Brain-Picking

Post-Selection Inference for fMRI Data

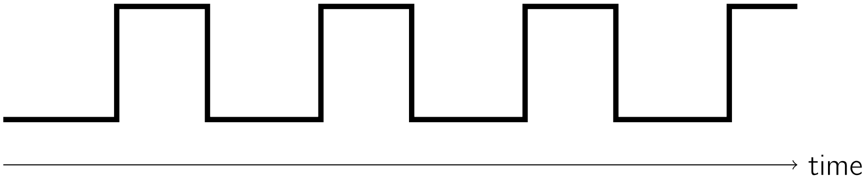
Aldo Solari

Joint work with: Livio Finos and Jelle Goeman

Padua, October 9, 2015

80 years after Bonferroni

An fMRI Experiment



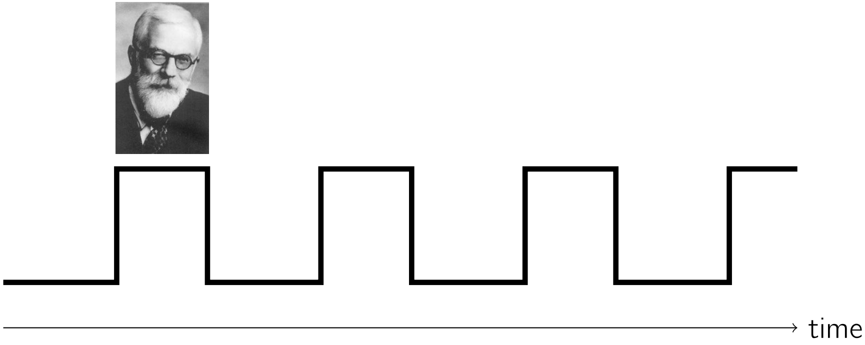
Stimulus

Famous or Nonfamous photograph

Task

Raise your hand at each Famous photograph

An fMRI Experiment



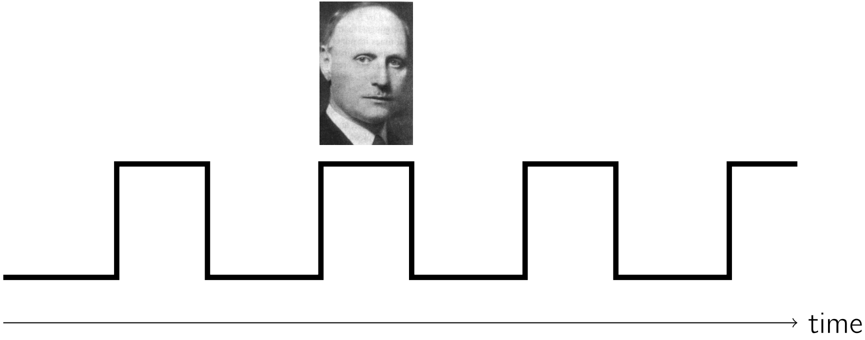
Stimulus

Famous or Nonfamous photograph

Task

Raise your hand at each Famous photograph

An fMRI Experiment



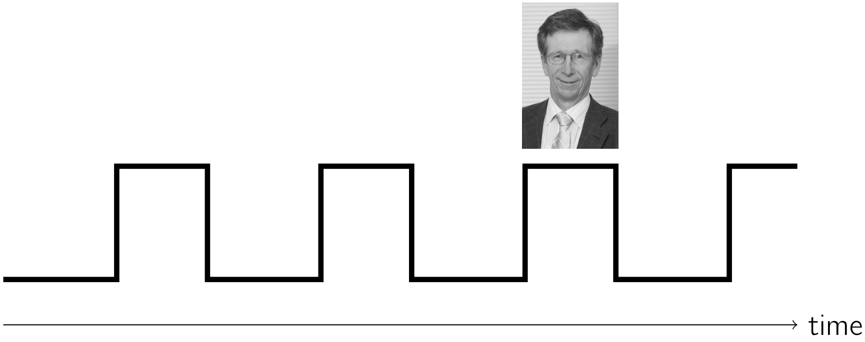
Stimulus

Famous or Nonfamous photograph

Task

Raise your hand at each Famous photograph

An fMRI Experiment



Stimulus

Famous or Nonfamous photograph

Task

Raise your hand at each Famous photograph

Simes' inequality

Prof. Robert John Simes

Famous inequality

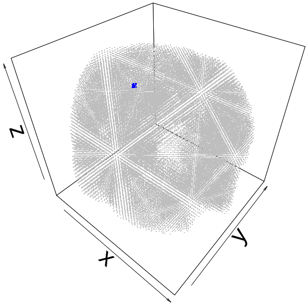
An Improved Bonferroni Procedure for Multiple Tests of Significance

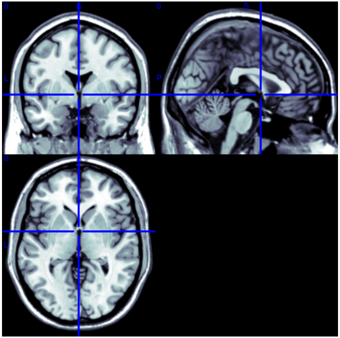
Biometrika 1986

Linked to

- Hommel procedure (for FWER control)
- Benjamini-Hochberg procedure (for FDR control)

Volumetric pixels





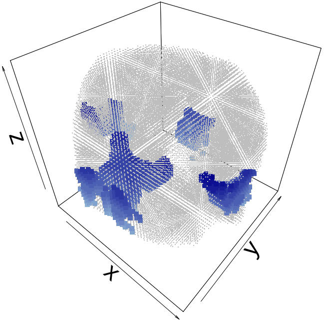
Response

Brain activity measured at each voxel

Question

Which regions respond differently to Famous and Nonfamous?

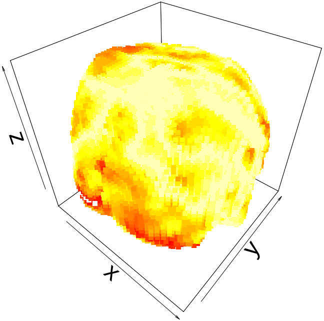
Voxel null hypothesis



H_v : voxel v is OFF, $\qquad v = 1, \ldots, m = 39787$

e.g. voxel v responds equally to Famous and Nonfamous photos

Activation map



For each voxel v, a p-value p_v testing H_v

Color $\to$ red as p-value $\to 0$

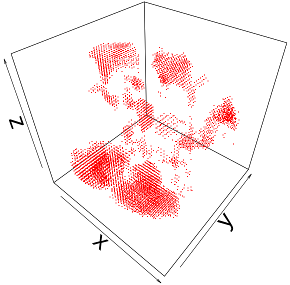
Popular approaches

Voxel-wise inference
Statements about voxels

Cluster-wise inference
Statements about clusters of voxels

Small volume correction
Statements about voxels $\in$ Region of interest

Voxel-wise inference



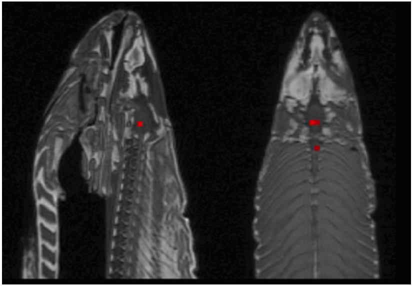
Statement

Voxel v is ON if $\{p_v \leq t\}$, OFF otherwise

Need for multiple testing correction

Set $t = t_\alpha$ to control the flood of false ON voxels (type I errors)

Risk of red herrings



Bennett et al. (2009)

fMRI scan of a *dead* salmon, $m = 8064$ voxels

Heuristic threshold

$t = 0.001 \rightarrow 16$ voxels ON $= 16$ type I errors

As expected

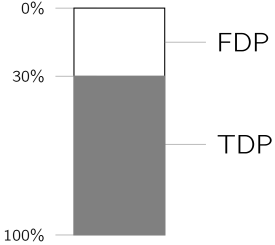
$\mathrm{E}(\#\ \text{type I errors}) = m \times t \approx 8$

False Discovery Proportion

$$\text{FDP} = \frac{\#\text{ false activations}}{\#\text{ activations}}$$

and $\text{FDP} = 0$ if $\#\text{ activations} = 0$

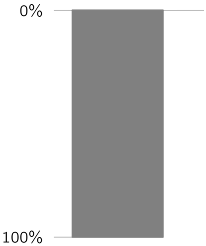
e.g. with $t = 0.001$ we may have



Familywise Error Rate control at 5%

Use $t_{5\%}^{\mathrm{FWER}}$ (e.g. Bonferroni $t_{5\%}^{\mathrm{FWER}} = 0.05/m$) to guarantee that

$\Pr(\mathrm{FDP} > 0) \leq 5\%$



$\mathrm{FDP} = 0\%$

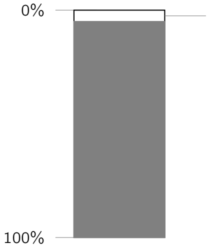
with probability $\geq 95\%$

Very strict $\rightarrow$ low power

False Discovery Rate control at 5%

Use $t^{\mathrm{FDR}}_{5\%}$ to guarantee that

$$\mathrm{FDR} = \mathrm{E}(\mathrm{FDP}) \leq 5\%$$

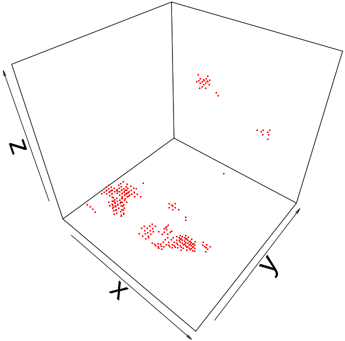


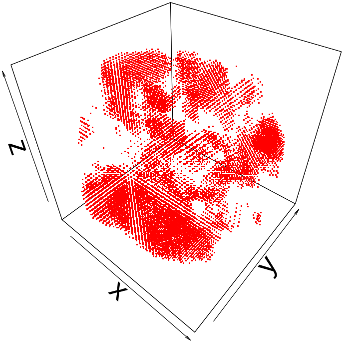
FDP $\leq 5\%$

on average

More powerful

FWER and FDR activations





Hommel
269 voxels ON

Benjamini-Hochberg
12747 voxels ON

Units of interest

The unit voxel

- Defined by spatial resolution, not a primary neural entity
- Interest on regions of activity
- Pre-processing: spatial smoothing

Active voxels

- Intermediary result, not an end result
- Use: screening

Post-processing

- Elimination of isolated active voxels
- Aggregation into clusters of contiguous active voxels
- Selection of active voxels for further analysis

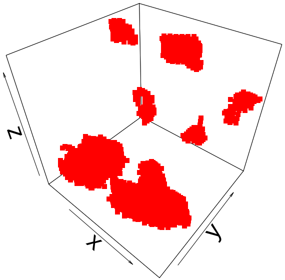
FWER control allows for simultaneous inference

- Robust against aggregation
- Robust against selection

FDR control doesn't allow for simultaneous inference

- Not robust against aggregation
- Not robust against selection

Cluster-wise inference



Cluster definition

$C = \{\text{contiguous voxels } v : p_v \leq t\}$

Statement

Cluster C is ON if $\{\#C \geq c\}$, OFF otherwise

Cluster null hypothesis

Interpretation of active clusters?

- Cluster C is ON $=$ at least one voxel within C is ON
- Locations and % of active voxels within an active cluster?

Cluster null hypothesis

$$H_C : \bigcap_{v \in C} H_v$$

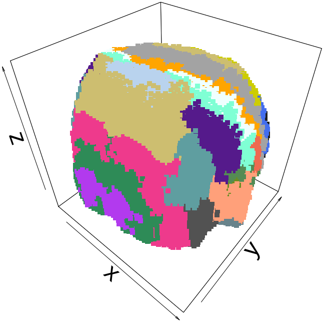
Cluster C is OFF $=$ all voxels within C are OFF

FWER control?

Set $c = c_\alpha^{\mathrm{FWER}}$ for FWER control of false ON clusters

> *I know no formal proof that cluster inference has such strong control of Familywise Error.*
> *T.E. Nichols (2012) NeuroImage*

Regions of interest



- Confirmatory (*a priori* defined)
  e.g. anatomical (Brodmann areas), previous studies, etc.
- Exploratory (*post hoc* selected)
  e.g. small spheres at the peaks of active clusters, etc.

Small volume correction

Single ROI

Choose any R^*

Statement

Voxel $v \in R^*$ is ON if $\{p_v \leq t_\alpha^{\mathrm{FWER}}\}$, OFF otherwise

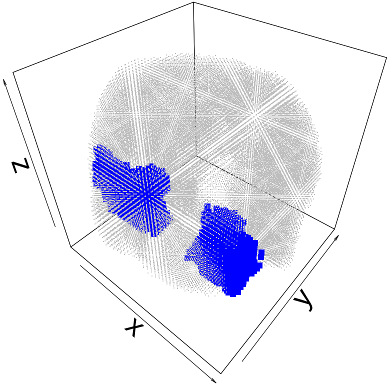
Advantages

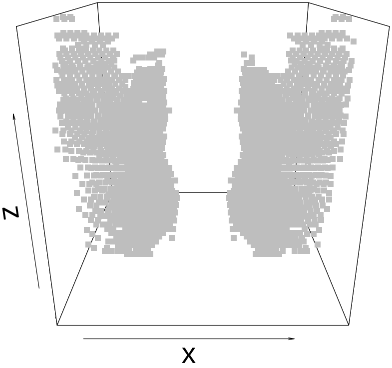
- Better interpretation
- Multiplicity $= \#R^* \ll m$
- Gives locations and % of active voxels within R^*

Drawback

R^* must be defined *a priori*

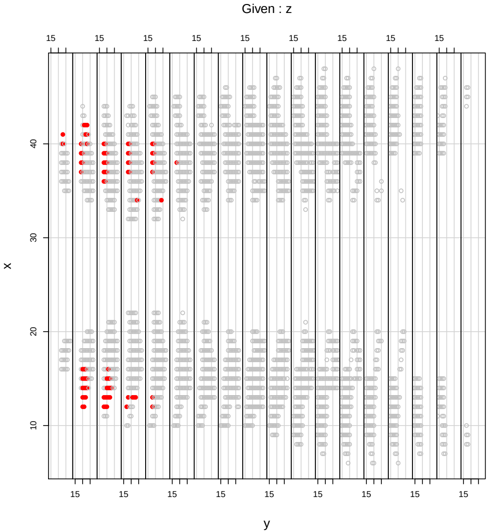
Fusiform gyrus





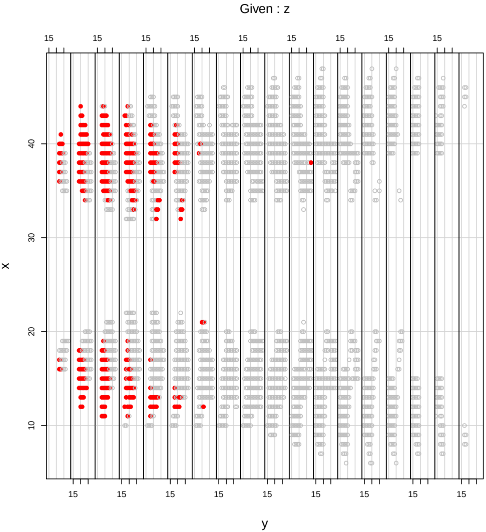
$R^* =$ fusiform gyrus (Brodmann area 37)

Activations (whole-brain correction)



Statement: 84 active voxels out of 2636 (3.2% of the region)

Activations (small volume correction)



Statement: 325 active voxels out of 2636 (12.3% of the region)

HARKing

*hark*ing = *h*ypothesizing *a*fter the *r*esults are *k*nown

= select which hypotheses to test after you've carried out the experiment and collected the data

A.K.A. *cherry-picking*, *double dipping*, *circular analysis*, *etc.*

1. Select any R^* by looking at the activation map
2. Pretend that R^* was chosen *a priori*
3. Make statements about voxels $\in R^*$

Consequence

Overoptimistic statements from *post hoc* selection of R^*

Exploratory data analysis

(10) Should data exploration be discouraged in favour of valid confirmatory analysis?

No!

Data exploration is essential to scientific discovery.

Kriegeskorte, Lindquist, Nichols, Poldrack, Vul (2010)
Journal of Central Blood Flow & Metabolism

Goal

Confirmatory analysis

A priori defined regions; valid statements

Exploratory analysis

Post hoc selected regions; overoptimistic statements

Can we have our cake and eat it too?

Post hoc selected regions and valid statements

Two ingredients

1. Simes' inequality
2. Closed testing

Our cake

Simultaneous inference

Selective and simultaneous inference

Selective inference
Adjusts for a pre-specified selection procedure

Simultaneous inference
Adjusts for any possible selection procedure

Valid post selection inference
Requires simultaneous inference

FDP estimation and confidence interval

Classic approach

- $\theta =$ parameter of interest
- Point estimate $\hat{\theta}$ for θ
- Confidence interval $[\underline{\theta}, \overline{\theta}]$ for θ

In our context

- For a fixed region $R \rightarrow$ statement: all voxels $\in R$ are active
- $\theta = \mathrm{FDP}(R) =$ proportion of non-active voxels
- $\hat{\theta} = \widehat{\mathrm{FDP}}(R)$
- $[\underline{\theta}, \overline{\theta}] = [0, \overline{\mathrm{FDP}}(R)]$

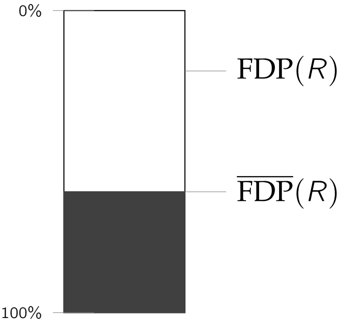
Post hoc selected R

Requires *simultaneous* point estimates and confidence intervals

Simultaneous upper bounds

$$\Pr\left\{\mathrm{FDP}(R) \leq \overline{\mathrm{FDP}}(R) \text{ for every } R\right\} \geq 95\%$$

$\overline{\mathrm{FDP}}(R)$ limits the true value $\mathrm{FDP}(R)$ at least 95% of the time, simultaneously for every R

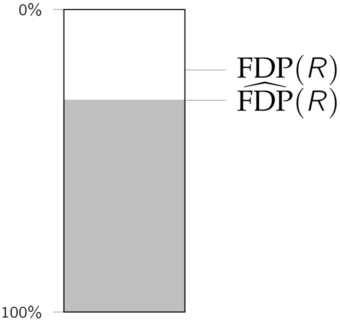


with probability $\geq 95\%$
for every R simultaneously

Simultaneous point estimates

$$\Pr\left\{\mathrm{FDP}(R) \geq \widehat{\mathrm{FDP}}(R) \text{ for every } R\right\} \leq 50\%$$

$\widehat{\mathrm{FDP}}(R)$ overestimates the true value $\mathrm{FDP}(R)$ at most 50% of the time, simultaneously for every R



with probability $\geq 50\%$
for every R simultaneously

Multilevel simultaneous

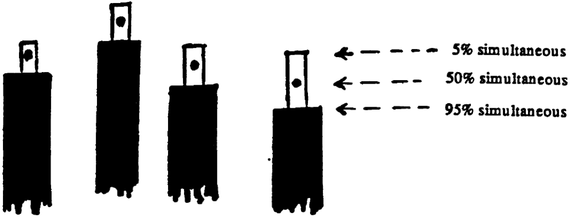
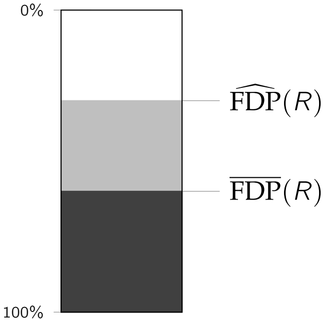


Fig. 2 in Tukey (1991) *Statistical Science*



Simes' inequality

With probability at least $1 - \alpha$

$$q_{(i)} > \frac{i\alpha}{m_0} \quad \text{for every } i$$

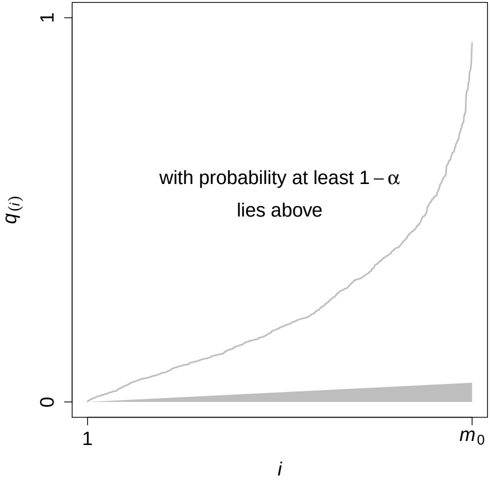
where

- $q_1, \ldots, q_{m_0}$ are the p-values of true null hypotheses
- $q_{(i)}$ is the ith ordered p-value

Assumption: Simes' inequality holds

- Same assumption as Hommel procedure
- Necessary condition for Benjamini-Hochberg procedure

Simes' inequality (graphically)



Simes' inequality = simultaneous lower bound for $(q_{(1)}, \ldots, q_{(m_0)})$

Closed testing

Marcus, Peritz and Gabriel (1976)

Fundamental principle of FWER control

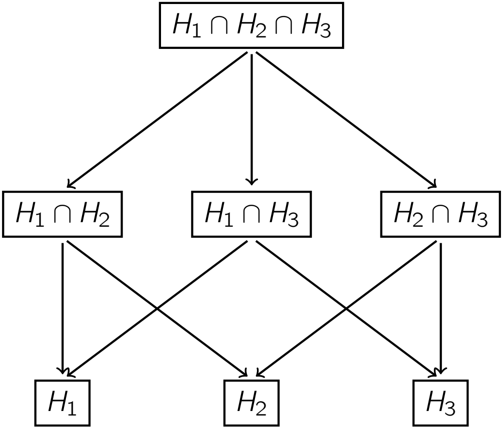
Requires

Testing the *closure*:

all $2^m - 1$ intersection hypotheses

$$H_R : \bigcap_{v \in R} H_v, \quad R \subseteq \{1, \ldots, m\}$$

Closed testing (graphically)



The closure

Fast algorithm

Closed testing

Requires $2^m - 1$ tests

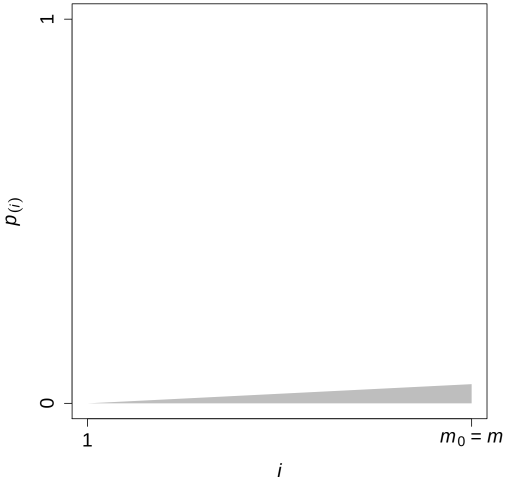
With $m = 39787$, requires $2^{39787} - 1$ tests!

Shortcut (Meijer et al., 2014)

- Use Simes' inequality as test (Simes' test)
- Reduces computational complexity to $m \log(m)$

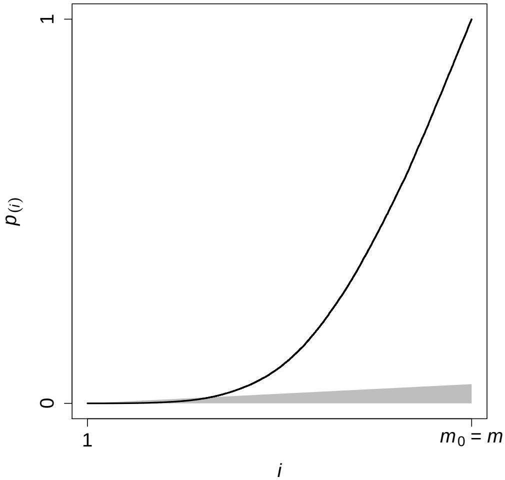
```
system.time(hF <- hommelFast(p.raw))

   user  system elapsed
 10.930   1.201  12.192
```

Closed testing with Simes' test



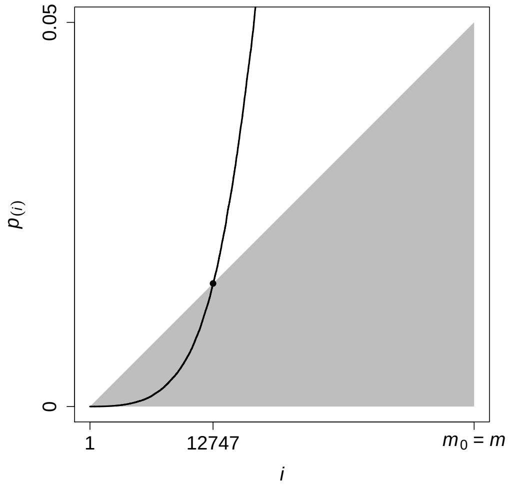
Assume that all hypotheses are true: $\widehat{m}_0 = m$

Closed testing with Simes' test



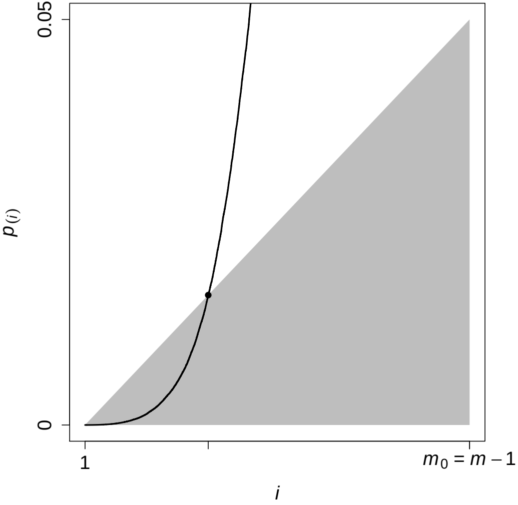
Compare p-value curve with lower bound

Closed testing with Simes' test



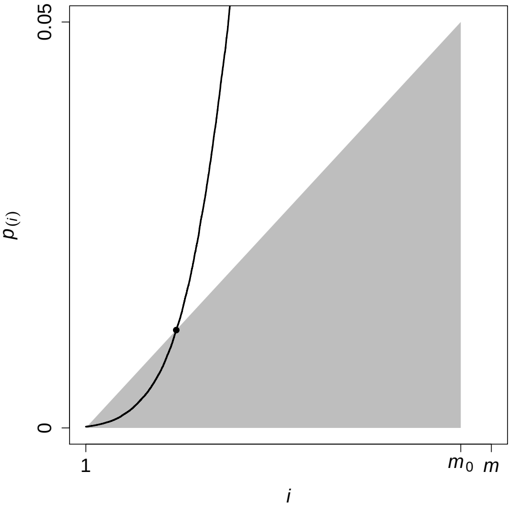
p-value curve crosses $\rightarrow$ at least 1 false (reject $\bigcap_{i=1}^{\widehat{m}_0} H_i$)

Closed testing with Simes' test



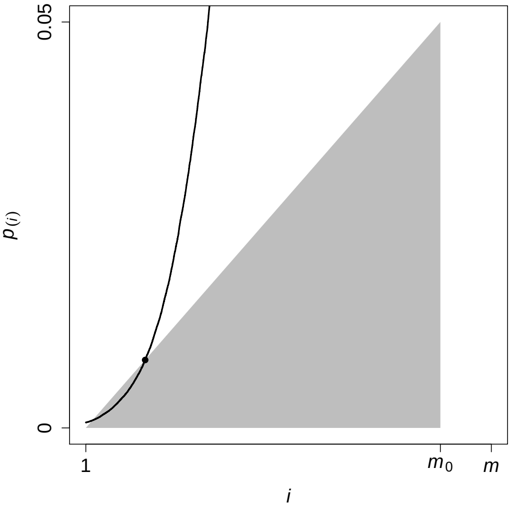
Discard the smallest p-value, set $\widehat{m}_0 = \widehat{m}_0 - 1$ and repeat

Closed testing with Simes' test



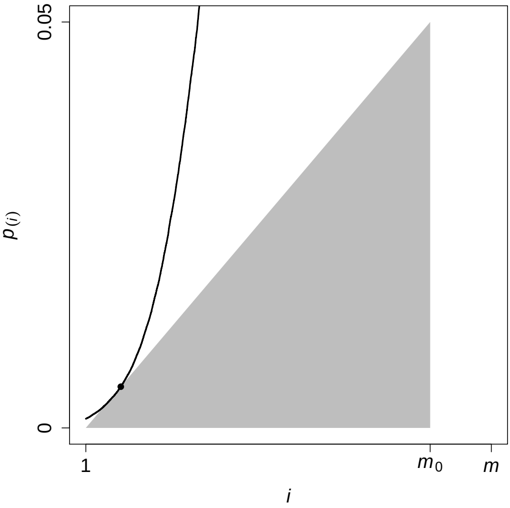
Still crossing

Closed testing with Simes' test



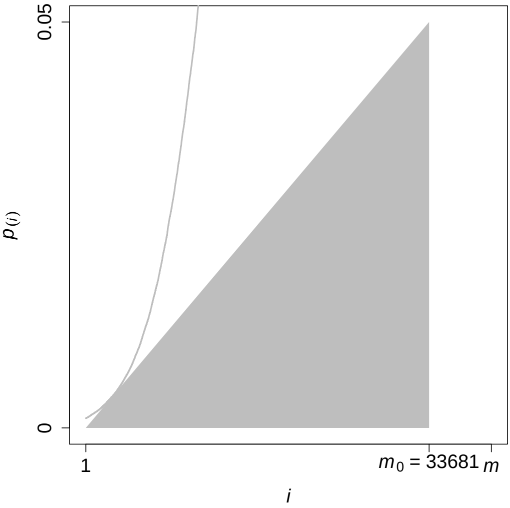
Still crossing

Closed testing with Simes' test



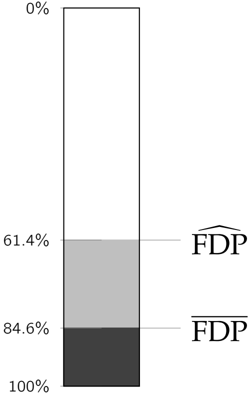
Still crossing

Closed testing with Simes' test



Stop: overall FDP $\leq \widehat{m}_0 / m = 33681/39787 = 84.6\%$

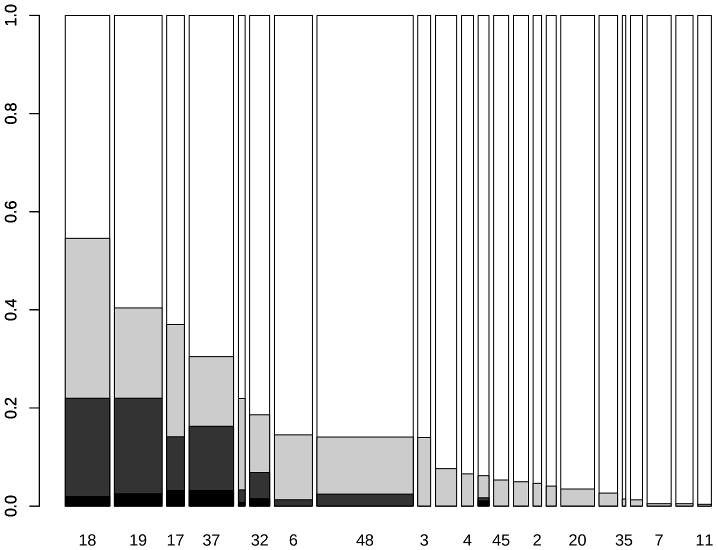
Overall activations



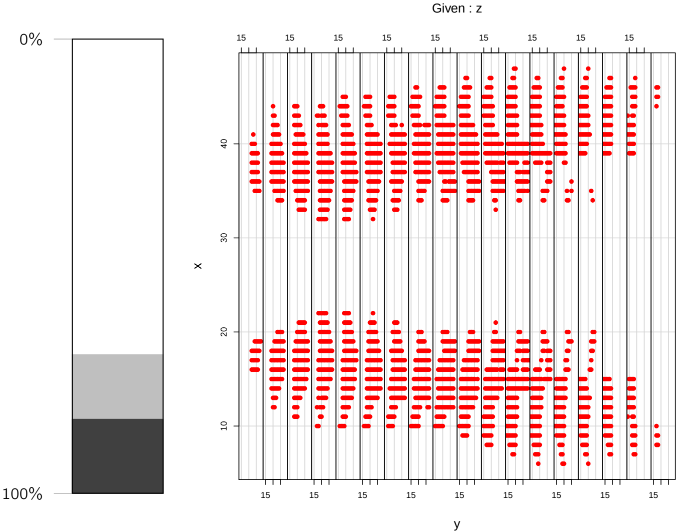
% of active voxels in total: 38.6% [**15.4%**, 100%]

Number of active voxels in total: 15367 [**6106**; 39787]

Brodmann areas



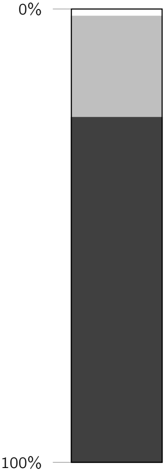
Fusiform gyrus

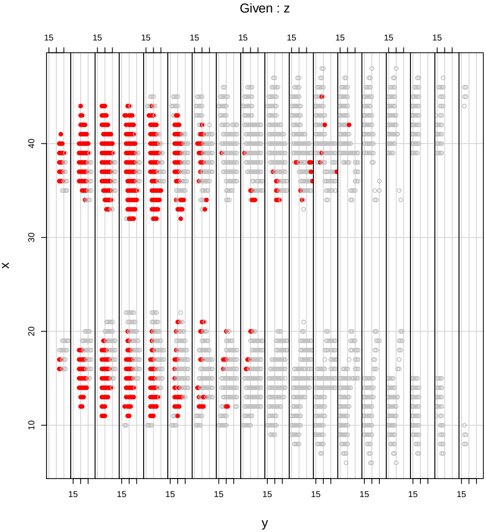


% of active voxels in Fusiform gyrus: 30.5% [**16.3%**, 100%]

Number of active voxels: 803 [**429**; 2636]

Fusiform gyrus

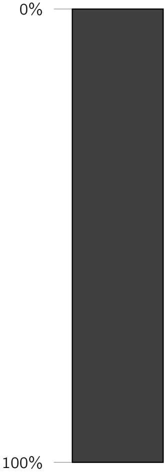


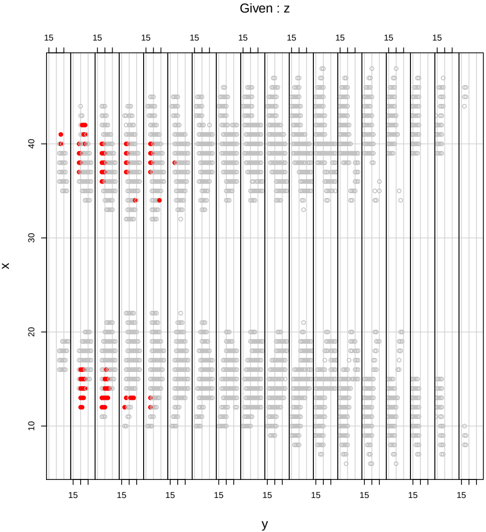


% of active voxels in red subregion: 98.4% [**76.1%**, 100%]

Number of active voxels: 555 [**429**; 564]

Whole-brain correction activations?

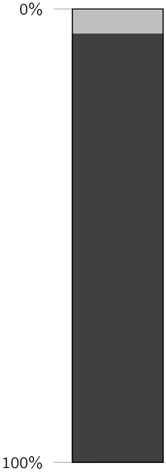


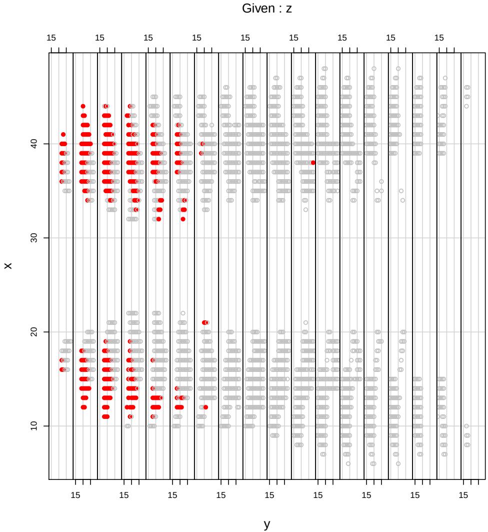


% of active voxels in red subregion: 100% [**100%**, 100%]

Number of active voxels: 84 [**84**; 84]

Small volume correction activations?





% of active voxels in red subregion: 99.7% [**94.5%**, 100%]

Number of active voxels: 324 [**307**; 325]

Link with other procedures

Link with Hommel

- Hommel activations have FDP $= 0$ with 95% confidence
- But obtain additional FDP statements for other regions

Link with Benjamini-Hochberg

- FDR activations have FDP estimate $\leq 10\%$
- FDR activations have FDP upper bound $< 100\%$

Summary

Closed testing

Test all intersection hypotheses

Simes' inequality

Makes this computationally feasible

Result

- Point estimate and confidence bound of FDP for any region
- Simultaneous $\rightarrow$ post hoc valid

Read more?

cherry
R package published on CRAN

Meijer RJ, Krebs T, Solari A and Goeman JJ (2015)
Extending Hommel's method
In preparation

Goeman JJ and Solari A (2011)
Multiple Testing for Exploratory Research.
Statistical Science 26:584–597 and 608–612

Goeman JJ and Solari A (2014)
Multiple hypothesis testing in genomics.
Statistics in Medicine 33:1946–78