



Permutation Testing & ETAC

Using neuropointillist to cluster-correct for multiple comparisons

John Flournoy

Research Scientist, Harvard University

9 Sept 2020

Roadmap

1. What is the multiple comparisons problem?
2. What is permutation testing?
3. What is ETAC?
4. How do I do it?

A photograph of a wooden staircase leading up a grassy hill. The stairs are made of light-colored wood and are partially covered with weeds and tall grass. The hill is covered in dense vegetation, including tall grasses and various weeds. The lighting is bright, suggesting a sunny day.

The Multiple Comparisons Problem

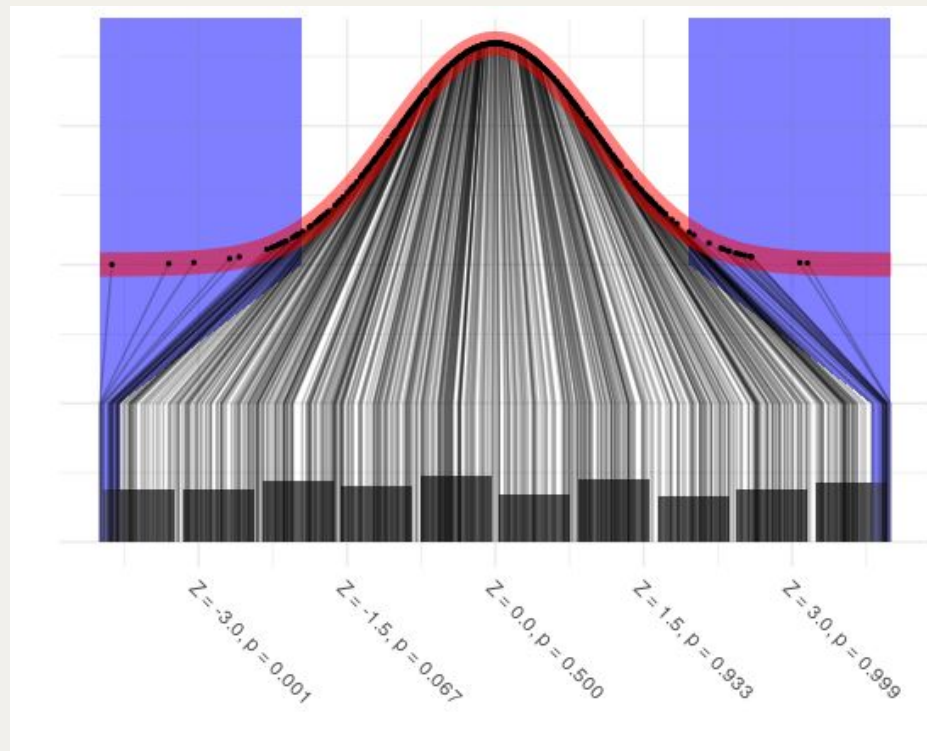
So many voxels: so many hypothesis tests

$\alpha = .05$ guarantees false positives
5% of the time...

...over the long run

For us “long run” = “so many tests”

Out of 200,000 voxels, 10,000
should be significant by chance




Corrections to maintain $\alpha = .05$

❖ Voxel-by-voxel

- Bonferroni, Holm's (Independent tests)
- Hochberg's, Hommel's (Independent, non-negatively correlated)
- False Discovery Rate

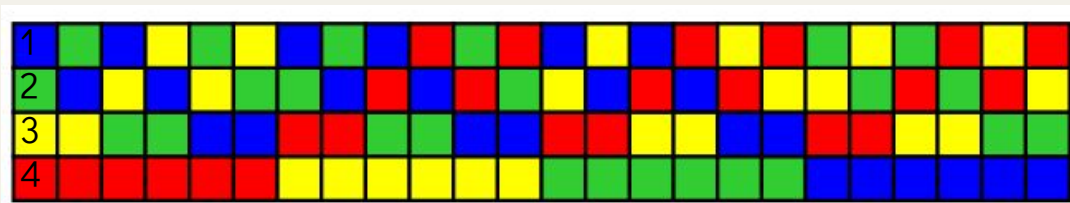
❖ Consider spatial autocorrelation of p -values

- Gaussian random field theory
- AFNI's improved autocorrelation function (Gaussian + mono-exponential)
- Non-parametric cluster correction
 - Voxel-wise p -value and cluster thresholds
- ETAC and TFCE

A photograph of a wooden staircase leading up a grassy hill. The stairs are made of light-colored wood and are partially covered with green grass and small plants. The hill is covered in dense, tall grass and various green plants. The lighting is bright, suggesting a sunny day. A white rectangular box with a red border is superimposed over the middle of the image, containing the text "What is permutation testing?".

What is permutation testing?

Permuting is a type of resampling



Resample without replacement (no duplicate values)

The number of *distinct* permutations gets big very fast:

$$n!$$

$$10! = 3,628,800$$

Permutation tests

The goal:

Construct the null distribution from the data

The method:

1. Permute the data
2. Compute the test statistic for each permutation
3. Use the resulting distribution to test the observed statistic

Permutation tests: the logic

1. Assume the null hypothesis is true *in your data*.
2. If the null is true,
 then rearranging the values of X with respect to Y doesn't matter.
3. Permuting X with respect to Y breaks any association.
4. Differences in the test statistic across permutations
 reflects random variation *under the null hypothesis*.
5. If your *observed* test statistic is extreme wrt the permutation distribution
 you can reject the null.

Very quick worked-example

```
set.seed(2391)
#Draw two samples from populations with different means
Y1 <- rnorm(100, mean = .4, sd = 1)
Y2 <- rnorm(100, mean = 0, sd = 1)
d <- data.frame(y = c(Y1, Y2),
                x = rep(c('g1', 'g2'), each = 100))

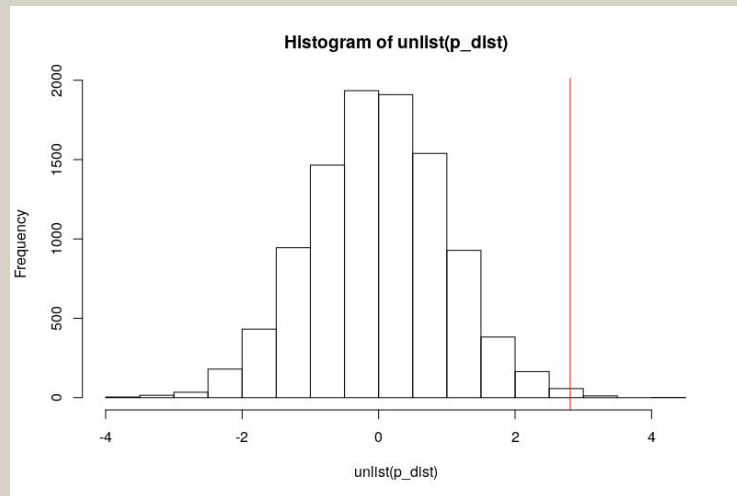
t_stat <- t.test(y ~ x, data = d)$statistic
# > t_stat
# t
# 2.808124

#Permute group labels using the row indices
p <- sample(1:200, replace = FALSE)
t_stat_p <- t.test(y ~ x[p], data = d)$statistic
# > t_stat_p
# t
# 0.2710259
```

Very quick worked example (cont)

```
#Permute 10,000 times
system.time({
  p_dist <- lapply(1:1e4, function(i){
    p <- sample(1:200, replace = FALSE)
    t_stat_p <- t.test(y ~ x[p], data = d)$statistic
    return(t_stat_p)
  })
})
# user  system elapsed
# 12.886    0.000   12.913

hist(unlist(p_dist))
abline(v = t_stat, col = 'red')
p_value <- mean(abs(unlist(p_dist)) > abs(t_stat))
# > p_value
# [1] 0.0054
```



Cluster correction by permutations

Apply the same logic to the *spatial distribution* under the null.

The method:

1. Permute the data
2. Compute the test statistic for each permutation *for each voxel*
3. Use the resulting ~~distribution to test the observed statistic~~
statistical parameter maps generated under the null
to find the cluster-size cutoff (for some voxel-wise threshold)
that maintains $\alpha = .05$ with respect to false positive clusters

Software that implements this stuff

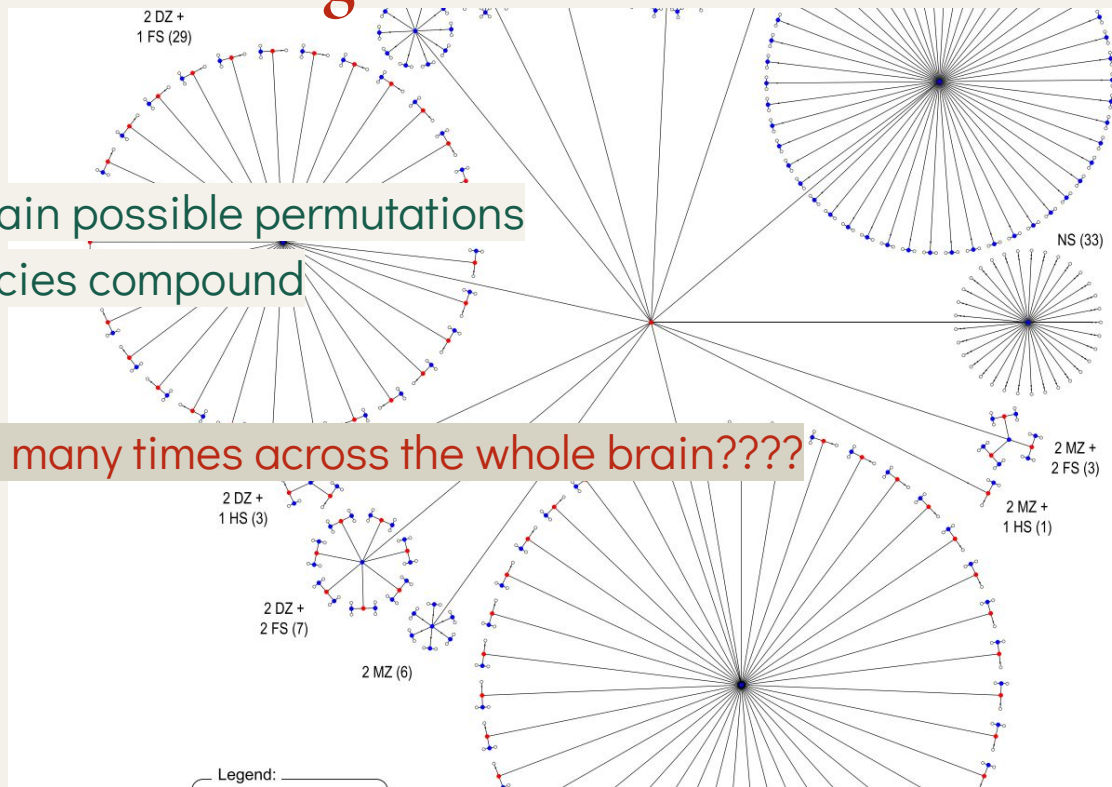
- ❖ AFNI 3dClustSim
 - Generates random maps using their ACF function + empirical parameters
 - Finds cluster sizes for voxel-wise cut-offs from these simulations
 - These simulations \neq permutations of the data
- ❖ AFNI 3dttest++
 - Cluster correction using permutations
 - Equitable Thresholding and Clustering (ETAC, which we'll talk about soon!)
- ❖ FSL Randomise
 - Cluster correction using permutations
 - Models specified same as FEAT
 - Wide range of designs (e.g., repeated measures), with covariates
 - Threshold-Free Cluster Enhancement (similar to ETAC)

Challenges in permutation testing

Complex data

- ❖ Repeated measures constrain possible permutations
- ❖ More difficult as dependencies compound
- ❖ COVARIATES 🤖

>>>You want me to do what how many times across the whole brain????



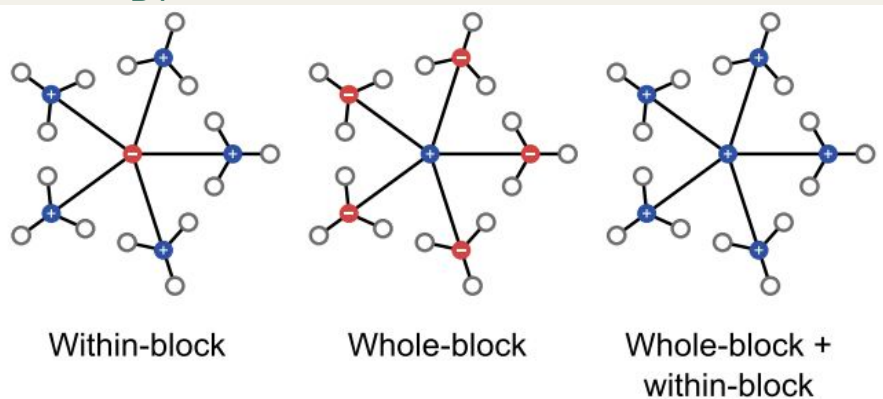
Complex designs

Example: $Y_{it} = \beta_{00} + \beta_{10}X_{it} + \nu_{0i} + \nu_{1i} + \epsilon_{it}$

- ❖ Y_{it} : At each time t , measure participant i task contrast (view fear – view blurred)
- ❖ X_{it} : Anxiety symptoms at time t for participant i .

Errors, ϵ_{it} are correlated within-person (“blocks”).

The permutation strategy should reflect this (Winkler et al. 2014, 2015).



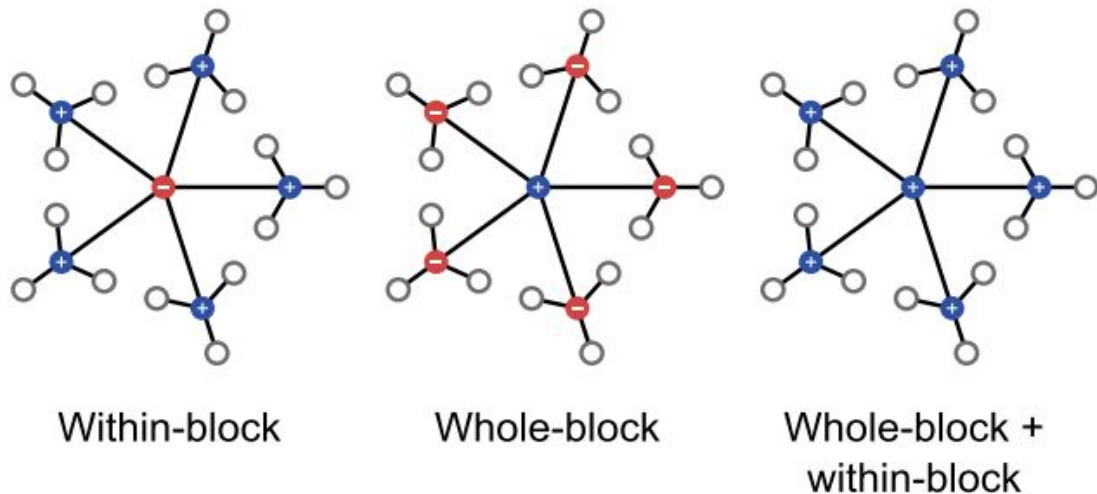
Complex designs

What's our null hypothesis?

Value of emotional faces contrast is not conditional on anxiety symptoms.

How should we permute in order to bring this into being?

Do we need to assume the null that between-person differences don't matter?



Handling Covariates

What to do with $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$?

Let's ask again, “what's our null hypothesis?”

*Value of emotional faces contrast is not conditional on anxiety symptoms
after we account for possible confounders.*

But how do we now break the relationship between \mathbf{X} and \mathbf{Y} ,
but not between \mathbf{X} and nuisance variables \mathbf{Z}
and not between \mathbf{Z} and \mathbf{Y} ?

The **Freedman–Lane procedure** (Freedman and Lane, 1983)! ^{*More on this later}

Freedman–Lane procedure (briefly)

1. Regress \mathbf{Y} against the full model that contains both the effects of interest and the nuisance variables, i.e. $\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\gamma + \epsilon$. Use the estimated parameters $\hat{\beta}$ to compute the statistic of interest, and call this statistic T_0 .
2. Regress \mathbf{Y} against a reduced model that contains only the nuisance effects, i.e. $\mathbf{Y} = \mathbf{Z}\gamma + \epsilon$, obtaining estimated parameters $\hat{\gamma}$ and estimated residuals $\hat{\epsilon}_Z$.
3. Compute a set of permuted data \mathbf{Y}_j^* . This is done by pre-multiplying the residuals from the reduced model produced in the previous step, $\hat{\epsilon}_Z$, by a permutation matrix, \mathbf{P}_j , then adding back the estimated nuisance effects, i.e. $\mathbf{Y}_j^* = \mathbf{P}_j\hat{\epsilon}_Z + \mathbf{Z}\hat{\gamma}$.
4. Regress the permuted data \mathbf{Y}_j^* against the full model, i.e. $\mathbf{Y}_j^* = \mathbf{X}\beta + \mathbf{Z}\gamma + \epsilon$, and use the estimated $\hat{\beta}_j^*$ to compute the statistic of interest. Call this statistic T_j^* .
5. Repeat Steps 2–4 many times to build the reference distribution of T^* under the null hypothesis.
6. Count how many times T_j^* was found to be equal to or larger than T_0 , and divide the count by the number of permutations; the result is the p-value.

A photograph of a wooden staircase leading up a grassy hill. The stairs are made of light-colored wood and are partially covered with green grass and small plants. The hill is covered in dense, tall grass and various green plants. The sky is bright and clear. The text "What is ETAC?" is overlaid in the center of the image in a red, serif font.

What is ETAC?

Equitable Thresholding and Clustering (ETAC)

Let's let Bob tell us*:

**there's also,
unsurprisingly,
a paper (Cox, 2019).*

Equitable Thresholding And Clustering (ETAC)

Robert W Cox



SSCC / NIMH / NIH / DHHS / USA / EARTH



Ahu Akivi

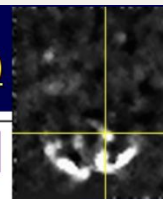
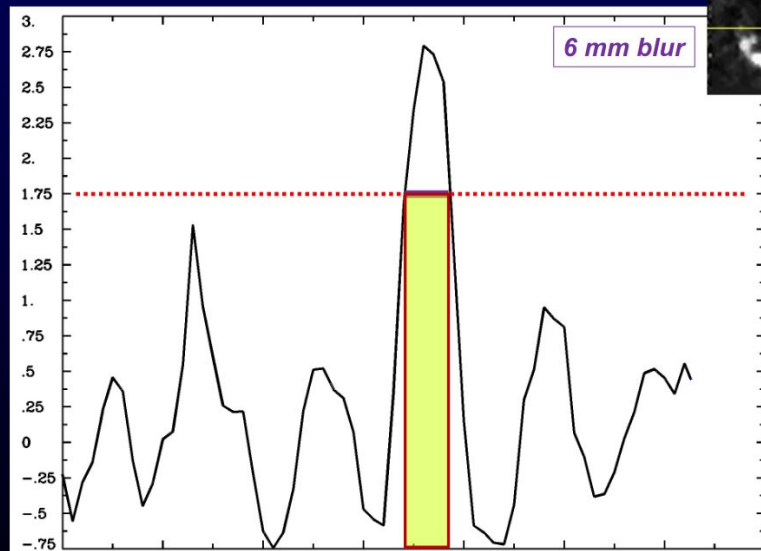
<https://afni.nimh.nih.gov>

2017

Choosing voxel-thresholds & cluster sizes

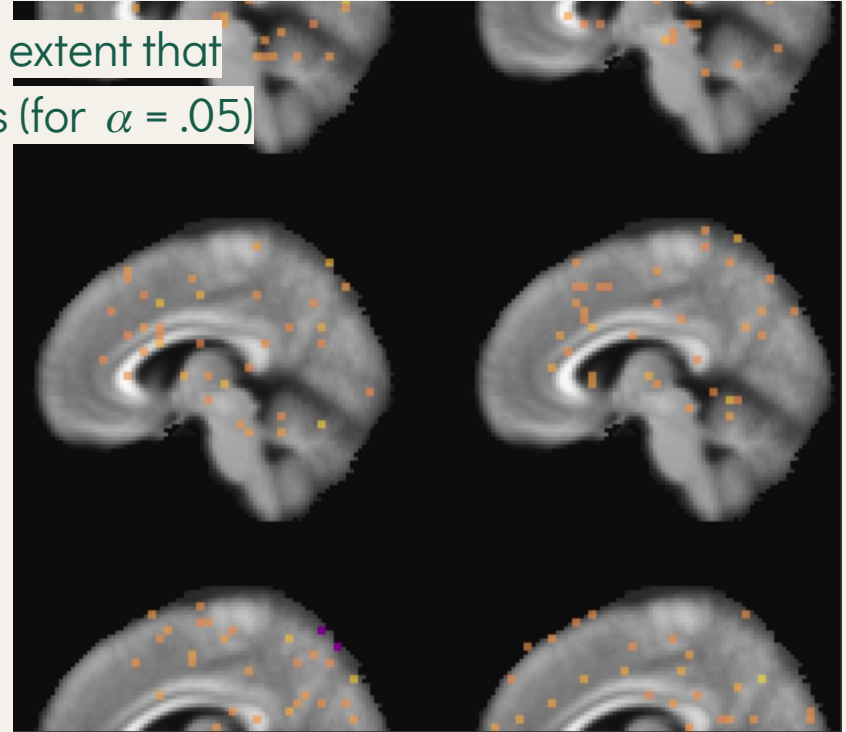
Select an arbitrary voxel-wise p -value threshold
and get its corresponding extent threshold (which will you choose!?)

1D Double Thresholding (real data)

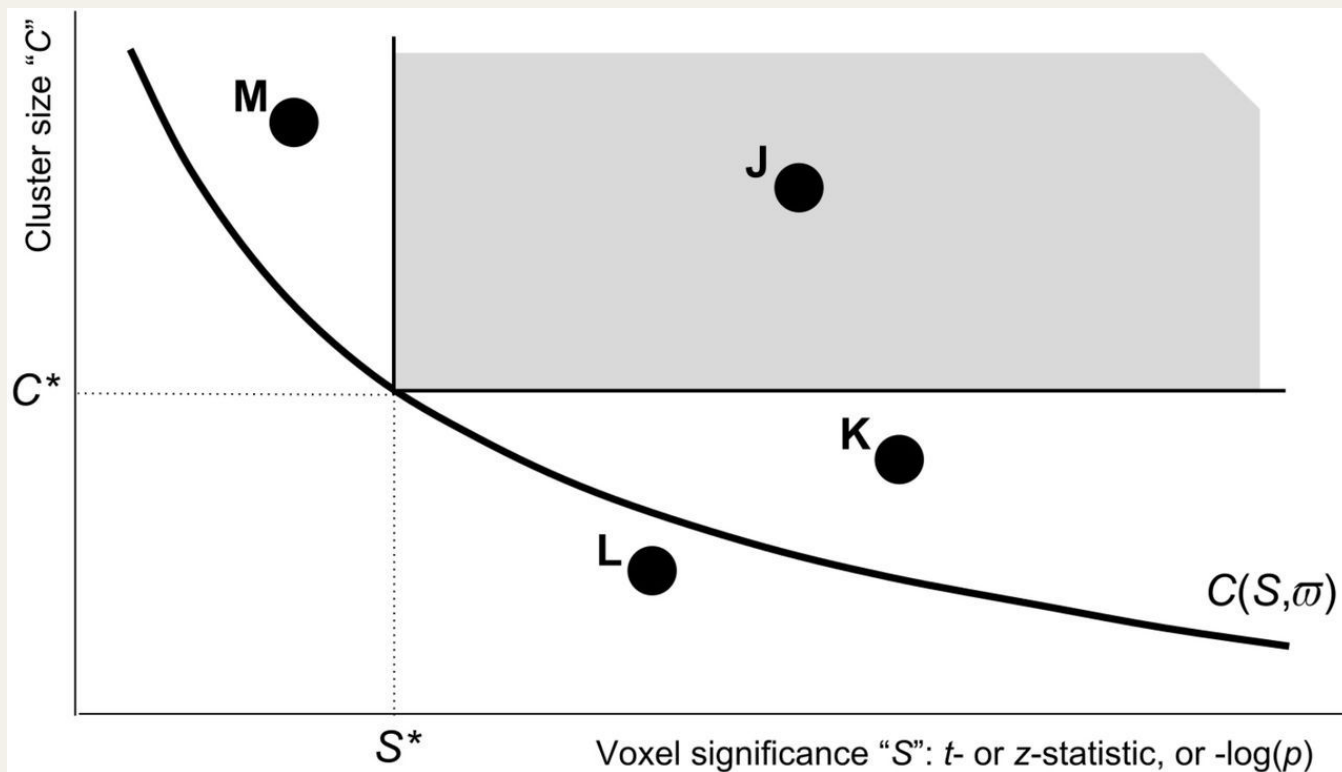


Reminder: using permuted SPMs

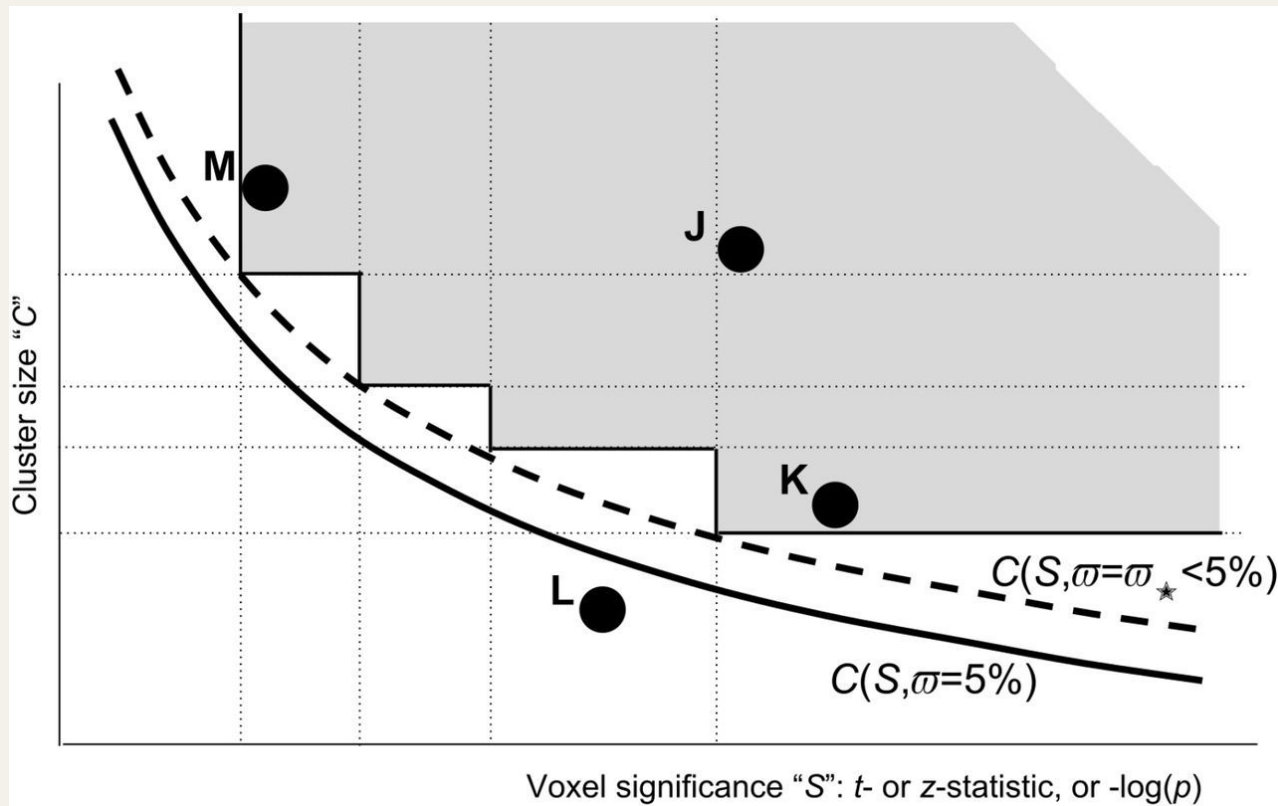
For a given voxel-wise threshold, find a cluster extent that yields a cluster in $< 5\%$ of permuted SPMs (for $\alpha = .05$)



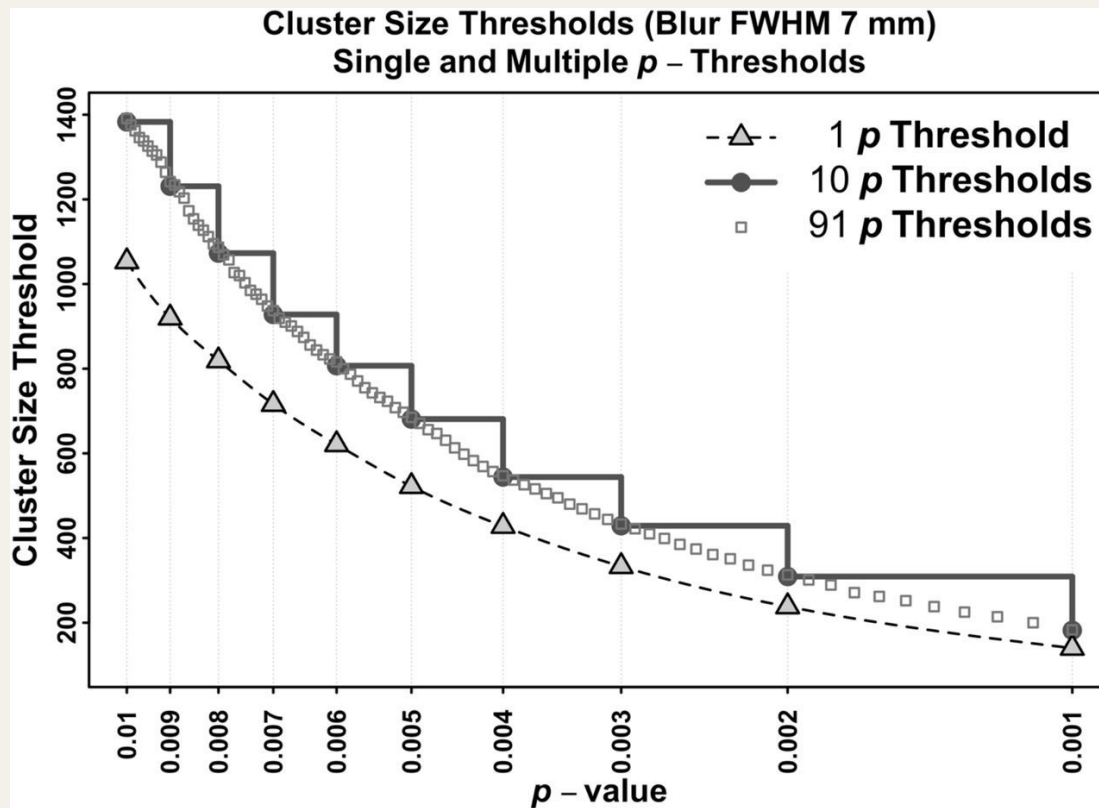
The “dual-thresholding trade-off” curve



Thresholding curve for multiple thresholds



Different numbers of thresholds



A photograph of a wooden staircase leading up a grassy hill. The stairs are made of light-colored wood and are partially covered with green grass and small plants. The hill is covered in dense, tall grass and various green plants. The sky is visible at the top of the frame, and the overall scene is bright and sunny. The text "How do I do it?" is overlaid in the center of the image in a red, serif font.

How do I do it?

Walkthrough fingerfootlips permutations...