

Permutation-based multiple testing with R - theory 2/2

Livio Finos

`livio.finos@unipd.it`

University of Padua

Ribno (Bled), Slovenia, 23/09/18

Outline

Introduction

Some examples

Some remarks

FamilyWise Error Rate (FWER)

Definition + Bonferroni

Holm (step-wise)

Closed Testing

FWER control by Permutations tests

Permutation Bonferroni

Westfall & Young: permutation Holm¹

Closed Testing

Conclusion

¹Westfall PH, Young SS (1993) Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment. Wiley

Microarray study

Compare 20,000 genes activity between disease and control group:

Top 5 genes out of 20,000

Gene	p-value
OCIAD2	5.5e-6
NEK3	6.7e-6
TAF5	7.1e-6
FOXD4L6	7.5e-6
ADIG	8.8e-6
⋮	⋮

Small p-value?

- a $p=5.5e-6$: very unlikely (evidende in favour of H_1)
- it is unlikely also on the light of 20,000 tested hypotheses?
- is OCIAD2 really differentially expressed?
- what about NEK3?

Surveys

e.g. self report inventory (web), psychological tests or social surveys

For each question we can get one or more tests
(e.g. comparing the opinions of two groups of people)

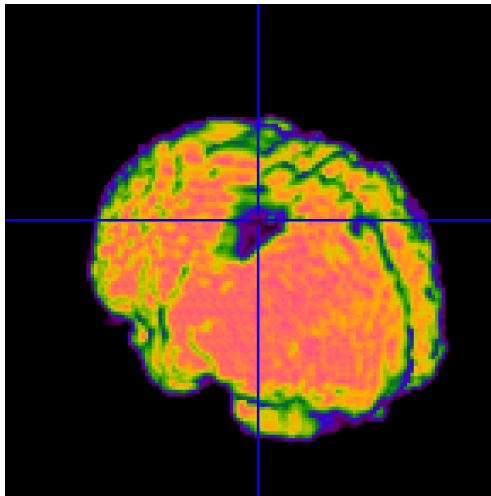
Surveys

e.g. self report inventory (web), psychological tests or social surveys

For each question we can get one or more tests
(e.g. comparing the opinions of two groups of people)

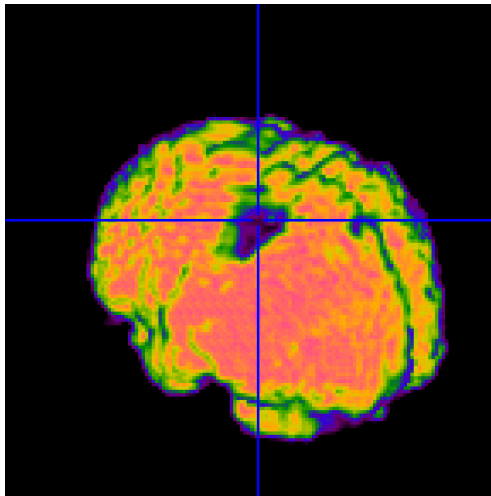
Question: Do we need to control for multiple type I errors?

A further example: fMRI studies



A map of brain activity for each subject

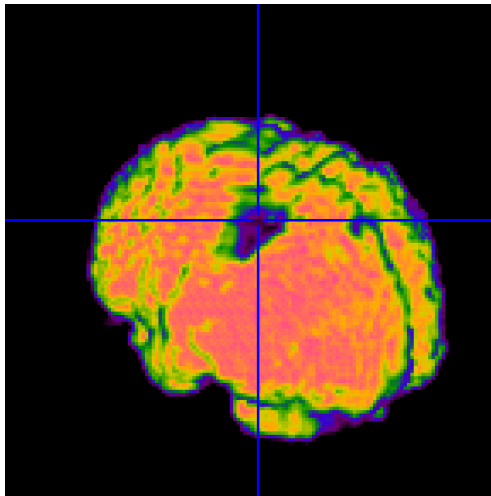
A further example: fMRI studies



A map of brain activity for each subject

A p-value for each voxel (=pixel, point)

A further example: fMRI studies

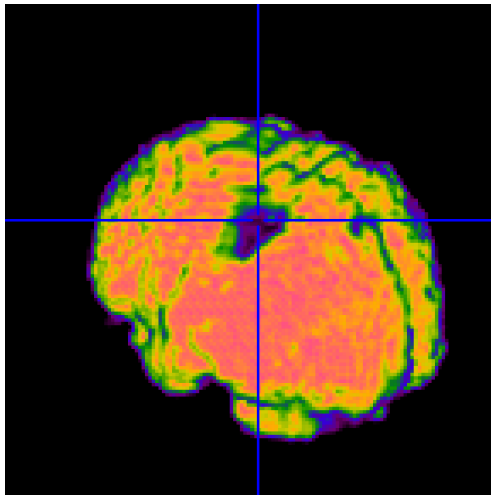


A map of brain activity for each subject

A p-value for each voxel (=pixel, point)

The output is a list of most significant voxels (among the thousands tested)

A further example: fMRI studies



A map of brain activity for each subject

A p-value for each voxel (=pixel, point)

The output is a list of most significant voxels (among the thousands tested)

Question: Do we need to control for multiple type I errors?

Further examples

Kinematics of human movements

a test for each aspect of the movement

Regression models (e.g. LM e GLM)

a t-test for each parameter of the model

Anova

All pairs comparisons (post-hoc)

Every times we get more than one p-value ...

Question: Do we need to control for multiple type I errors?

Further examples

Kinematics of human movements

a test for each aspect of the movement

Regression models (e.g. LM e GLM)

a t-test for each parameter of the model

Anova

All pairs comparisons (post-hoc)

Every times we get more than one p-value ...

Question: Do we need to control for multiple type I errors?

Answer: Yes, of course!

Hypotesis testing, a single test

compare the two hypos:

- H_0 : two groups are Equal, Y doesn't depend on X , nothing to publish :(
- H_1 : two groups are Different, Y does depend on X , publish it! :)

Each test produces a p-value p ,

if $p \leq .05$ ($\alpha = .05$) reject H_0 (and incline to H_1)

Errors

- **Type I** (false positive): Reject H_0 when it is TRUE
 $P(\text{Type I error}) = P(p \leq .05 | H_0) = .05$
- **Type II** (false negative): Don't reject H_0 when it is FALSE
 $P(\text{Type II error}) = P(p > .05 | H_1)$

Power:

$$\begin{aligned} P(p \leq .05 | H_1) &= 1 - P(p > .05 | H_1) \\ &= 1 - P(\text{Type II error}) \end{aligned}$$

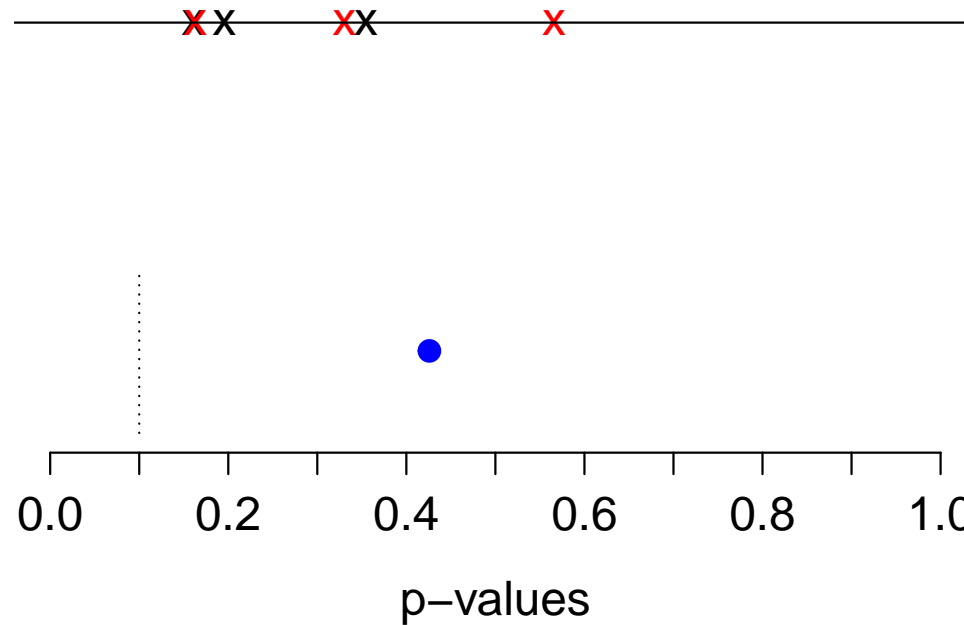
Power is nothing without control
(Asymmetric relevance of errors)

We control the $P(\text{type I Error})$ (ie $\leq .05$)
and (try) to maximize the power (i.e. minimize *type II error*)

Type I Error:

$$P(p \leq .05 | H_0 = \text{groups are equal}) = ?$$

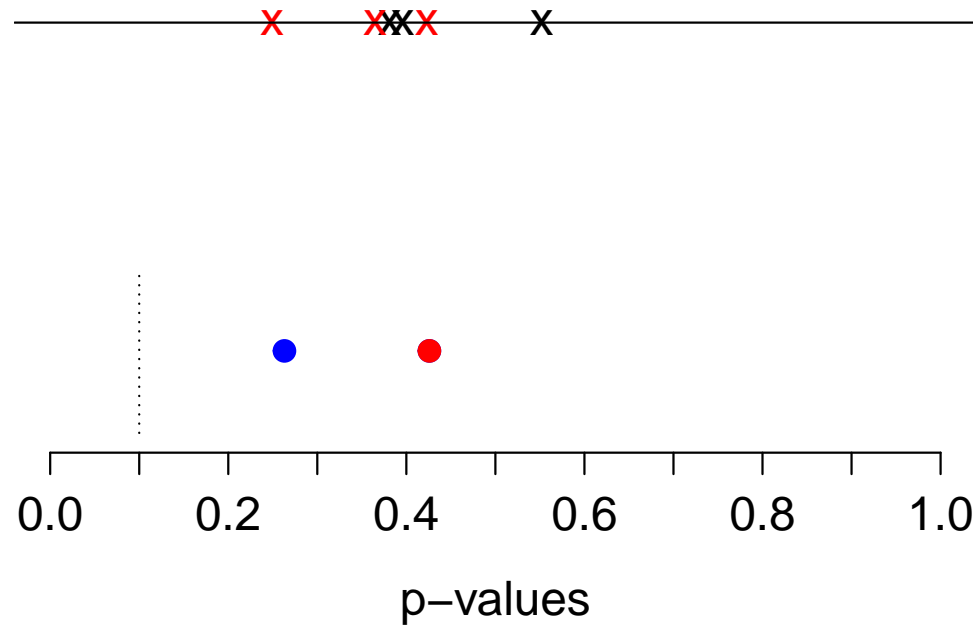
$t = -0.886$, $p = 0.426$



Type I Error:

$$P(p \leq .05 | H_0 = \text{groups are equal}) = ?$$

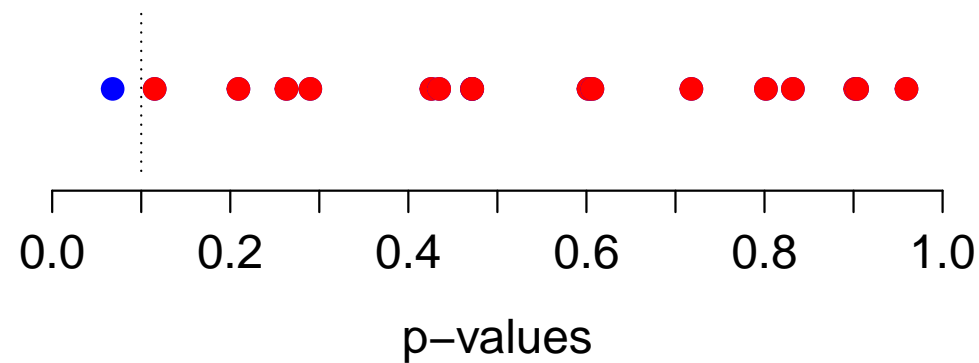
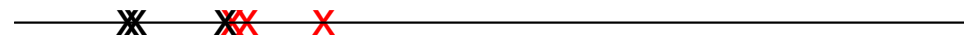
$t = 1.301$, $p = 0.263$



Type I Error:

$$P(p \leq .05 | H_0 = \text{groups are equal}) = ?$$

$t = -2.484$, $p = 0.068$

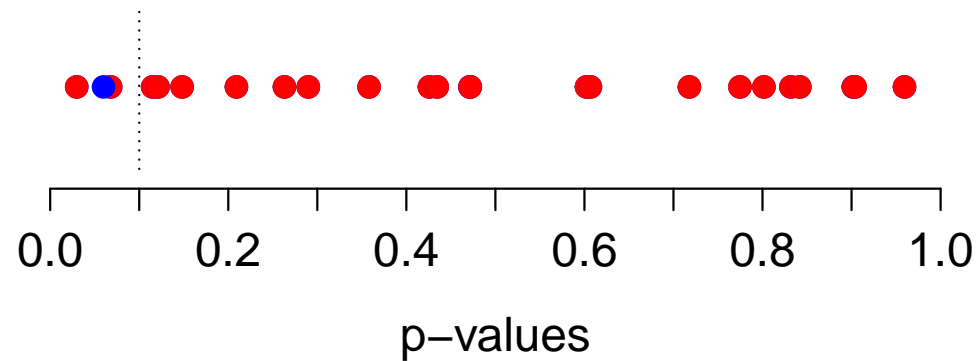


Type I Error:

$$P(p \leq .05 | H_0 = \text{groups are equal}) = ?$$

$t = -2.602$, $p = 0.06$

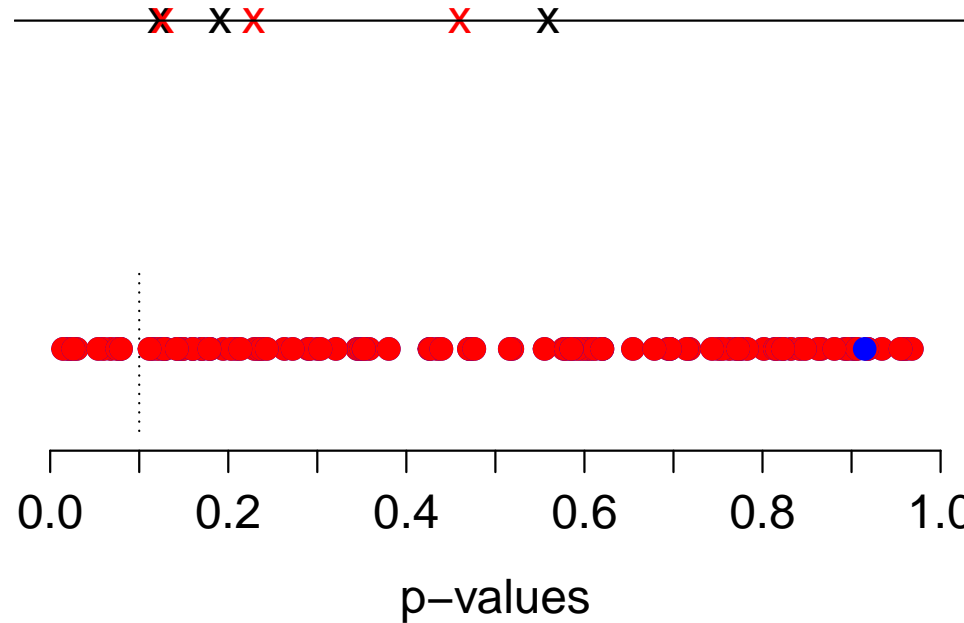
x x x x x x



Type I Error:

$$P(p \leq .05 | H_0 = \text{groups are equal}) = 0.05$$

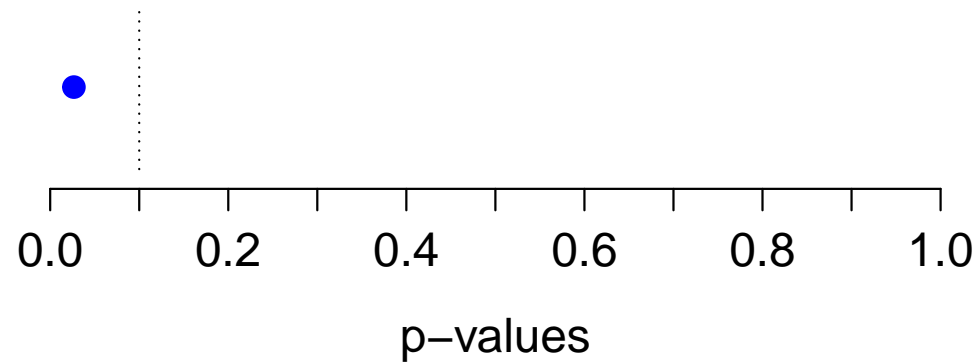
t= 0.114 , p= 0.915



Power:

$$P(p \leq .05 | H_1 = \text{groups are NOT equal})$$

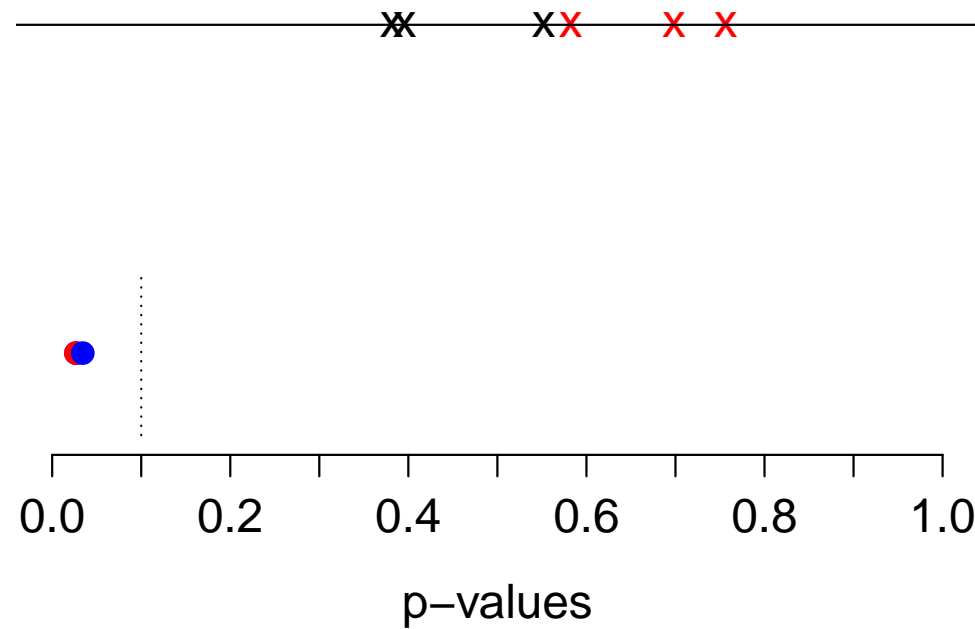
$t = -3.426$, $p = 0.027$



Power:

$$P(p \leq .05 | H_1 = \text{groups are NOT equal})$$

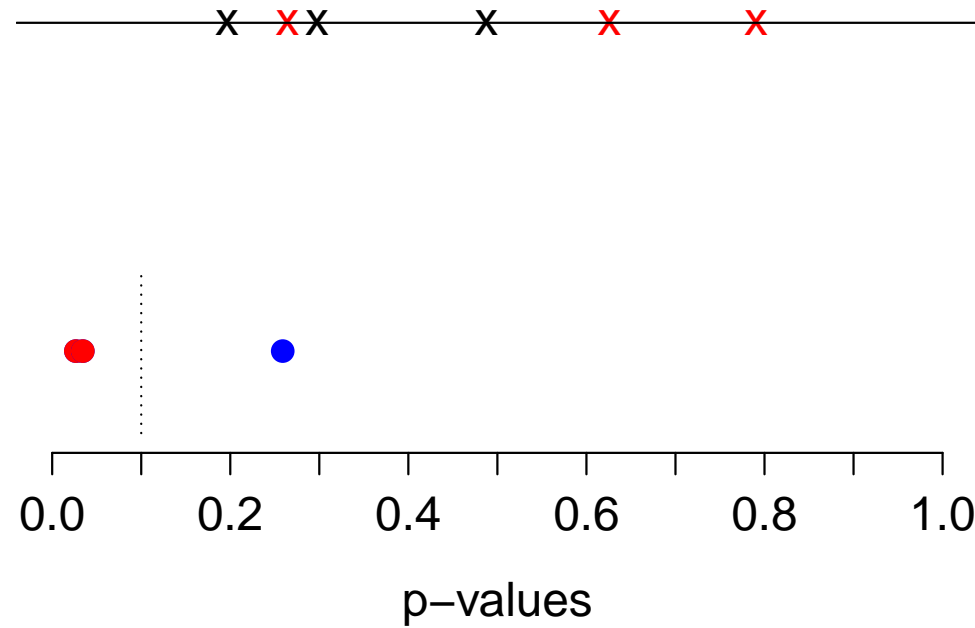
$t = -3.154$, $p = 0.034$



Power:

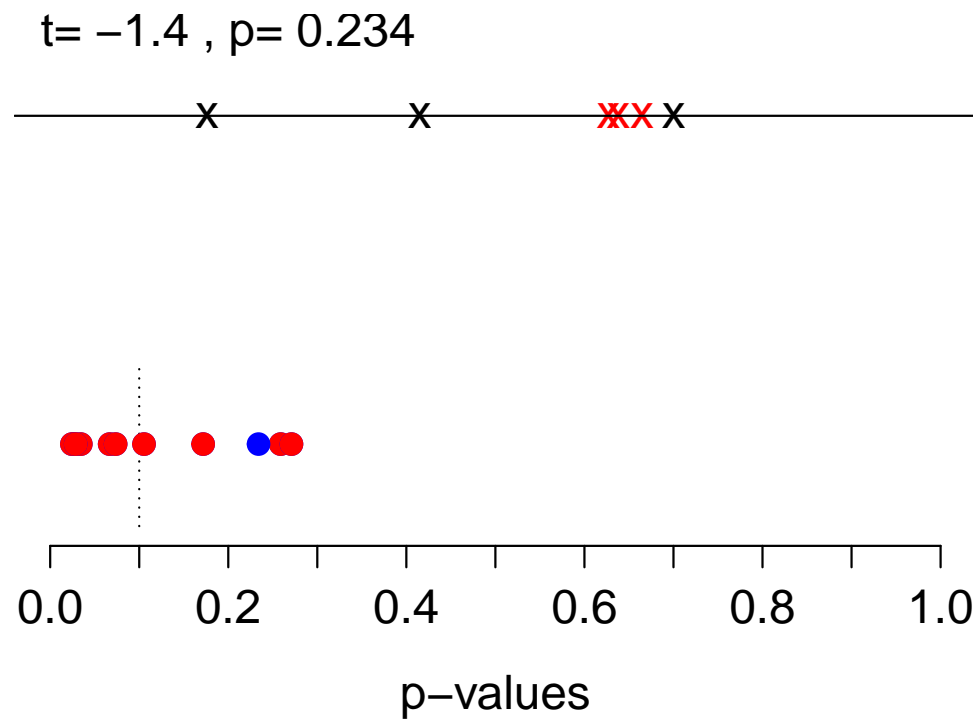
$$P(p \leq .05 | H_1 = \text{groups are NOT equal})$$

$t = -1.315$, $p = 0.259$



Power:

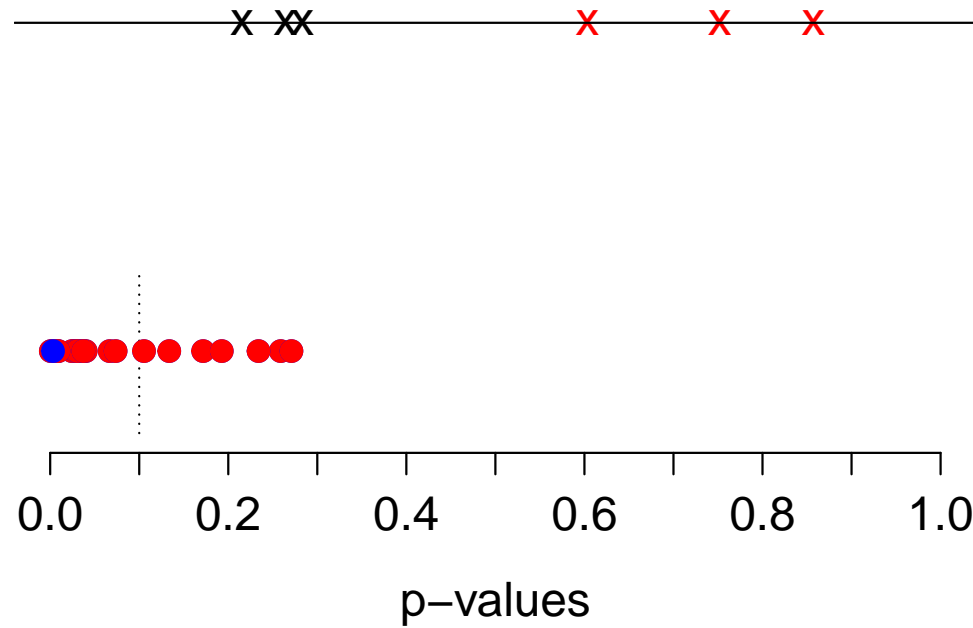
$P(p \leq .05 | H_1 = \text{groups are NOT equal})$



Power:

$$P(p \leq .05 | H_1 = \text{groups are NOT equal})$$

$t = -6.331$, $p = 0.003$

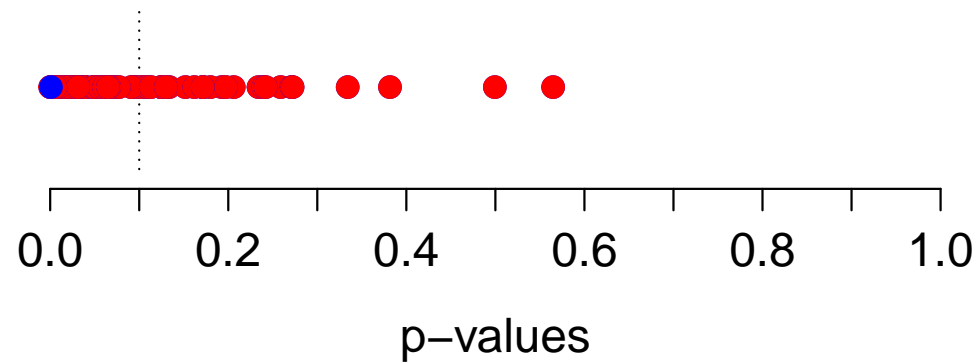
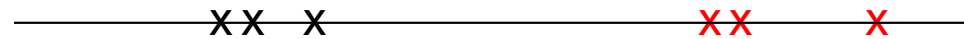


Power:

$P(p \leq .05 | H_1 = \text{groups are NOT equal})$

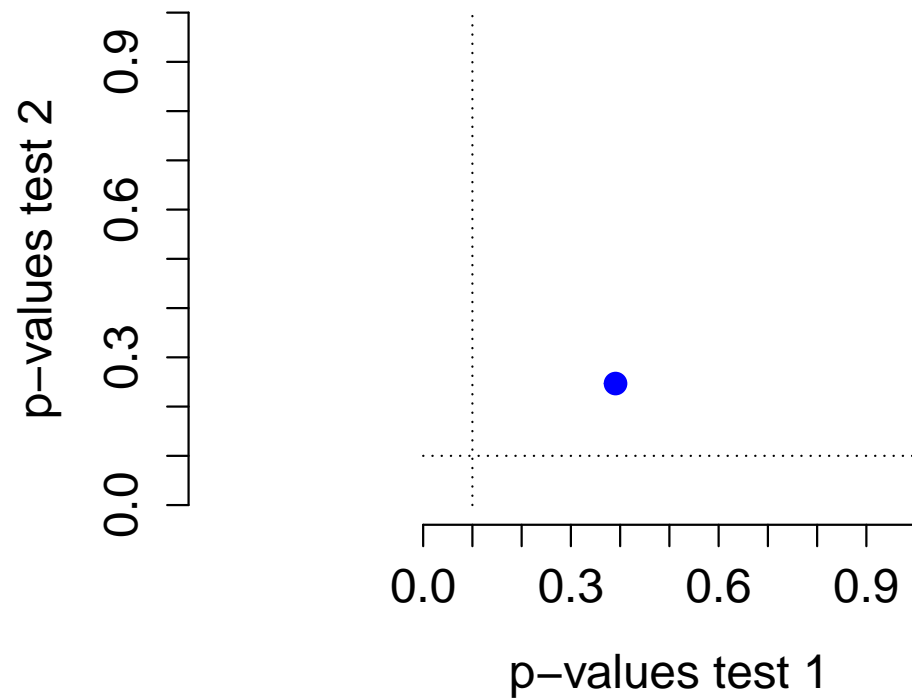
e.g.: *Power* : $P(p \leq 0.05 | H_1) = 0.75$

$t = -8.81$, $p = 0.001$



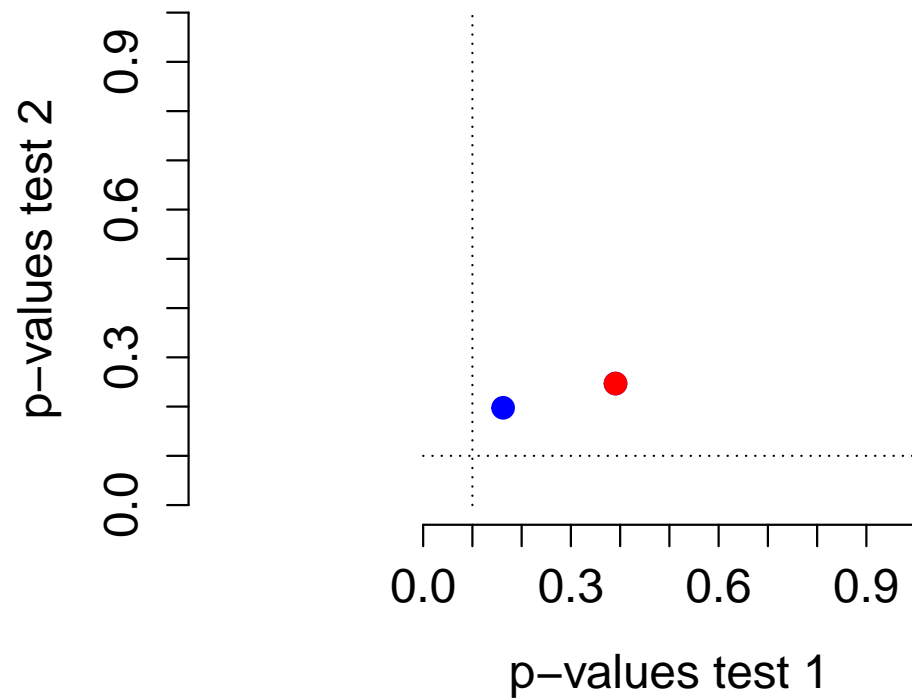
Type I Error, Two Tests

Probability of at LEAST one (false) rejection?



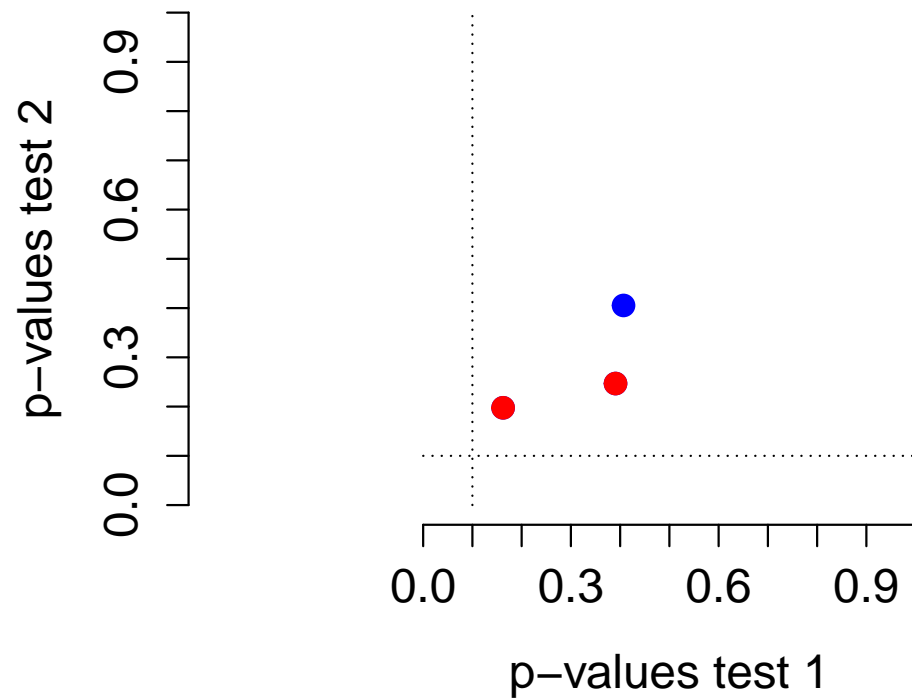
Type I Error, Two Tests

Probability of at LEAST one (false) rejection?



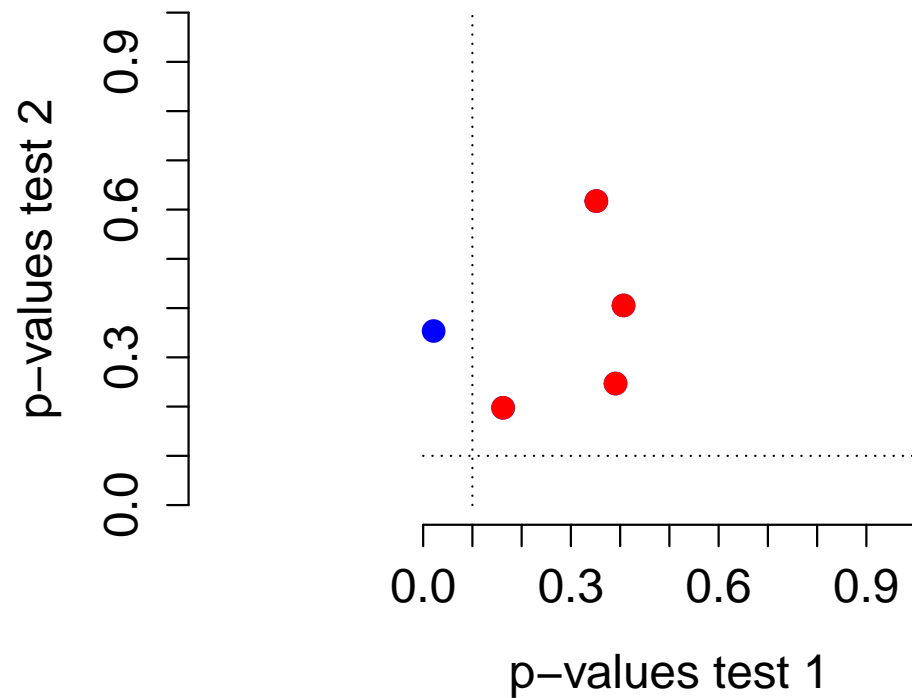
Type I Error, Two Tests

Probability of at LEAST one (false) rejection?



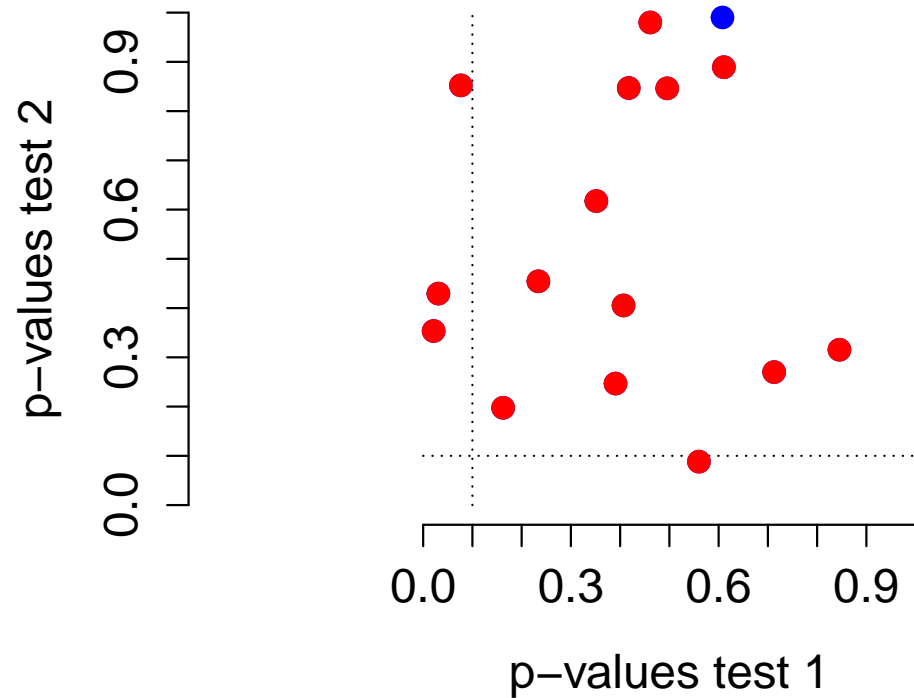
Type I Error, Two Tests

Probability of at LEAST one (false) rejection?



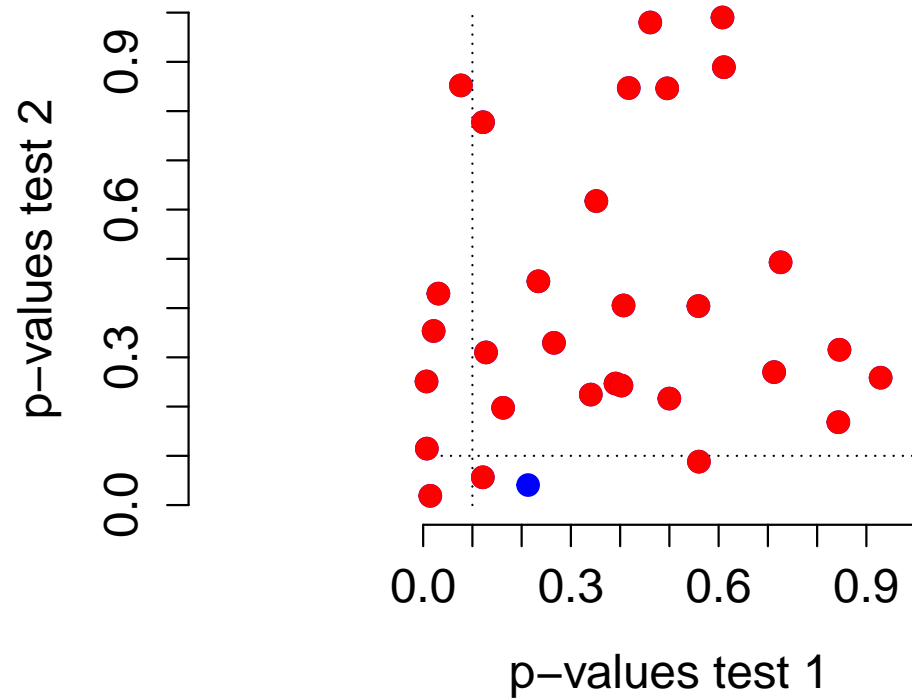
Type I Error, Two Tests

Probability of at LEAST one (false) rejection?



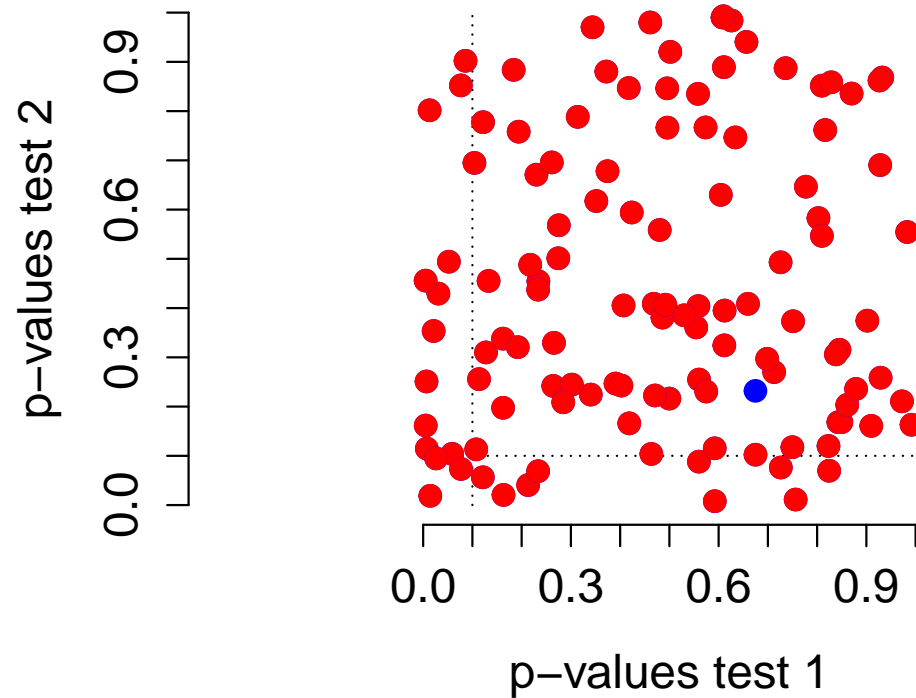
Type I Error, Two Tests

Probability of at LEAST one (false) rejection?



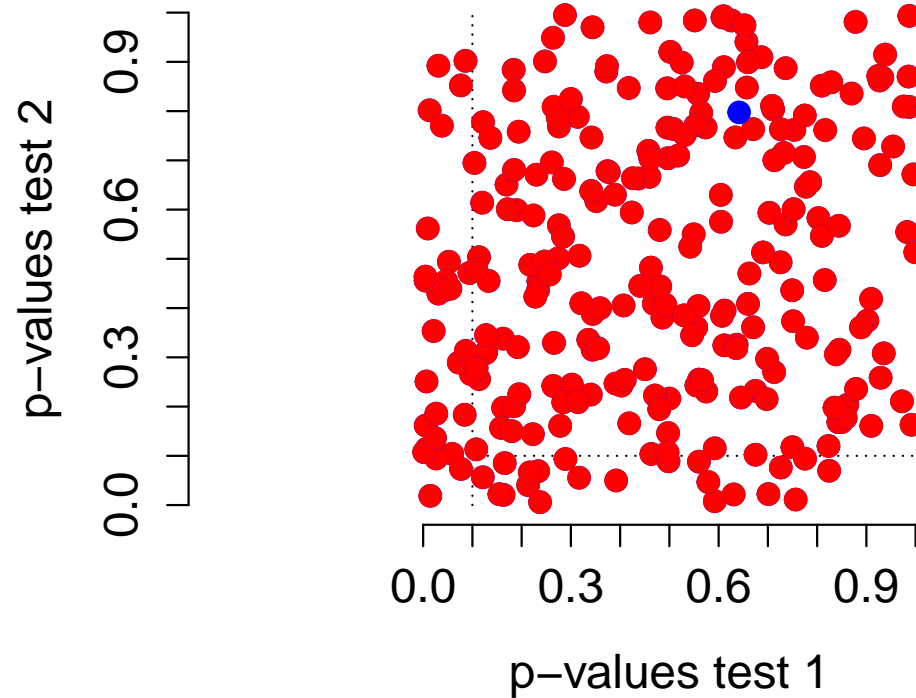
Type I Error, Two Tests

Probability of at LEAST one (false) rejection?



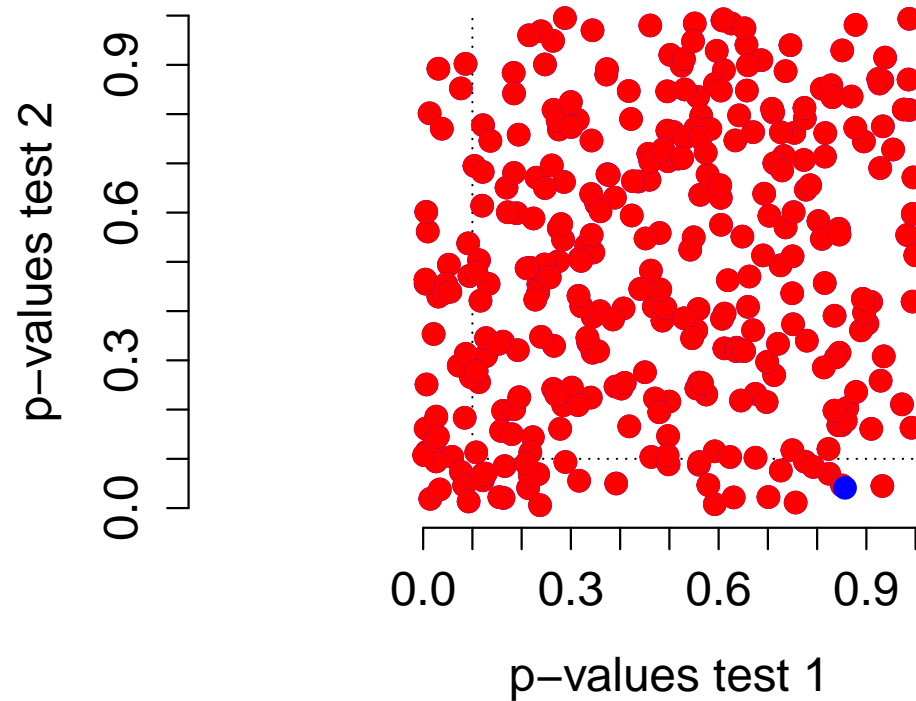
Type I Error, Two Tests

Probability of at LEAST one (false) rejection?



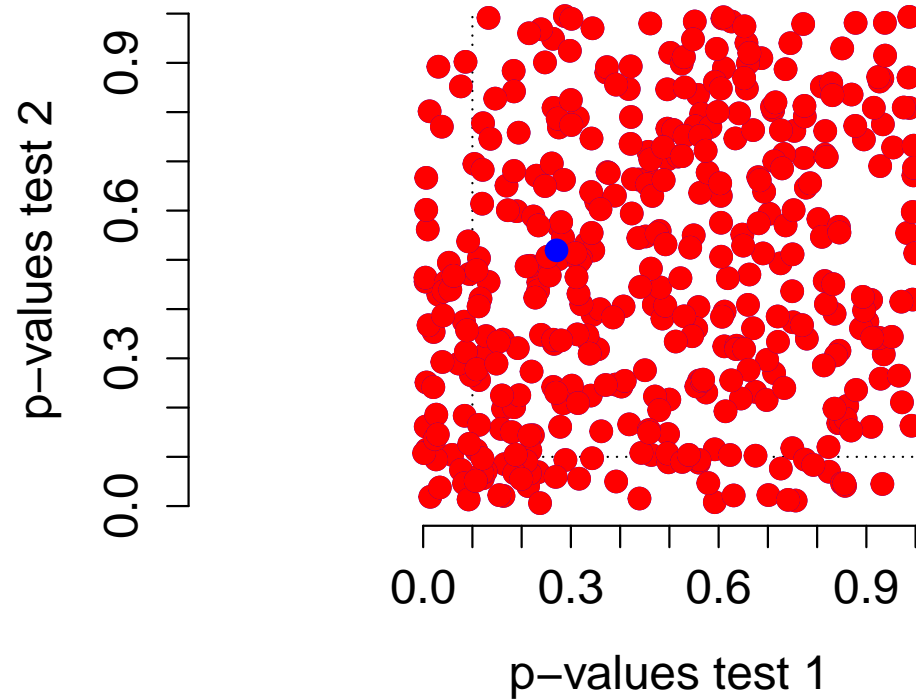
Type I Error, Two Tests

Probability of at LEAST one (false) rejection?



Type I Error, Two Tests

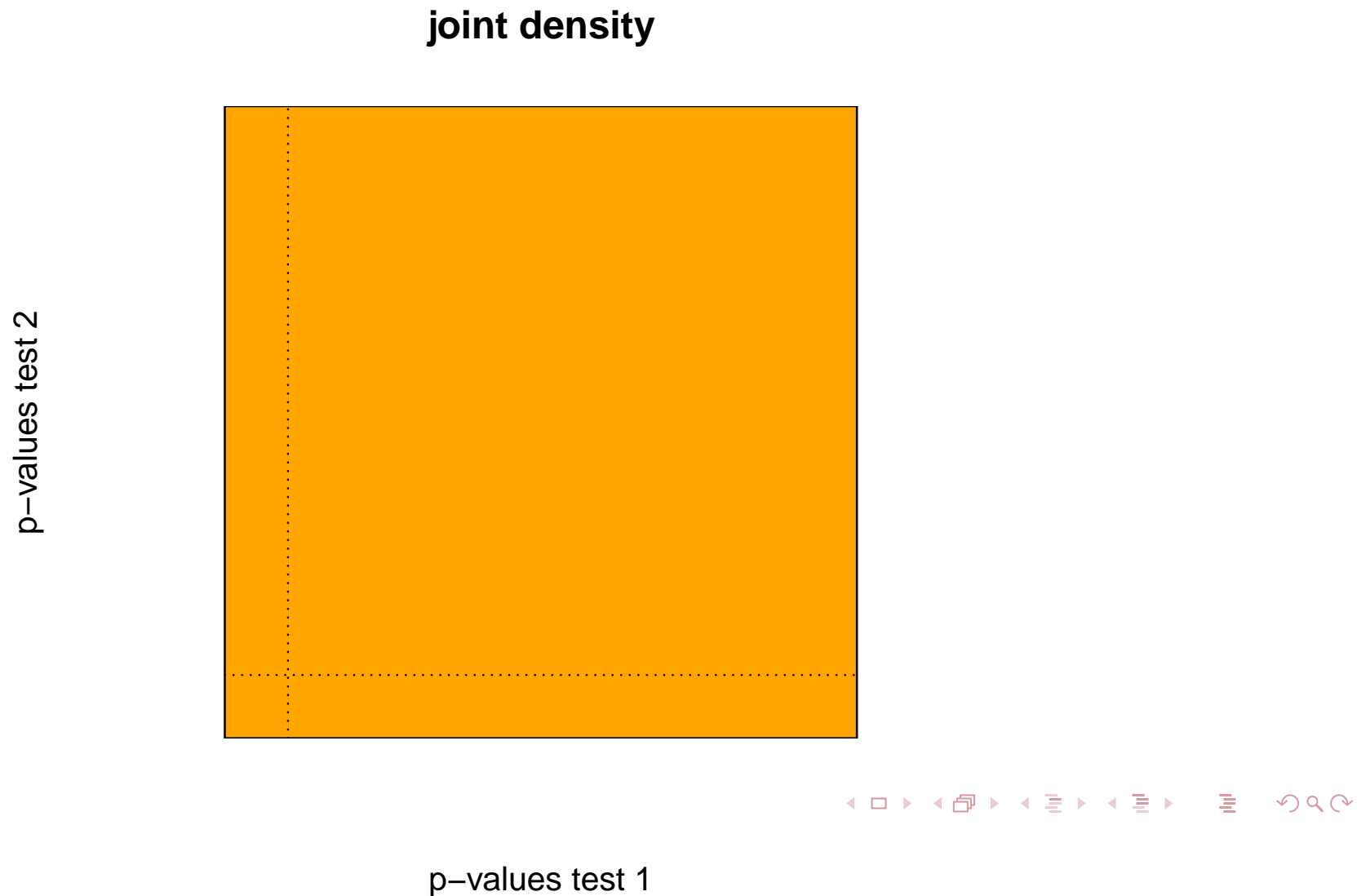
Probability of at LEAST one (false) rejection?



Type I Error, Two Tests

Probability of at LEAST one (false) rejection?

$$= .05 + .05 - (.05 * .05) = 1 - (1 - .05)^2 = .0975 = 1 - (1 - \alpha)^2$$



Probability of false rejections

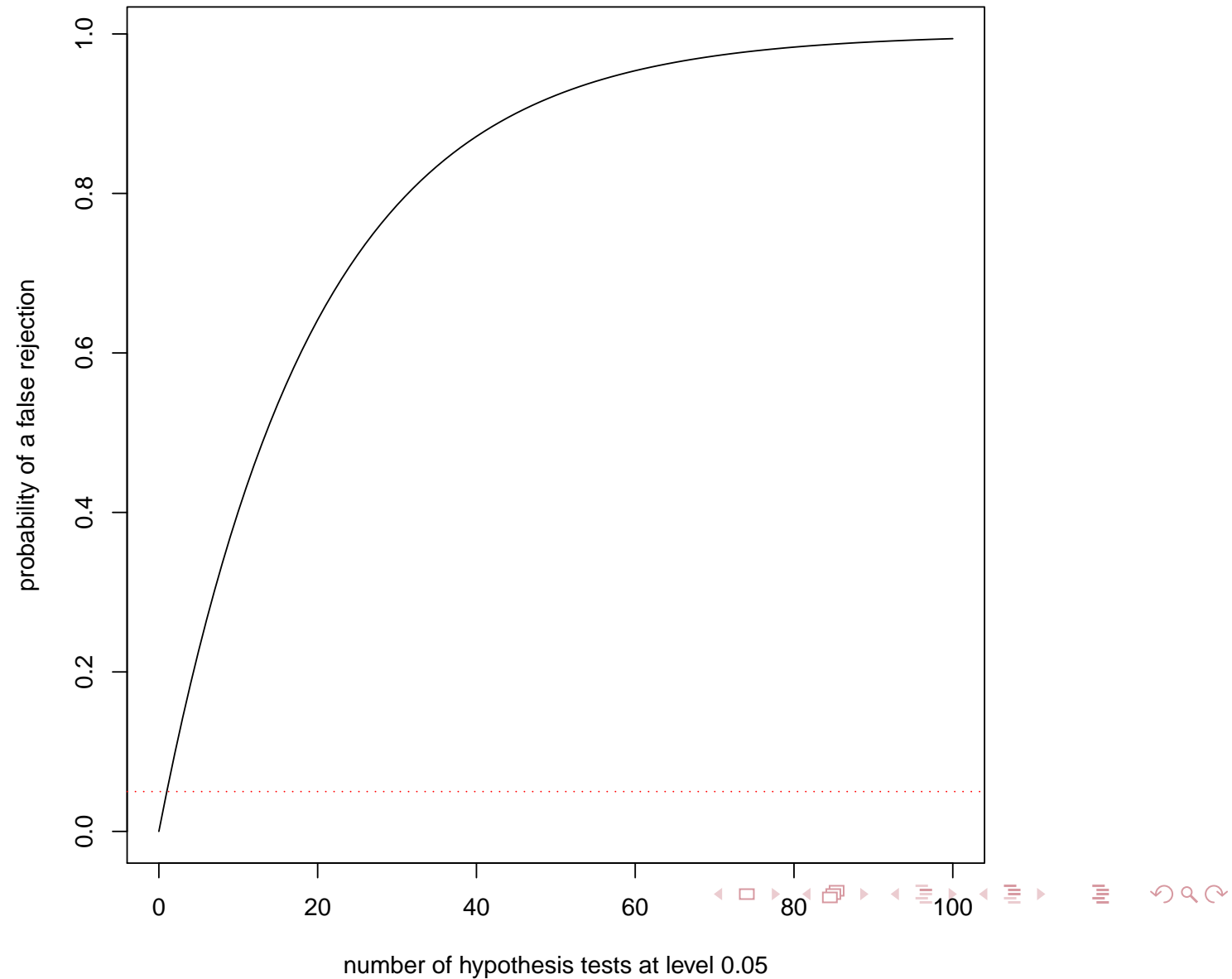
m independent tests (p-values)

Reject H_0 if $p \leq \alpha$

Probability of AT LEAST a rejection

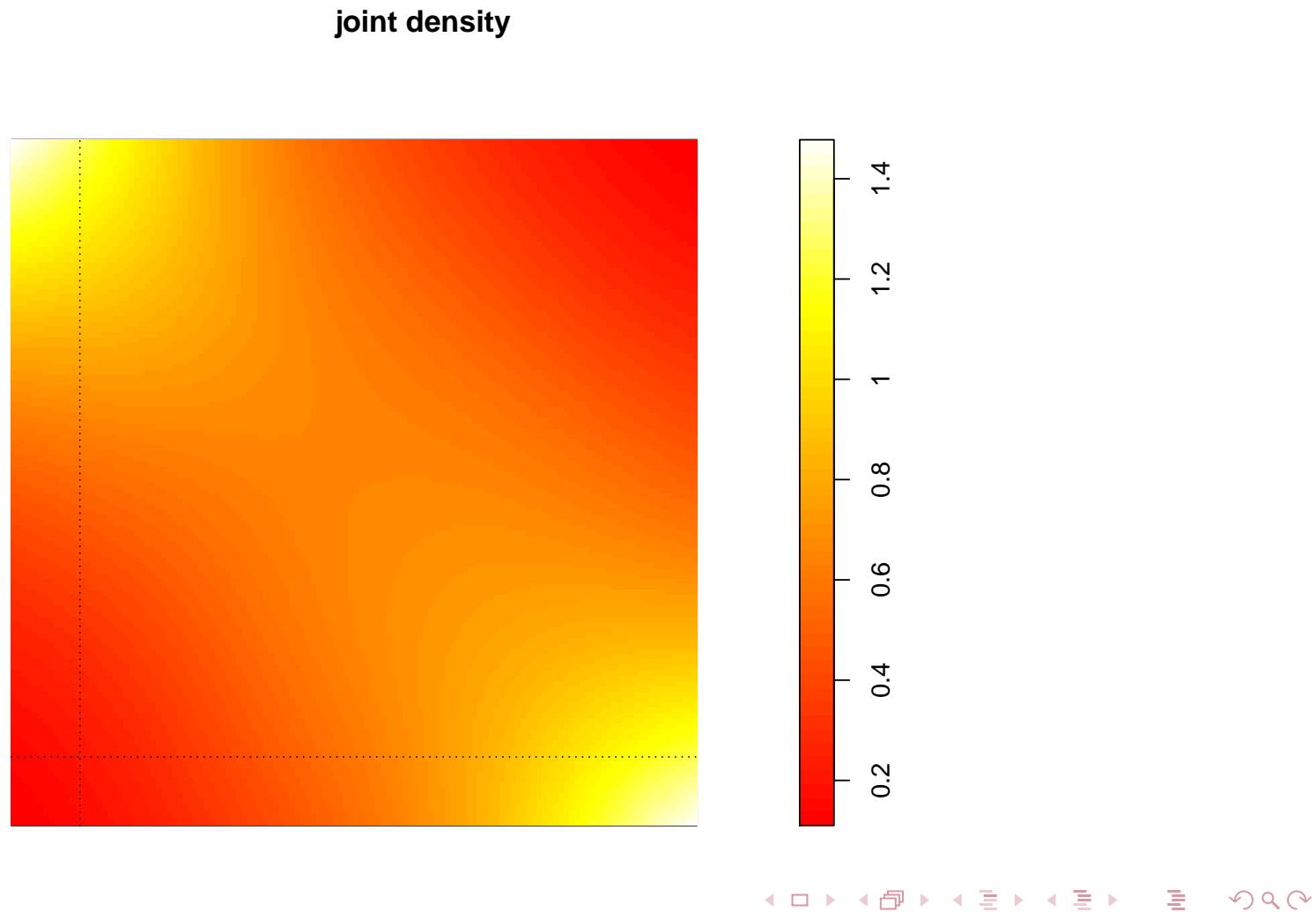
$$P = 1 - (1 - \alpha)^m$$

Probability of AT LEAST one Type I error as a function of number of tests



Dependent tests

In most of the real cases:

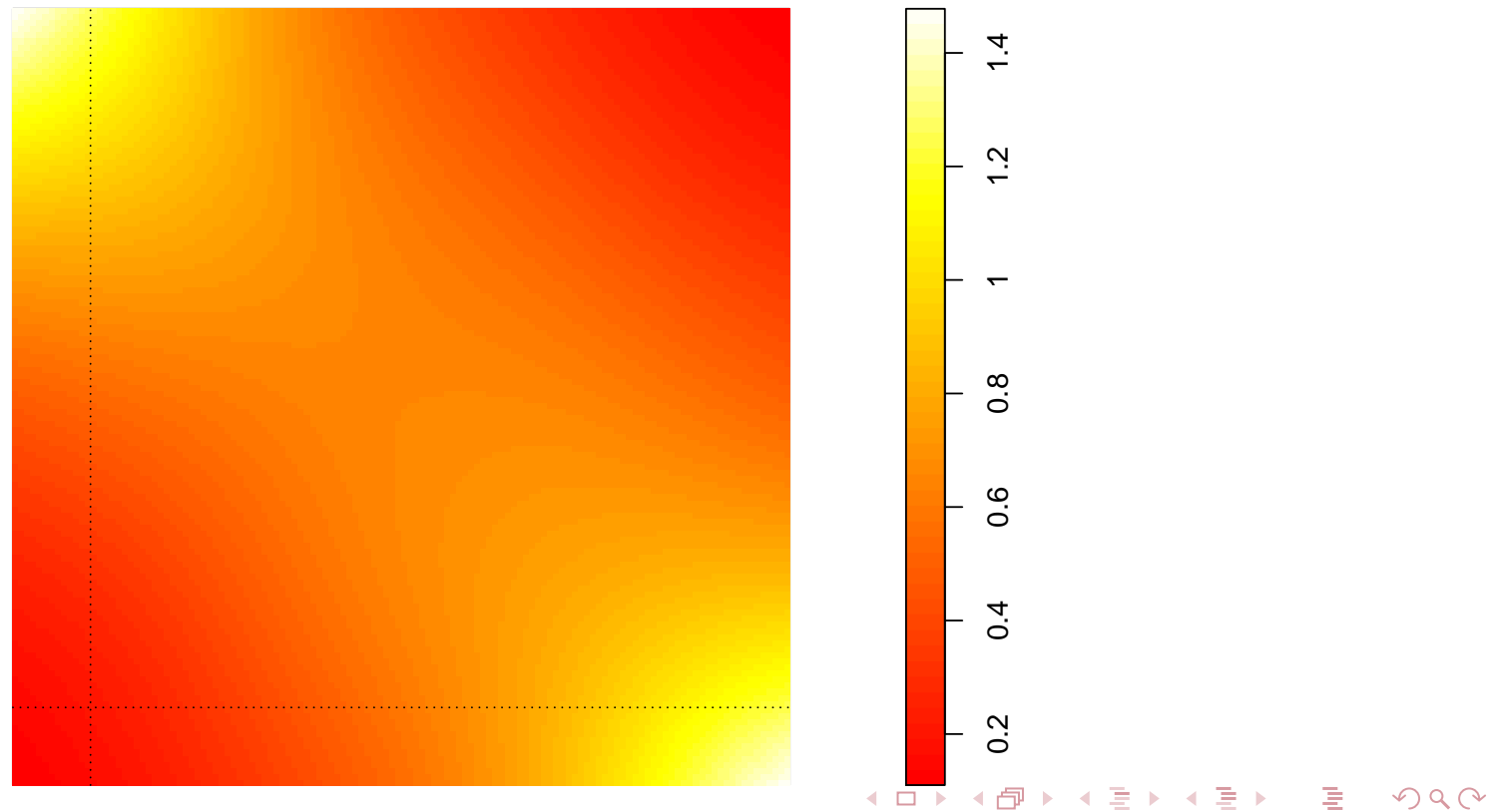


Dependent tests

In most of the real cases:

$$P(\text{At least one false rejection}) > (!) 1 - (1 - \alpha)^2$$

joint density



Type I errors

How to define the type I error when there are many hypotheses?

Which procedures control this error?

Outline

Introduction

Some examples

Some remarks

FamilyWise Error Rate (FWER)

Definition + Bonferroni

Holm (step-wise)

Closed Testing

FWER control by Permutations tests

Permutation Bonferroni

Westfall & Young: permutation Holm²

Closed Testing

Conclusion

²Westfall PH, Young SS (1993) Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment. Wiley

FamilyWise Error Rate (FWER)

Probability of AT LEAST one false rejection

Bonferroni Inequality

Reject H_i if $p_i \leq \alpha/m$ (m = number of hypotheses)

control FWER

$$\begin{aligned}\text{FWER} &= P(p_i \leq \alpha/m \text{ for at least one True null hypo}) \\ &= P\left(\bigcup_{i \in \{\text{true null hypotheses}\}} \{p_i \leq \alpha/m\}\right) \\ &\leq \sum_{i \in \{\text{true null hypotheses}\}} P(p_i \leq \alpha/m) \\ &\leq \#\{\text{true null hypotheses}\} \frac{\alpha}{m} \leq \alpha\end{aligned}$$

Bonferroni procedure

Multiplicity-adjusted p-value

$\tilde{p}_i = mp_i$ $i = 1, \dots, m$ and rejects it if $\leq \alpha$

Pros

- Very easy
- Control the FWER under any dependence

Cons

Conservative (Adj. P-values very high, few rejections)

Holm's procedures ³

1. First step: adjusted p-value: $p \cdot m$; reject if $\leq \alpha$
2. After r rejections, adjusted p-value: $p \cdot (m - r)$
3. just do not reject anything, Stop.

Bonferroni

Adj. p-value: p_A^5 p_B^5 p_C^5 p_D^5 p_E^5 $\leq ?\alpha$ $\mathcal{H} \setminus \mathcal{R}: \quad \boxed{A} \quad \boxed{B} \quad \boxed{C} \quad \boxed{D} \quad \boxed{E}$ $\mathcal{R}:$

³Holm S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*; 6 (2): 65-70. ◀ ▶

Holm's procedures ³

1. First step: adjusted p-value: $p \cdot m$; reject if $\leq \alpha$
2. After r rejections, adjusted p-value: $p \cdot (m - r)$
3. just do not reject anything, Stop.

Suppose p_A and p_C significant

Adj. p-value: $p_A^5 \quad p_B^5 \quad p_C^5 \quad p_D^5 \quad p_E^5 \leq ?\alpha$ $\mathcal{H} \setminus \mathcal{R}: \quad \boxed{\text{A}} \quad \boxed{\text{B}} \quad \boxed{\text{C}} \quad \boxed{\text{D}} \quad \boxed{\text{E}}$ $\mathcal{R}:$

³Holm S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*; 6 (2): 65-70. ◀ ▶

Holm's procedures ³

1. First step: adjusted p-value: $p \cdot m$; reject if $\leq \alpha$
2. After r rejections, adjusted p-value: $p \cdot (m - r)$
3. just do not reject anything, Stop.

Adjusted p-value: $p \cdot 3$

Adj. p-value: $-p_{B^3} - p_{D^3} - p_{E^3} \leq \alpha$

 $\mathcal{H} \setminus \mathcal{R}:$ B D E
$$\mathcal{R}: \quad \boxed{A} \quad \boxed{C}$$

³Holm S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*; 6 (2): 65-70. ◀ ▶

Holm's procedures ³

1. First step: adjusted p-value: $p \cdot m$; reject if $\leq \alpha$
2. After r rejections, adjusted p-value: $p \cdot (m - r)$
3. just do not reject anything, Stop.

Suppose p_D significant

Adj. p-value: $-p_B^3 - p_D^3 - p_E^3 \leq ?\alpha$

 $\mathcal{H} \setminus \mathcal{R}: \quad \boxed{\text{B}} \quad \boxed{\text{D}} \quad \boxed{\text{E}}$
$$\mathcal{R}: \quad \boxed{A} \quad \boxed{C}$$

³Holm S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*; 6 (2): 65-70. ◀ ▶

Holm's procedures ³

1. First step: adjusted p-value: $p \cdot m$; reject if $\leq \alpha$
2. After r rejections, adjusted p-value: $p \cdot (m - r)$
3. just do not reject anything, Stop.

Adjusted p-value: $p \cdot 2$

Adj. p-value: - p_{B^2} - - $p_{E^2} \leq ?\alpha$

$$\mathcal{H} \setminus \mathcal{R}: \quad \boxed{\text{B}} \quad \boxed{\text{E}}$$
$$\mathcal{R}: \quad \boxed{A} \quad \boxed{C} \quad \boxed{D}$$

³Holm S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*; 6 (2): 65-70. ◀ ▶

Holm's procedures ³

1. First step: adjusted p-value: $p \cdot m$; reject if $\leq \alpha$
2. After r rejections, adjusted p-value: $p \cdot (m - r)$
3. just do not reject anything, Stop.

No rejections: Stop

Adj. p-value: - p_{B^2} - - $p_{E^2} \leq ?\alpha$

$$\mathcal{H} \setminus \mathcal{R}: \quad \boxed{\text{B}} \quad \boxed{\text{E}}$$
$$\mathcal{R}: \quad \boxed{A} \quad \boxed{C} \quad \boxed{D}$$

³Holm S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*; 6 (2): 65-70. ◀ ▶

Closed Testing⁴

Closure of hypos (all possible intersections)

hypos to test

A

B

C

⁴R Marcus, E Peritz, KR Gabriel (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 63: 655-660. ▶

Closed Testing⁴

Test the overall null hypo (eg MANOVA)

Closure

ABC

AB

AC

BC

A

B

C

⁴R Marcus, E Peritz, KR Gabriel (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 63: 655-660.

Closed Testing⁴

test at level α

ABC α

AB

AC

BC

A

B

C

⁴R Marcus, E Peritz, KR Gabriel (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 63: 655-660.

Closed Testing⁴

Suppose significant

ABC -

AB

AC

BC

A

B

C

⁴R Marcus, E Peritz, KR Gabriel (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 63: 655-660.

Closed Testing⁴

Go on

ABC -

AB α AC α BC α

A

B

C

⁴R Marcus, E Peritz, KR Gabriel (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 63: 655-660.

Closed Testing⁴

Test at level α

ABC -

AB - AC - BC α

A

B

C

⁴R Marcus, E Peritz, KR Gabriel (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 63: 655-660.

Closed Testing⁴

Go on

$$\boxed{ABC} -$$

$$\boxed{AB} - \boxed{AC} - \boxed{BC} \alpha$$

$$\boxed{A} \alpha \quad \boxed{B} \quad \boxed{C}$$

⁴R Marcus, E Peritz, KR Gabriel (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 63: 655-660.

Closed Testing⁴

A is signif. if all above are signif.

ABC -

AB - AC - BC α

A - B C

⁴R Marcus, E Peritz, KR Gabriel (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 63: 655-660.

Closed Testing⁴

Cons: too much tests very quickly: $= 2^{\#hypotheses} - 1$

A is signif. if all above are signif.

ABC -

AB - AC - BC α

A - B C

⁴R Marcus, E Peritz, KR Gabriel (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 63: 655-660.

Outline

Introduction

Some examples

Some remarks

FamilyWise Error Rate (FWER)

Definition + Bonferroni

Holm (step-wise)

Closed Testing

FWER control by Permutations tests

Permutation Bonferroni

Westfall & Young: permutation Holm⁵

Closed Testing

Conclusion

⁵Westfall PH, Young SS (1993) Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment. Wiley

Bonferroni is conservative

Bonferroni bound

Reject for p-values at most α/m

By Boole's inequality

Guaranteed: $\text{FWER} \leq \alpha$, but often $\text{FWER} < \alpha$

Can we improve?

Reject for p-values at most $\tilde{\alpha} > \alpha/m$, while keeping FWER control

Yes we can

By permutations

Improved Bonferroni

Reduced α

Reject H_i if $p_i \leq \tilde{\alpha}$

Control of FWER?

$$\begin{aligned}\text{FWER} &= P(p_i \leq \tilde{\alpha} \text{ for at least one } i \text{ with } H_i \text{ true}) \\ &= P\left(\bigcup_{i \in T} \{p_i \leq \tilde{\alpha}\}\right) \\ &= P\left(\min_{i \in T} p_i \leq \tilde{\alpha}\right) \leq \alpha\end{aligned}$$

How can we determine the value of $\tilde{\alpha}$?

Using permutations to find the distribution of the minimum p-value

Multiple testing using permutations

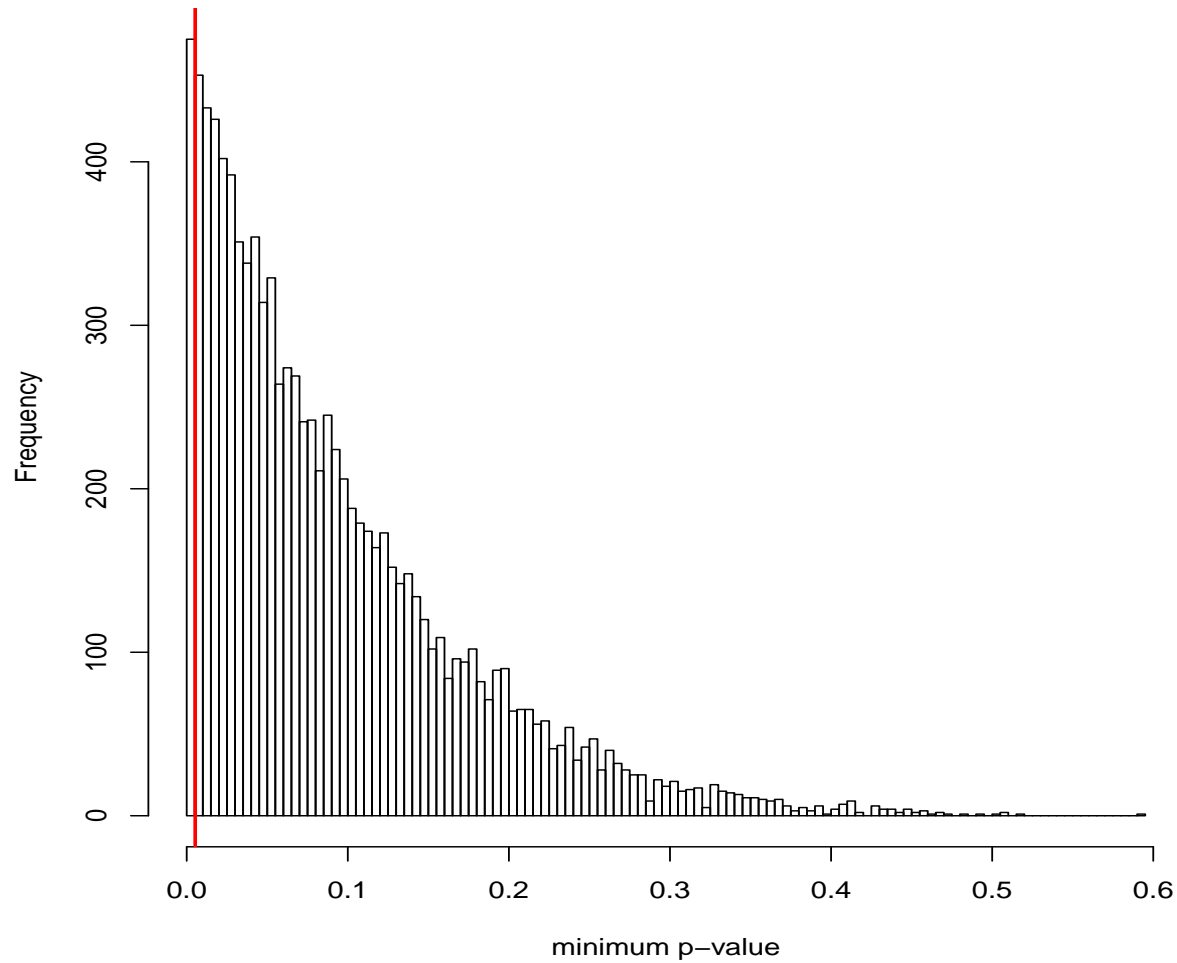
The single step min-P method

1. Calculate the smallest p-value m for the real data
2. Randomly permute the data
3. Calculate new p-values for all tests based on permuted data
4. Calculate the smallest p-value m^π for permuted data
5. Repeat permutation many (say $k=1000$) times: m_1^π, \dots, m_k^π
6. Calculate $\tilde{\alpha}$ as the α -quantile of m_1^π, \dots, m_k^π

Multiple testing result

Reject all hypotheses with (non-permuted) p-values at most $\tilde{\alpha}$

Histogram of the smallest p-value of 10 independent tests



Correlation structure of p-values

Permutation

- Destroys correlation between covariates and response
- Retains correlation among covariates

Consequence

- P-values of correlated genes remain correlated in permutations
- Distribution of minimum p-value correctly takes correlations into account

Gain relative to Bonferroni

Min-P permutation critical value

In the example data: $\tilde{\alpha} = 0.00527$ vs $\alpha/m = 0.005$

When is the gain large?

Negatively correlated p-values: typically no gain

Independent p-values: minimal gain

Positively correlated p-values: gain can be large

Genomic data

Genes typically strongly positively correlated

Sequential permutation multiple testing

Single step

Single step min-P is permutation equivalent of Bonferroni

What about Holm?

Permutation equivalent of Holm's method: Westfall & Young

Westfall & Young min-P

The min-P algorithm

- Start with all hypotheses
- Repeat
 - Do single step min-P to calculate $\tilde{\alpha}$
 - Reject hypotheses with p-value $\leq \tilde{\alpha}$
 - Remove rejected hypotheses
- Until no new rejections occur

Westfall & Young min-p (and max-t)

Westfall & Young min-p: shortcut using min-p combining function (m tests).

Westfall & Young min-p (and max-t)

Westfall & Young min-p: shortcut using min-p combining function (m tests).

Suppose three hypotheses tested and $p_A \leq p_B \leq p_C$

- Test H_A , H_B and H_C using min-p: p_{ABC}
 - if $p_{ABC} \leq \alpha$ reject H_A and go on
 - if $p_{ABC} > \alpha$ STOP
- Test H_B and H_C using min-p: p_{BC}
 - if $p_{BC} \leq \alpha$ reject H_B and go on
 - if $p_{BC} > \alpha$ STOP
- Test H_C : p_C
 - if $p_C \leq \alpha$ reject H_C and STOP
 - if $p_C > \alpha$ STOP

For max-t procedure substitute $p_{ABC} \leq \alpha$ with $t_{ABC} \geq t_\alpha$

Closed Testing⁶

Test in each node: any multivariate permutation test
(eg alternative to Manova)

Closure Set

ABC

AB

AC

BC

A

B

C

$$\text{Adjusted } \tilde{p}_A = \max(p_A, p_{AB}, p_{AC}, p_{ABC})$$

⁶R Marcus, E Peritz, KR Gabriel (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 63: 655-660.

Outline

Introduction

Some examples

Some remarks

FamilyWise Error Rate (FWER)

Definition + Bonferroni

Holm (step-wise)

Closed Testing

FWER control by Permutations tests

Permutation Bonferroni

Westfall & Young: permutation Holm⁷

Closed Testing

Conclusion

⁷Westfall PH, Young SS (1993) Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment. Wiley

Accounting for dependences: adjusted p-value lower (i.e. more rejections)

When?

Negative correlation: generally no gain

p-value Independents: little or no gain

Positive correlation: big gain, usually

(NB: a test with bi-directional alternative and with negative correlation produce p-value positively correlated)

How?

```
in R: library(flip); flip(); flip.adjust()
```

Real data

The variables of real datasets are often correlated
then ...

Accounting for dependences: adjusted p-value lower (i.e. more rejections)

When?

Negative correlation: generally no gain

p-value Independents: little or no gain

Positive correlation: big gain, usually

(NB: a test with bi-directional alternative and with negative correlation produce p-value positively correlated)

How?

```
in R: library(flip); flip(); flip.adjust()
```

Real data

The variables of real datasets are often correlated
then ... **permutations are (often) convenient**

Summary

Error FamilyWise

- Generalize the Type I error to the case of multiple hypotheses

Summary

Error FamilyWise

- Generalize the Type I error to the case of multiple hypotheses
- Control the probability of at LEAST a false rejection

Summary

Error FamilyWise

- Generalize the Type I error to the case of multiple hypotheses
- Control the probability of at LEAST a false rejection
- corrects the p-value (adjusted p-value always equal to or greater than the unadjusted p-value)

Summary

Error FamilyWise

- Generalize the Type I error to the case of multiple hypotheses
- Control the probability of at LEAST a false rejection
- corrects the p-value (adjusted p-value always equal to or greater than the unadjusted p-value)

R Software

- Bonferroni and Holm `library(stats); p.adjust()`
- Testing Closed `library(cherry); closed()`
- Permutations - Westfall & Young
`library(flip); flip.adjust()`