrix} \end{array} \begin{array}{c} \text{age} \\ 28 \\ 35 \\ 21 \\ \\ \\ \end{array}$$



[0 1] whether the task-related effect increases with age

[0 -1] whether the task-related effect decreases with age

[1 0] the task-related effect at the mean age (when mean-centered)

[0 1] these brain regions response increases with age

Two sample T-test

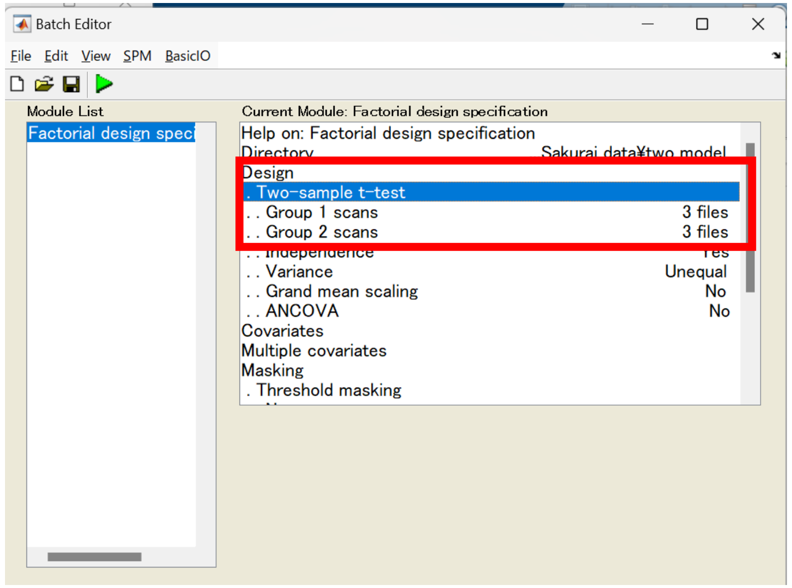
Test the difference between two independent groups

$$\begin{array}{l} \text{young} \\ \text{young} \\ \text{elderly} \\ \text{elderly} \\ \\ \\ \end{array}\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \vdots \end{bmatrix} = \underset{\text{young}}{\beta_{young} X_{young}} \begin{bmatrix} \text{sub1} & 1 \\ \text{sub2} & 1 \\ \text{sub3} & 0 \\ \text{sub4} & 0 \\ \vdots & \end{bmatrix} + \underset{\text{elderly}}{\beta_{elderly} X_{elderly}} \begin{bmatrix} \text{sub1} & 0 \\ \text{sub2} & 0 \\ \text{sub3} & 1 \\ \text{sub4} & 1 \\ \vdots & \end{bmatrix} + u_N$$

sub1
young

sub3
elderly

Null hypothesis
$\mathbf{H_0}$: $\beta_{young} = \beta_{elderly}$

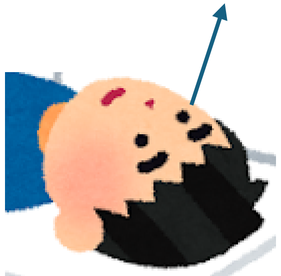


Two sample T-test

Test the difference between two independent groups

$$\begin{array}{l} \text{young} \\ \text{young} \\ \text{elderly} \\ \text{elderly} \\ \\ \end{array} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \vdots \end{bmatrix} = \beta_{young} X_{young} \begin{bmatrix} \text{sub1} & 1 \\ \text{sub2} & 1 \\ \text{sub3} & 0 \\ \text{sub4} & 0 \\ \vdots & \end{bmatrix} + \beta_{elderly} X_{e}$$

elderly

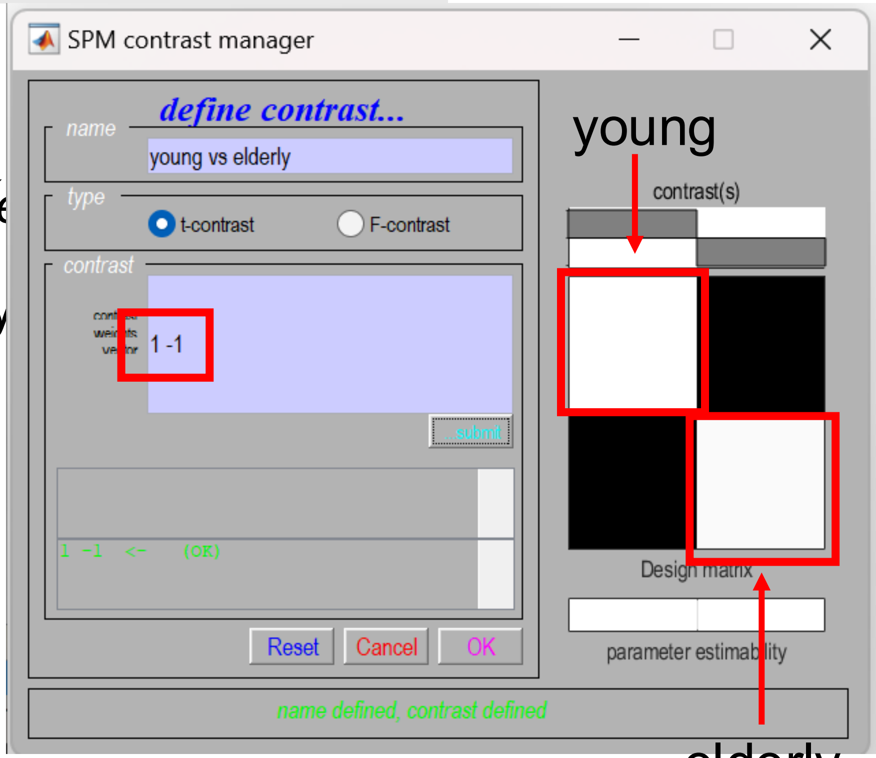


sub1
young

sub3
elderly

Null hypothesis

$$\mathbf{H_0}: \beta_{young} = \beta_{elderly}$$



elderly

[1 -1] test greater task-related effect in young group than in elderly group

[-1 1] test greater task-related effect in elderly group than in young group

The other options

One-way ANOVA

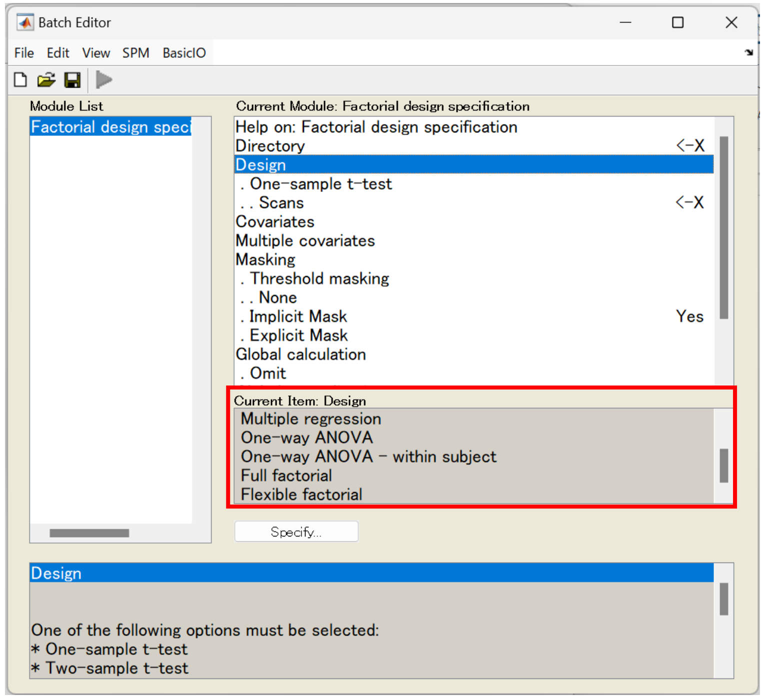
One-way ANOVA – within subject

Full factorial model

Flexible factorial model

Multiple regression

Paired T-test



=> Research question and experimental design

Topics

01

Overview

2^{nd} level GLM

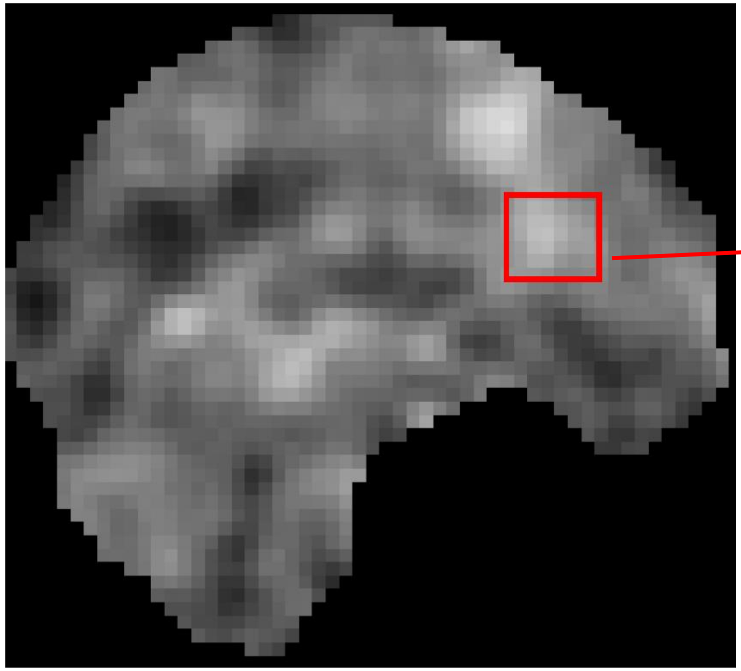
02

2^{nd} level designs

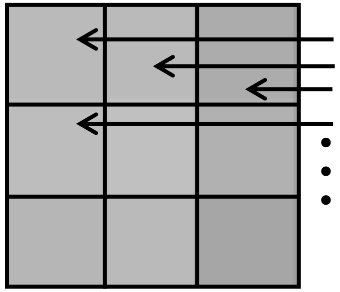
03

Multiple comparisons problem

Multiple comparisons problem



Approx 100k – 300k voxels



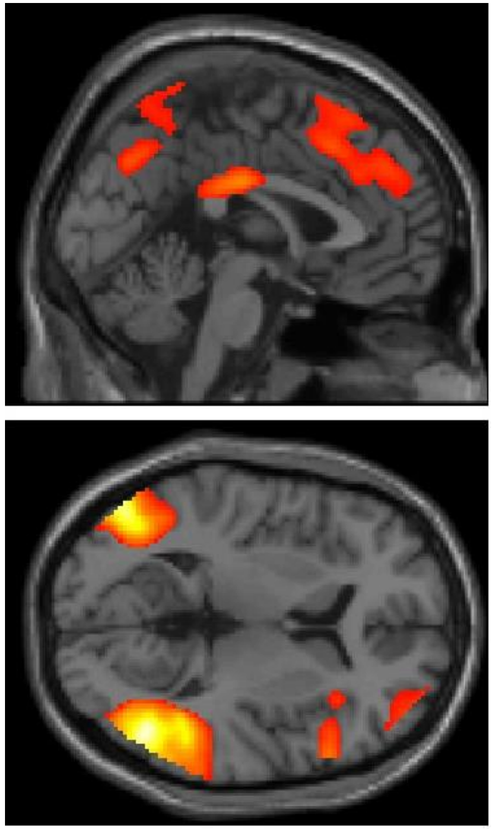
$$\beta_{sub} = \beta_{grp} X_{grp} + u_{grp}$$

***α* = 0.05** → If we run 100 independent test, we could have 5 false positives on average (type I error)

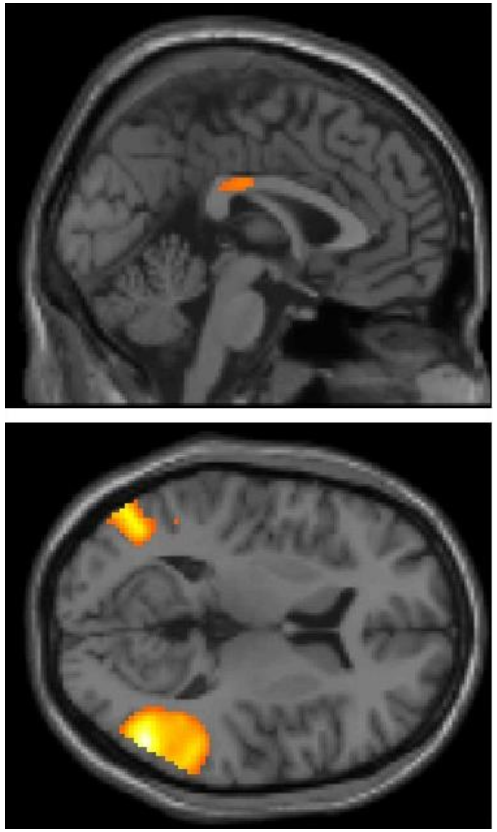
With 100,000 tests at α = .05, we could expect up to **~5,000** false-positive voxels under the simplified example

→ **Multiple-comparison correction is necessary**

Example of correction



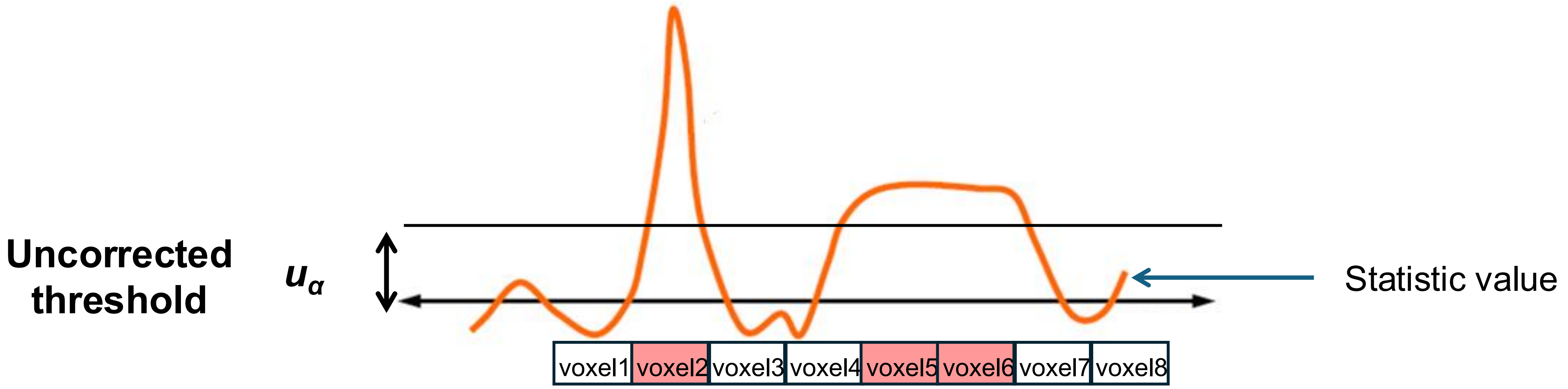
uncorrected
***p* < .001**



Family wise error corrected
***p* < .05**

Voxel-level correction

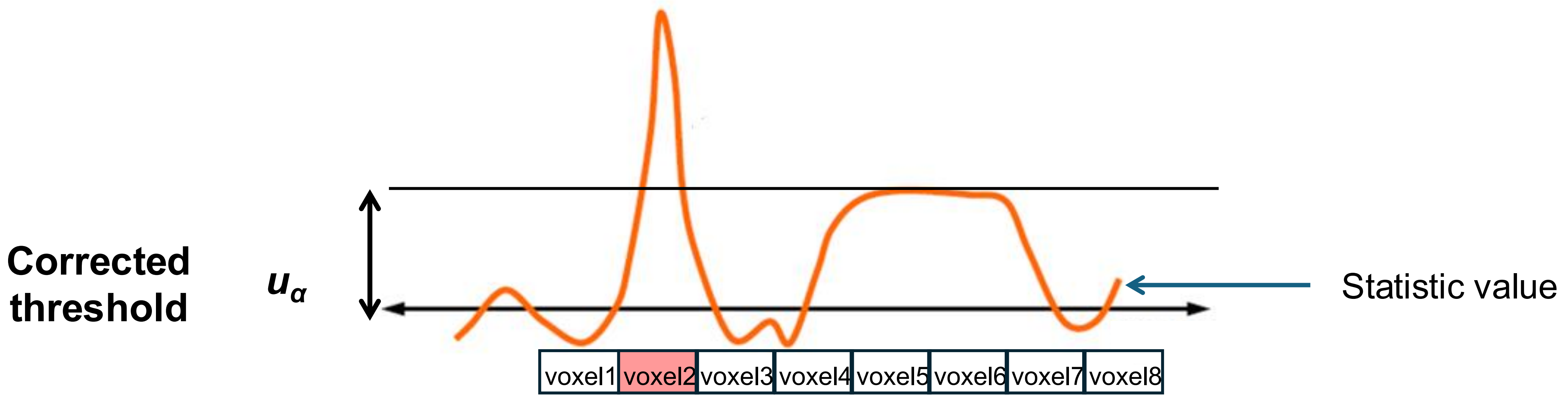
Voxels above α-level threshold u_α are significant



Without correction, voxel 2,5, and 6 are significant

Voxel-level correction

Voxels above α-level threshold u_α are significant



With correction, just voxel2 is significant

How do we set the corrected threshold?

Family-Wise Error Rate (FWE)

FWE-corrected *p* <.05

Keep the probability of having at least one false-positive voxel across the whole brain at 5%.

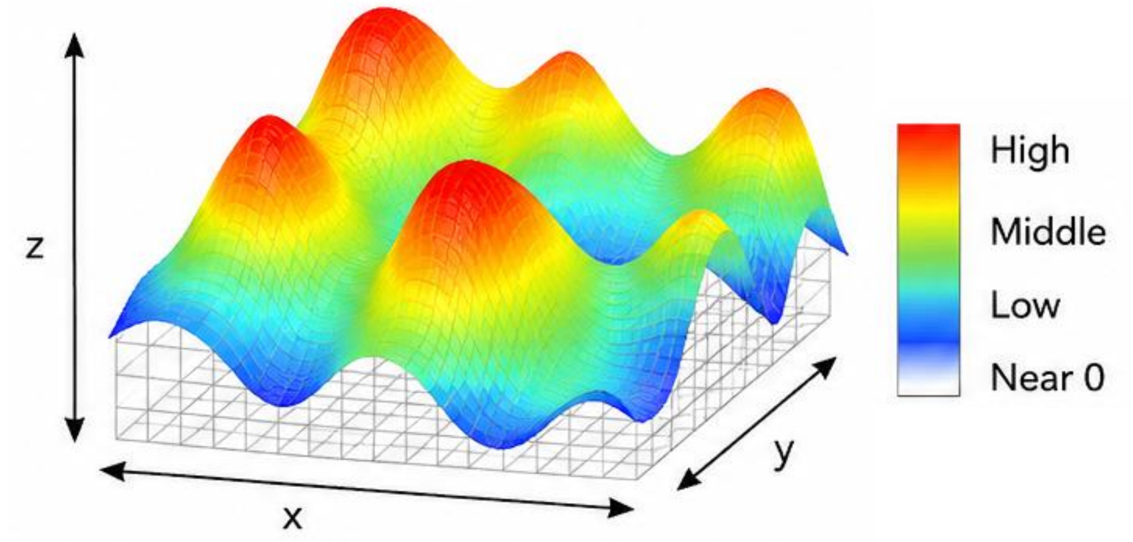
↓

Bonferroni correction

α = 0.05/100,000 (number of voxel)
= 0.0000005

FWE - Random field theory (RFT)

RFT

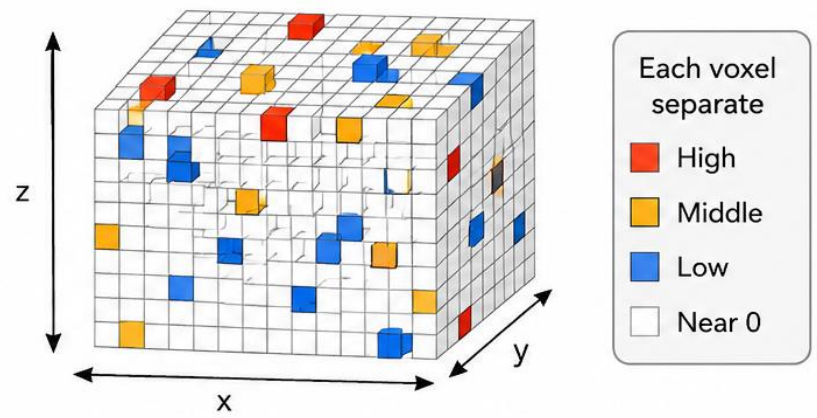


Neighboring voxels are spatially correlated, fewer independent information
Search volume + smoothness controls FWE

↓

More appropriate FWE-corrected threshold

Bonferroni

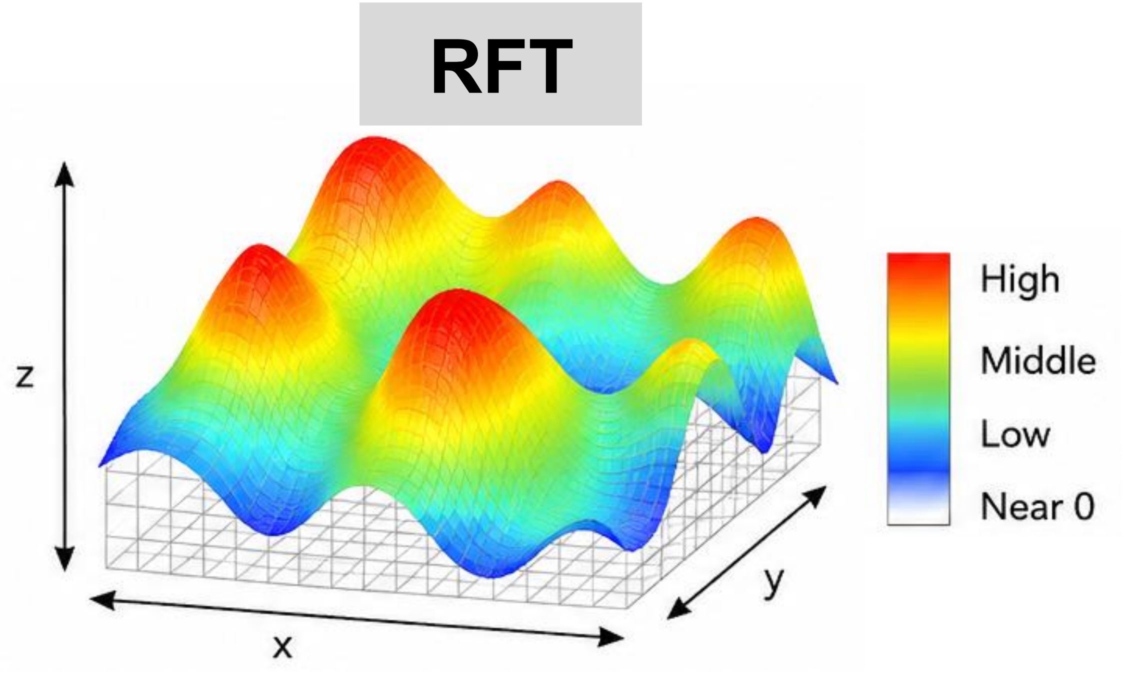


Treats each voxel as independent
Corrects for the total number of voxels

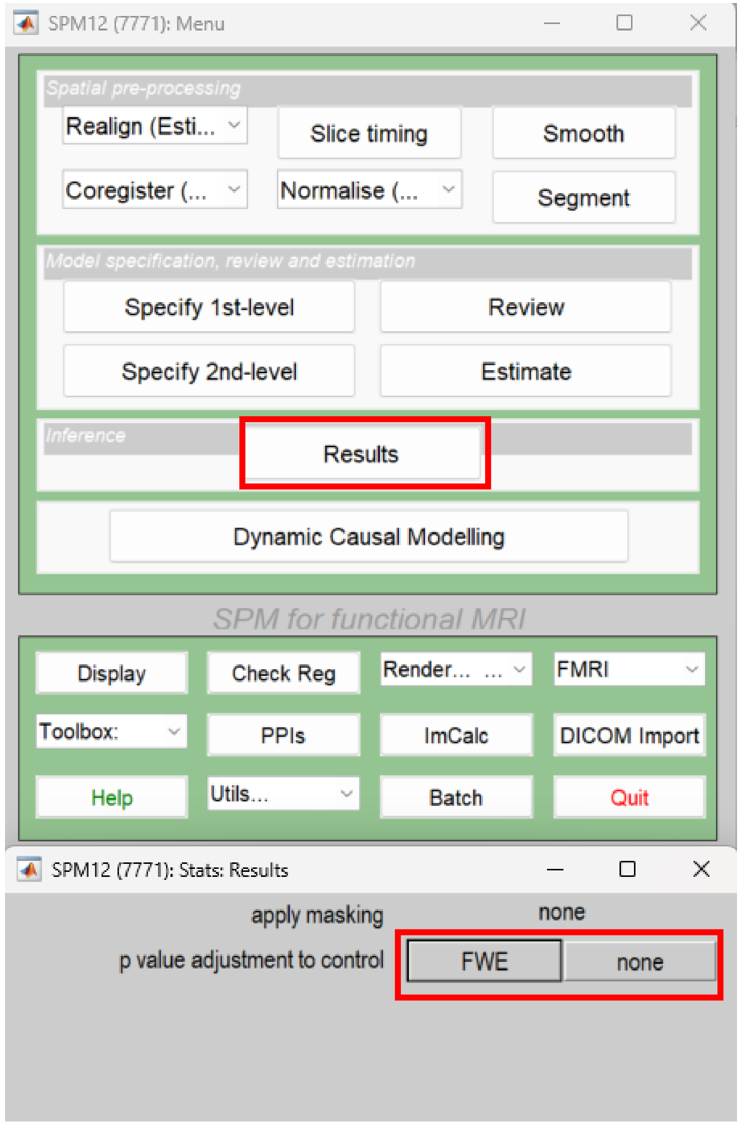
↓

Very conservative

FWE - Random field theory (RFT)



SPM uses RFT for FWE correction



False Discovery Rate (FDR)

FDR controls the expected proportion of false positives at 5% among the voxels declared significant : FDR controlled at q =.05"

↓

Benjamini–Hochberg (BH) procedure

1. Rank all voxel-wise p-values from smallest to largest.

①	②	③	④	⑤
voxel 3	voxel 5	voxel 1	voxel 2	voxel 4
p = .001	p = .003	p = .012	p = .033	p = .064

2. Compare each p value with its BH threshold

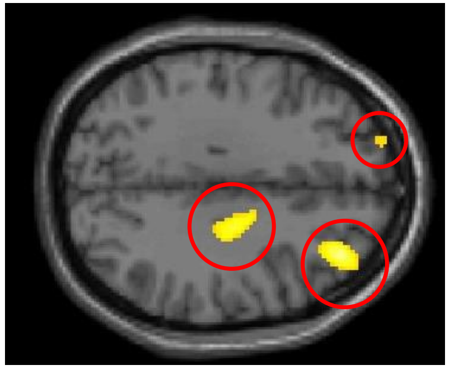
$$\text{BH threshold} = \frac{\text{rank}}{\text{number of test}} \times q$$

3. Find the largest p-value that passes the threshold

	①	②	③	④	⑤
BH threshold	p = .010	p = .020	p = .030	p = .040	p = .050

Cluster-level correction

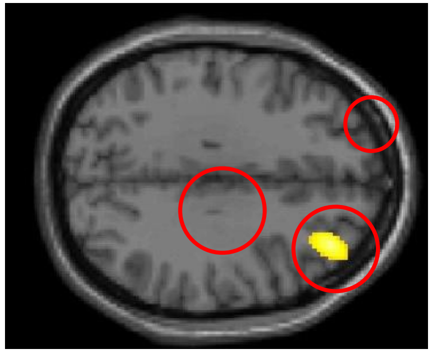
We can use an **uncorrected voxel-wise threshold (e.g., *p* <. 001)**
→ then apply **cluster-level correction**
a cluster = contiguous supra-threshold voxels



Voxel-wise uncorrected *p* < .001

Could clusters of these sizes occur by chance?

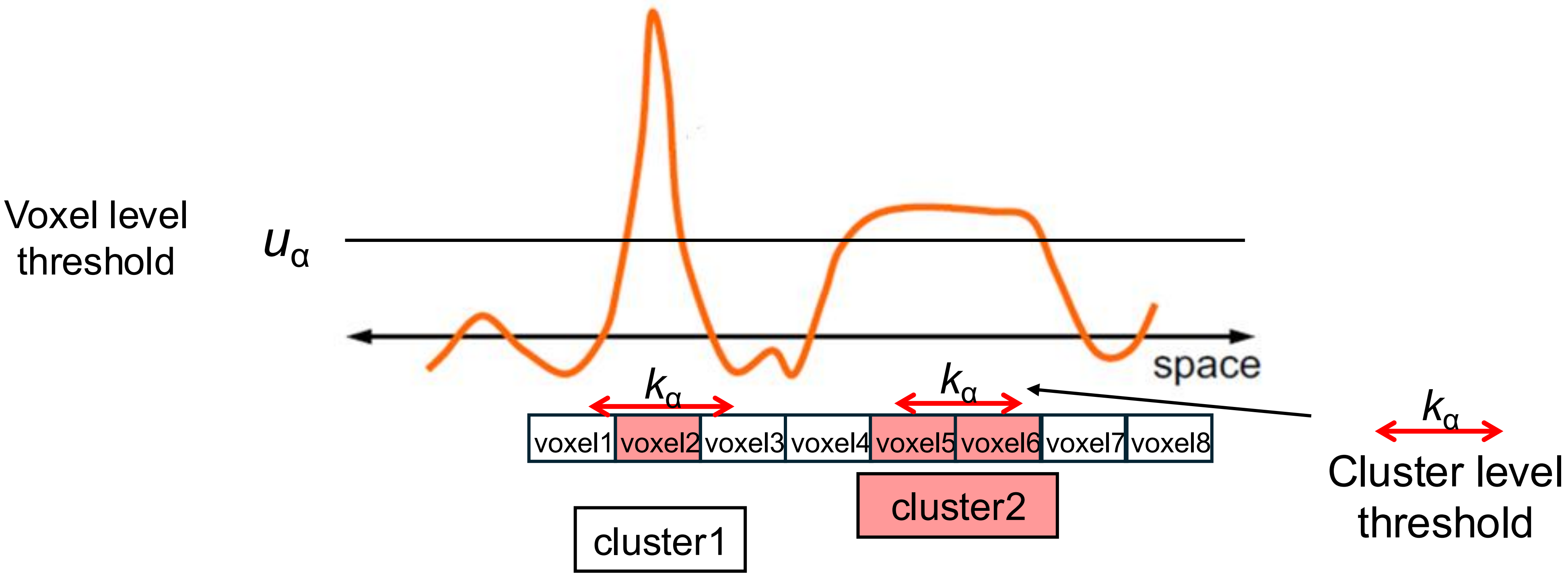
H_0: The observed cluster is not larger than what would be expected by chance.



FWE cluster-wise corrected *p* < .05

Only one cluster survived the correction.
=> Significant at the cluster level

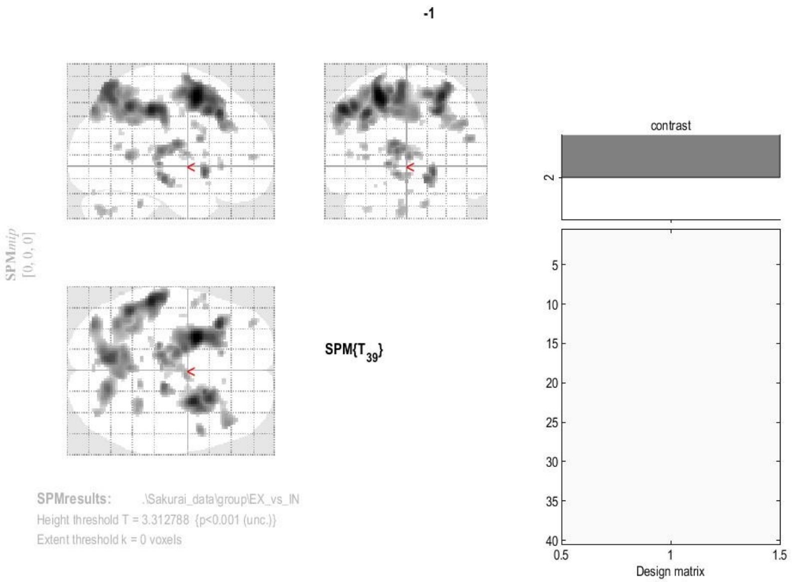
Cluster-level correction



Identify clusters consisting of contiguous voxels that exceed the voxel-level threshold
Test whether each cluster exceeds the cluster-extent threshold k_α

→ Cluster wise FWE/FDR correction

Examples



Statistics: *p-values adjusted for search volume*

set-level p	set-level c	cluster-level $p_{FWE-corr}$	cluster-level $q_{FDR-corr}$	cluster-level k_E	cluster-level p_{uncorr}	peak-level $p_{FWE-corr}$	peak-level $q_{FDR-corr}$	peak-level T	peak-level (Z_E)	peak-level p_{uncorr}	mm	mm	mm
0.000	30	0.000	0.000	1551	0.000	0.047	0.409	5.39	4.63	0.000	-24	6	54
						0.170	0.409	4.87	4.28	0.000	-36	24	40
						0.387	0.450	4.49	4.00	0.000	-16	-4	58
		0.000	0.000	1484	0.000	0.100	0.409	5.09	4.43	0.000	-54	-32	44
						0.192	0.409	4.82	4.24	0.000	-46	-50	46
						0.635	0.460	4.19	3.78	0.000	-34	-60	58
		0.000	0.000	893	0.000	0.122	0.409	5.01	4.38	0.000	28	2	54
						0.213	0.409	4.77	4.21	0.000	24	16	42
						0.348	0.450	4.54	4.04	0.000	32	8	60
		0.000	0.000	1576	0.000	0.269	0.419	4.67	4.13	0.000	-8	-68	56
						0.450	0.450	4.41	3.94	0.000	18	-68	56
						0.528	0.460	4.31	3.88	0.000	-34	-80	44
		0.691	0.524	52	0.175	0.366	0.450	4.52	4.03	0.000	20	12	-6
		0.008	0.007	399	0.001	0.455	0.450	4.40	3.94	0.000	-12	-10	16
						0.509	0.460	4.33	3.89	0.000	-10	-24	10
						0.665	0.460	4.15	3.76	0.000	-18	-26	0
		0.489	0.374	79	0.100	0.459	0.450	4.39	3.94	0.000	16	-22	-12
		0.675	0.524	54	0.167	0.673	0.460	4.14	3.75	0.000	0	-50	6
		0.805	0.607	38	0.243	0.751	0.460	4.05	3.68	0.000	50	8	20
		0.349	0.319	103	0.064	0.765	0.464	4.03	3.66	0.000	44	32	32
						0.979	0.766	3.58	3.31	0.000	28	26	30
		0.438	0.367	87	0.086	0.856	0.569	3.90	3.56	0.000	28	-42	40
		0.952	0.848	16	0.452	0.857	0.569	3.90	3.56	0.000	-12	26	-2
		0.994	0.848	3	0.771	0.892	0.611	3.84	3.52	0.000	-2	-14	-14
		0.970	0.848	12	0.519	0.920	0.629	3.78	3.47	0.000	14	-42	-32
		0.957	0.848	15	0.467	0.933	0.629	3.75	3.45	0.000	-36	52	6
		0.732	0.534	47	0.196	0.935	0.629	3.75	3.44	0.000	-8	-72	-38
		0.931	0.848	20	0.397	0.940	0.629	3.73	3.43	0.000	-38	-60	14
		0.851	0.653	32	0.283	0.940	0.629	3.73	3.43	0.000	-52	4	32
						0.969	0.726	3.64	3.35	0.000	-58	12	26
		0.993	0.848	4	0.729	0.977	0.766	3.60	3.32	0.000	-8	-40	-32

table shows 3 local maxima more than 8.0mm apart

Height threshold: T = 3.31, p = 0.001 (0.999)
Extent threshold: k = 0 voxels
Expected voxels per cluster, <k> = 30.000
Expected number of clusters, <c> = 0.79
FWEp: 5.367, FDRp: Inf, FWEc: 399, FDRc: 399

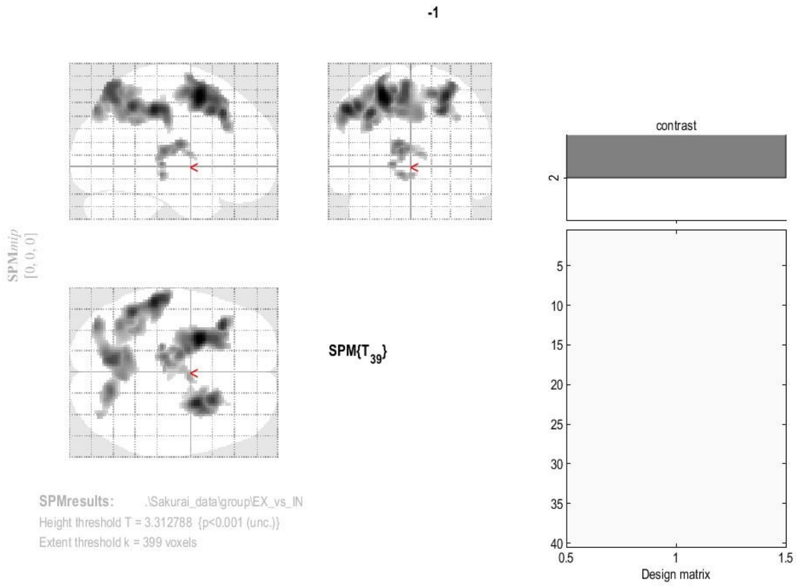
Degrees of freedom = [1.0, 39.0]
FWHM = 13.7 13.7 13.0 mm mm mm; 6.9 6.8 6.5 {voxels}
Volume: 1343488 = 167936 voxels = 505.5 resels
Voxel size: 2.0 2.0 2.0 mm mm mm; (resel = 305.88 voxels)
Page 1

FWEc: 399, FDRc: 399

SPM12 (7771): Stats: Results

apply masking	none
threshold {T or p value}	0.001
& extent threshold {voxels}	399

The cluster size is equal to or larger than 399, it's significant



Statistics: *p-values adjusted for search volume*

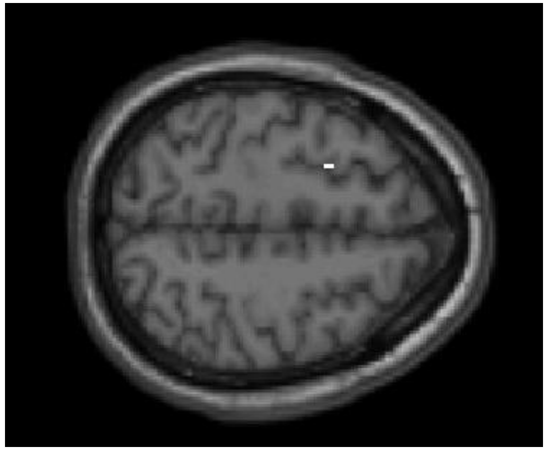
set-level p	set-level c	cluster-level $p_{FWE-corr}$	cluster-level $q_{FDR-corr}$	cluster-level k_E	cluster-level p_{uncorr}	peak-level $p_{FWE-corr}$	peak-level $q_{FDR-corr}$	peak-level T	peak-level (Z_E)	peak-level p_{uncorr}	mm	mm	mm
0.000	5	0.000	0.000	1551	0.000	0.047	0.409	5.39	4.63	0.000	-24	6	54
						0.170	0.409	4.87	4.28	0.000	-36	24	40
						0.387	0.450	4.49	4.00	0.000	-16	-4	58
		0.000	0.000	1484	0.000	0.100	0.409	5.09	4.43	0.000	-54	-32	44
						0.192	0.409	4.82	4.24	0.000	-46	-50	46
						0.635	0.460	4.19	3.78	0.000	-34	-60	58
		0.000	0.000	893	0.000	0.122	0.409	5.01	4.38	0.000	28	2	54
						0.213	0.409	4.77	4.21	0.000	24	16	42
						0.348	0.450	4.54	4.04	0.000	32	8	60
		0.000	0.000	1576	0.000	0.269	0.419	4.67	4.13	0.000	-8	-68	56
						0.450	0.450	4.41	3.94	0.000	18	-68	56
						0.528	0.460	4.31	3.88	0.000	-34	-80	44
		0.008	0.007	399	0.001	0.455	0.450	4.40	3.94	0.000	-12	-10	16
						0.509	0.460	4.33	3.89	0.000	-10	-24	10
						0.665	0.460	4.15	3.76	0.000	-18	-26	0

table shows 3 local maxima more than 8.0mm apart

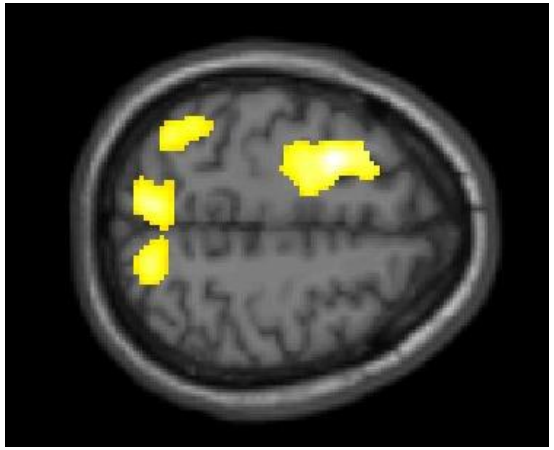
Height threshold: T = 3.31, p = 0.001 (0.999)
Extent threshold: k = 399 voxels, p = 0.001 (0.008)
Expected voxels per cluster, <k> = 30.000
Expected number of clusters, <c> = 0.01
FWEp: 5.367, FDRp: Inf, FWEc: 399, FDRc: 399

Degrees of freedom = [1.0, 39.0]
FWHM = 13.7 13.7 13.0 mm mm mm; 6.9 6.8 6.5 {voxels}
Volume: 1343488 = 167936 voxels = 505.5 resels
Voxel size: 2.0 2.0 2.0 mm mm mm; (resel = 305.88 voxels)

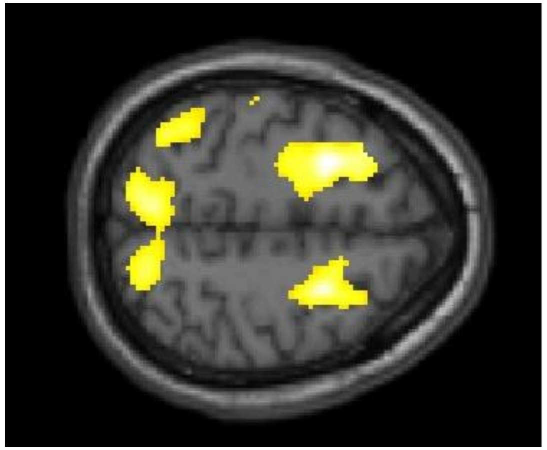
Examples



Voxel-wise
FWE corrected p < .05



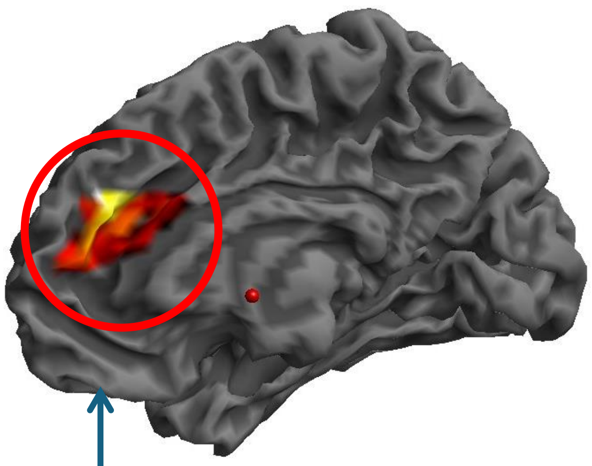
Voxel-wise
uncorrected p < .001
+
Cluster-wise
FWE corrected p < .05



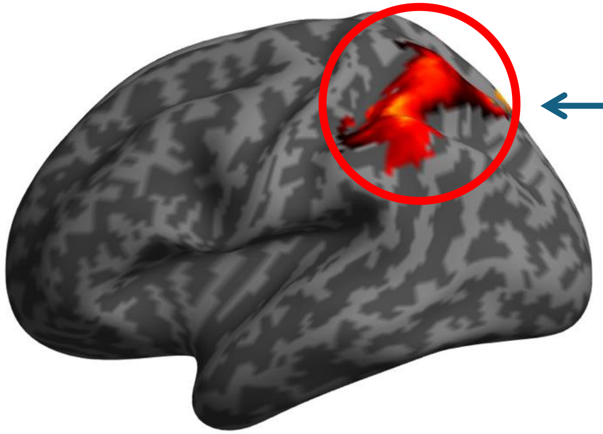
Voxel-wise
uncorrected p < .001

No cluster-level
correction

ROI (region of interest)



Anterior cingulate cortex
Approx 1100voxels



Posteior parietal cortex
Approx 3000voxels

The number of voxels being tested is much smaller,
the corrected threshold can be less strict than in a whole-brain analysis

Summary

If you want to generalize your result to the population,
conduct second level analysis

Choose the appropriate second-level designs for your research design,
one sample t-test, two-sample, etc.

Correct for multiple comparisons:
voxel-wise FWE - corrected $p < .05$
or
voxel-wise uncorrected $p < .001$ with cluster-wise correction $p < .05$

Thank you so much!!!