

Permutation tests for Clinical Trials

Livio Finos

Outline

Introduction

A toy example

A Naive approach to Permutation Testing

Theory (very short)

Other cases

Multivariate Testing

Introduction

- Well established nonparametric approach to **inference**: Fisher, 1935; Pitman, 1937; Pitman, 1938.
- (In general) it requires less assumptions about the data generating process than the parametric counterpart.
- Very good inferential properties, typically:
 - exactness (i.e. exact control of the type I error)
 - asymptotically optimality and convergence to the parametric counterpart when it does exist.

Introduction

- Fisher exact test is a prototypical example, but
- the general approach has restricted applicability without the support of a computer.

Renewed interest toward permutation testing

- A milestone: Westfall and Young (1993). Resampling-Based Multiple Testing: Examples and Methods for p-value Adjustment. Wiley.
- Many active areas of research adopt these methods in their daily statistical analysis (e.g. genetics and neuroscience: Nichols and Holmes (2002); Pantazis et al. (2009); Winkler et al. (2014)).
- Permutation approach:
 - Ideal for **randomized experimental design**
 - deals with very complex models, without formal definition of the data generating process.

Outline

Introduction

A toy example

A Naive approach to Permutation Testing

Theory (very short)

Other cases

Multivariate Testing

A Naive approach to Permutation Testing

Comparison of Two Samples (i.e. one factor with two levels)

- **Control:** 3 observations,
- **Treated:** 3 observations:

1.025, 1.949, 3.477, 2.391, 3.676, 4.816

Hypothesis testing

- H_0 : Two groups are equal
- H_1 : Treated is greater than Control (on average)

p -value: probability to get the observed evidence against H_0 if the two groups were equal (i.e. H_0 were true)

Test: if $p \leq \alpha$ (e.g. $\alpha = .05$): we decide for H_1 ,
otherwise: we stay with H_0

Parametric approach

Assumptions on y_1, y_2, \dots, y_6

- independent
 - identically distributed
 - normally distributed
- OR
- have finite mean and variance (but inference is only asymptotically valid in this case!)

Parametric approach

Assumptions on y_1, y_2, \dots, y_6

- independent
- identically distributed
 - normally distributed
- OR
- have finite mean and variance (but inference is only asymptotically valid in this case!)

We can perform a t-test:

$$T = \frac{\bar{y}(Treated) - \bar{y}(Control)}{\widehat{sd(\bar{y}(Treated) - \bar{y}(Control))}} \sim t_4$$

(i.e. T test statistic follow a t distribution with $n-2=4$ d.f.)

Parametric approach

With toy data:

$t = -1.4545$,

$df = 4$,

$p\text{-value} = 0.1098$

Remark

The hypotheses tested are:

- $H_0 : \mu_{Treated} = \mu_{Control}$
- $H_1 : \mu_{Treated} > \mu_{Control}$ (only a difference in mean is allowed)

Outline

Introduction

A toy example

A Naive approach to Permutation Testing

Theory (very short)

Other cases

Multivariate Testing

A Naive approach to Permutation Testing

The p -value is computed **under** H_0 :
Controls and **Treated** have the **same distribution**.

Collection of equally likely outcomes:

$$f(1.025, 1.949, 3.477, 2.391, 3.676, 4.816) =$$

A Naive approach to Permutation Testing

The p -value is computed **under** H_0 :
Controls and **Treated** have the **same distribution**.

Collection of equally likely outcomes:

$$\begin{aligned} & f(1.025, 1.949, 3.477, 2.391, 3.676, 4.816) = \\ = & f(1.025, 1.949, 3.477, 2.391, 3.676, 4.816) = \end{aligned}$$

A Naive approach to Permutation Testing

The p -value is computed **under** H_0 :
Controls and **Treated** have the **same distribution**.

Collection of equally likely outcomes:

$$\begin{aligned} & f(1.025, 1.949, 3.477, 2.391, 3.676, 4.816) = \\ = & f(1.025, 1.949, 3.477, 2.391, 3.676, 4.816) = \\ = & f(1.025, 1.949, 3.477, 2.391, 3.676, 4.816) = \end{aligned}$$

A Naive approach to Permutation Testing

The p -value is computed **under** H_0 :
Controls and **Treated** have the **same distribution**.

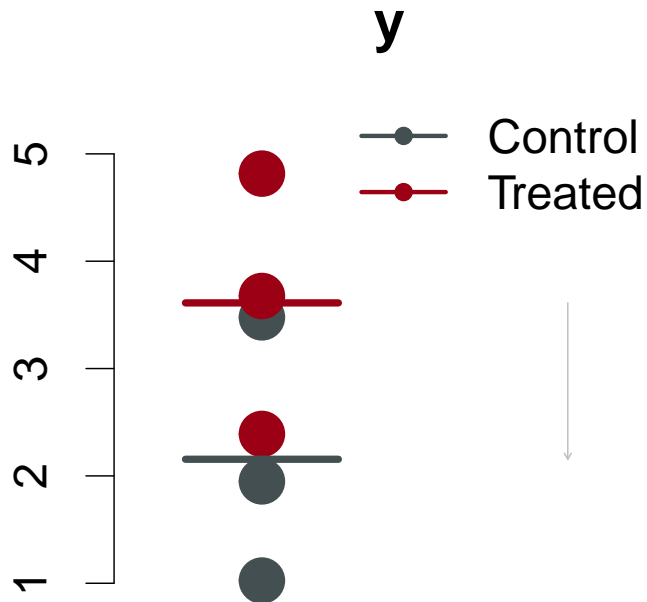
Collection of equally likely outcomes:

$$\begin{aligned} & f(1.025, 1.949, 3.477, 2.391, 3.676, 4.816) = \\ = & f(1.025, 1.949, 3.477, 2.391, 3.676, 4.816) = \\ = & f(1.025, 1.949, 3.477, 2.391, 3.676, 4.816) = \\ = & \dots \end{aligned}$$

There are $\binom{6}{3} = \frac{6!}{3!3!} = 20$ equally likely outcomes

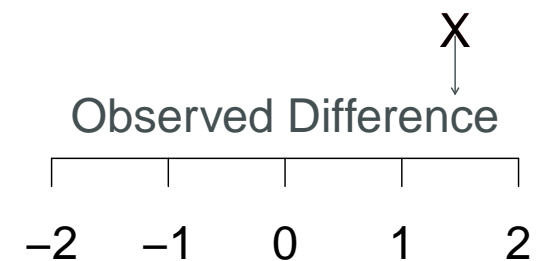
A Naive approach to Permutation Testing

Compute the difference in mean of the two samples



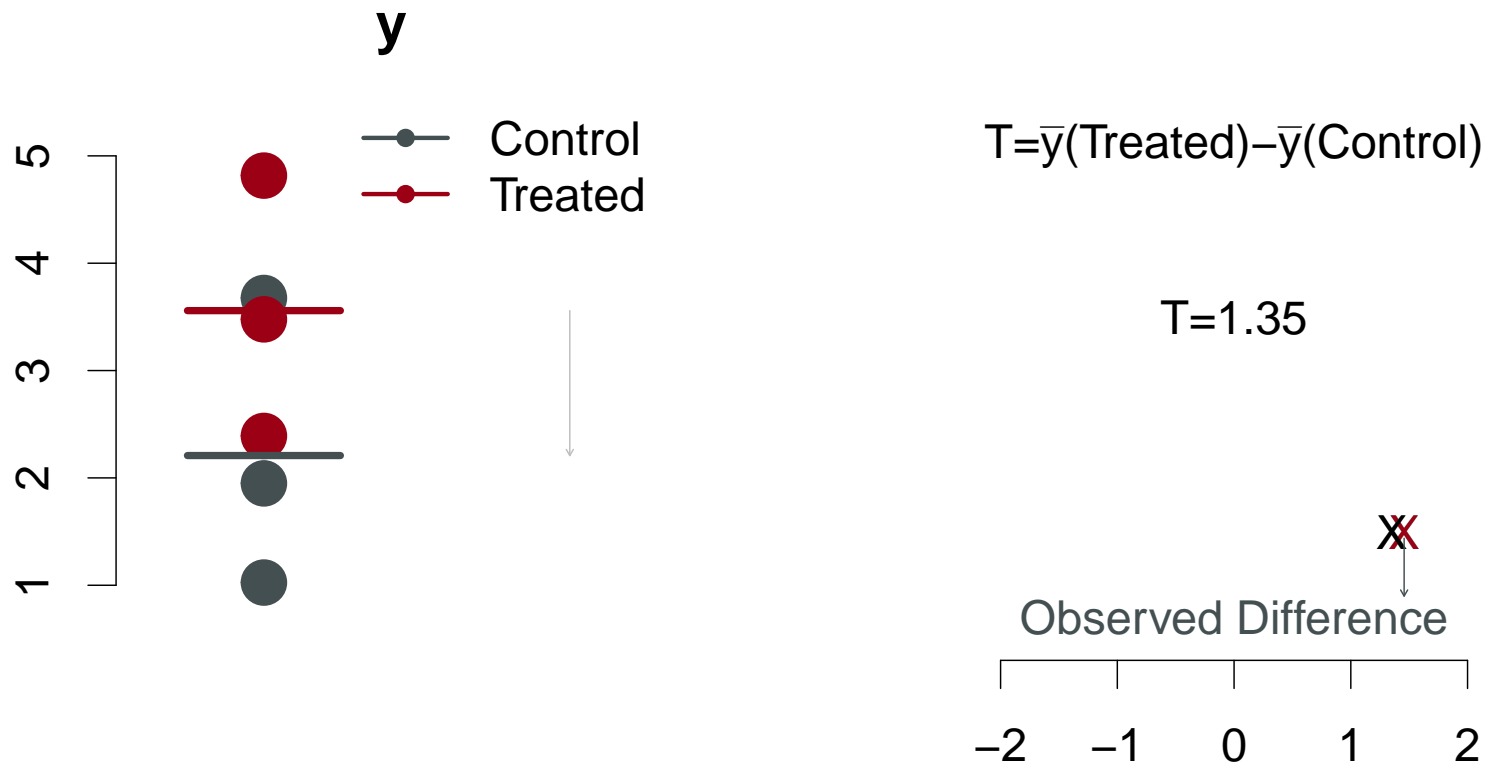
$$T = \bar{y}(\text{Treated}) - \bar{y}(\text{Control})$$

$$T = 1.46$$



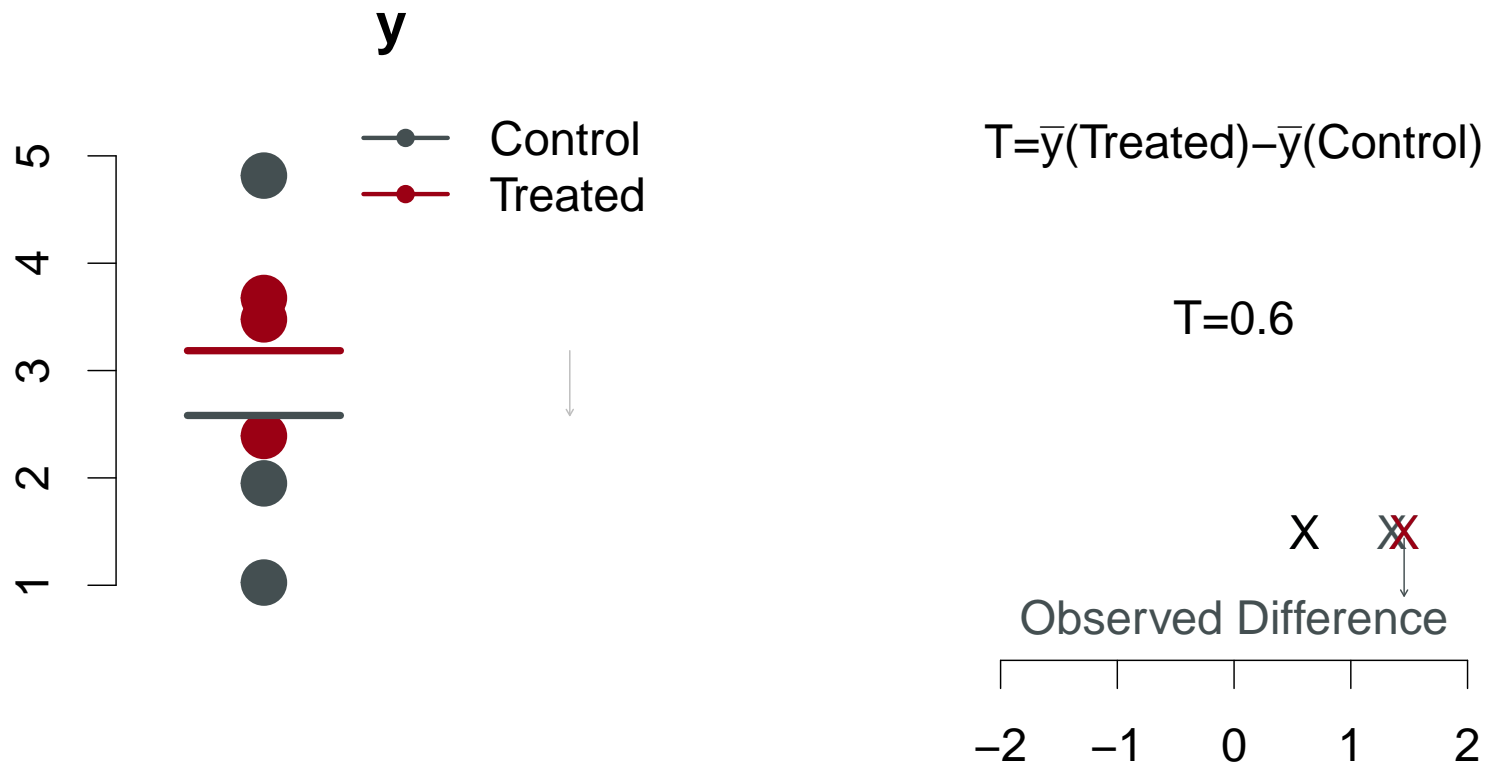
A Naive approach to Permutation Testing

Compute the same difference on another hypothetical experiment



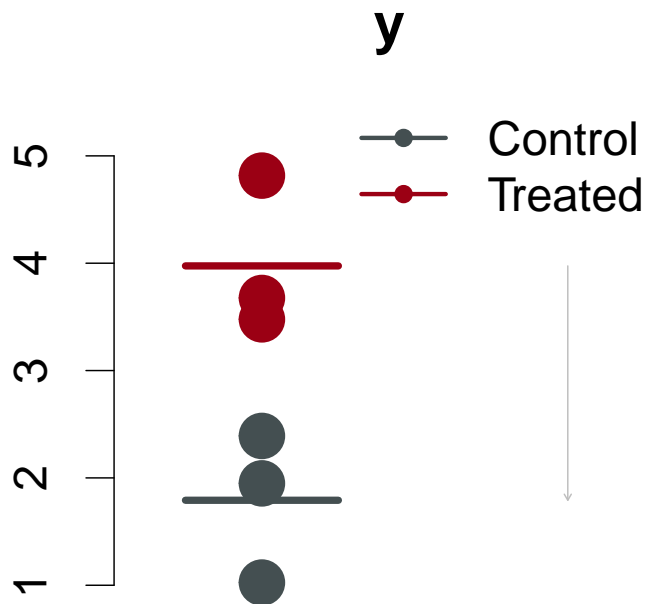
A Naive approach to Permutation Testing

...and go on with all hypothetical experiments...



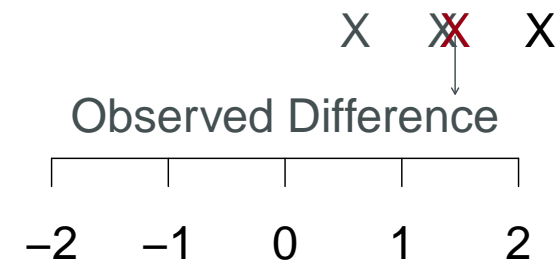
A Naive approach to Permutation Testing

...and go on with all hypothetical experiments...



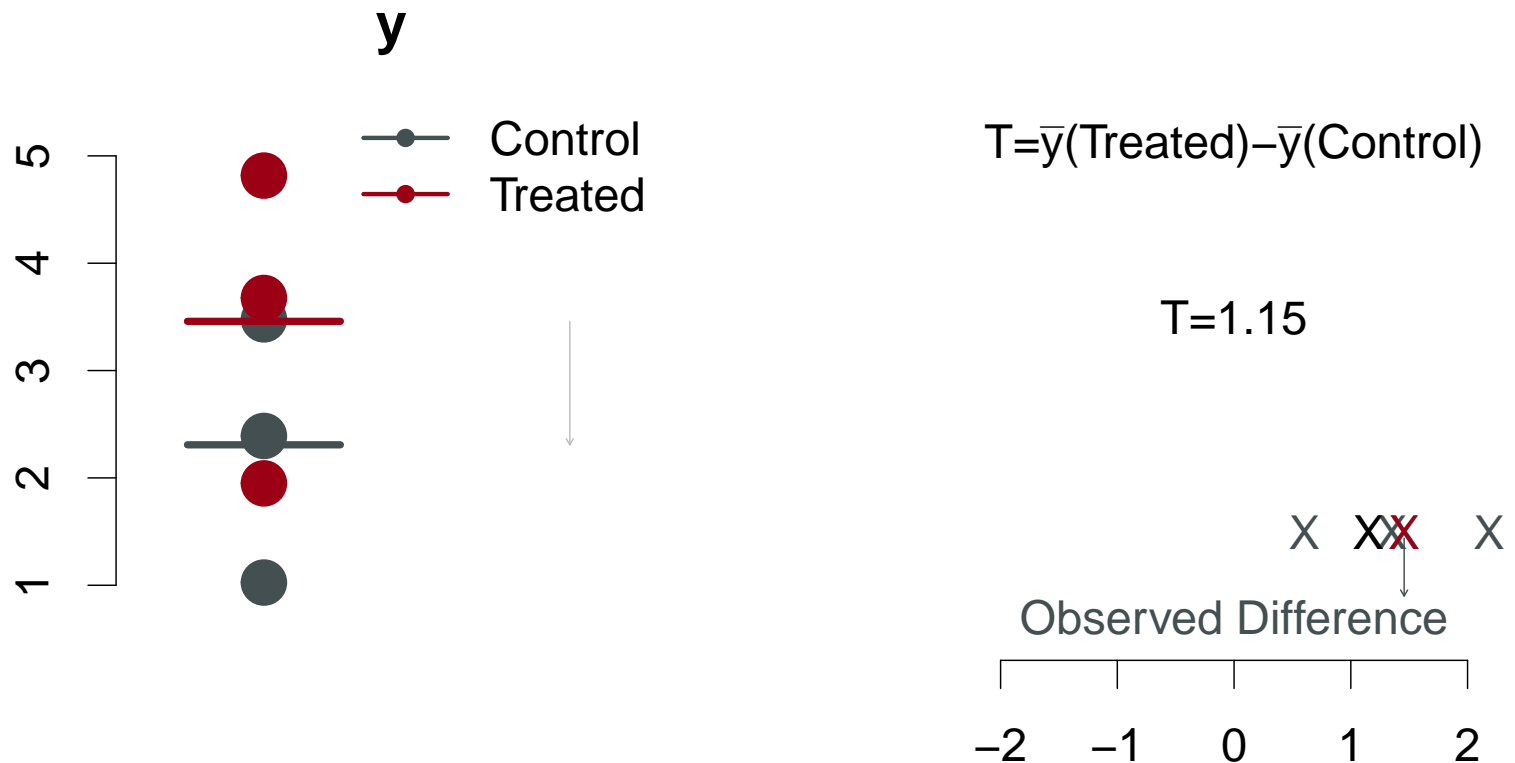
$$T = \bar{y}(\text{Treated}) - \bar{y}(\text{Control})$$

$$T = 2.18$$



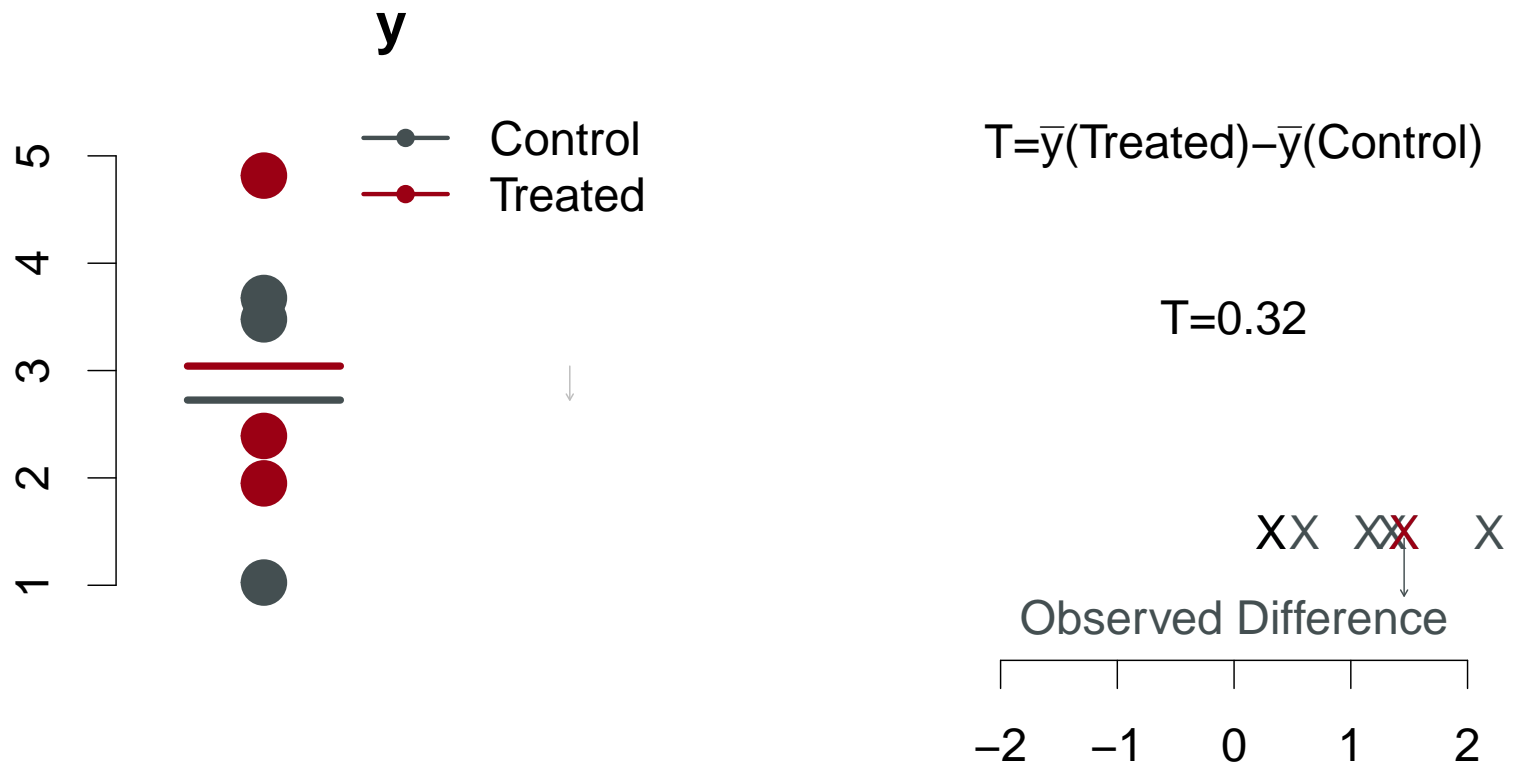
A Naive approach to Permutation Testing

...and go on with all hypothetical experiments...



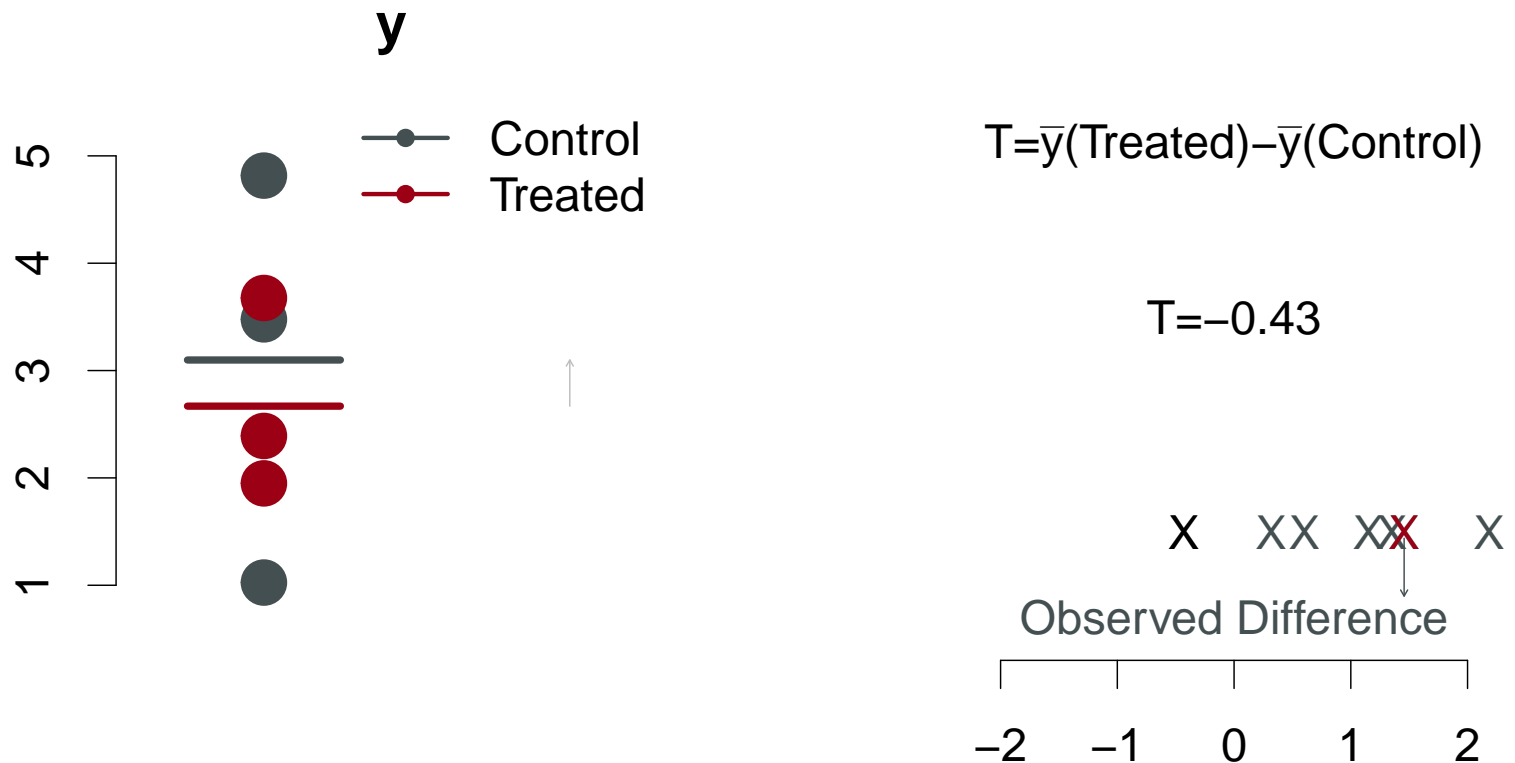
A Naive approach to Permutation Testing

...and go on with all hypothetical experiments...



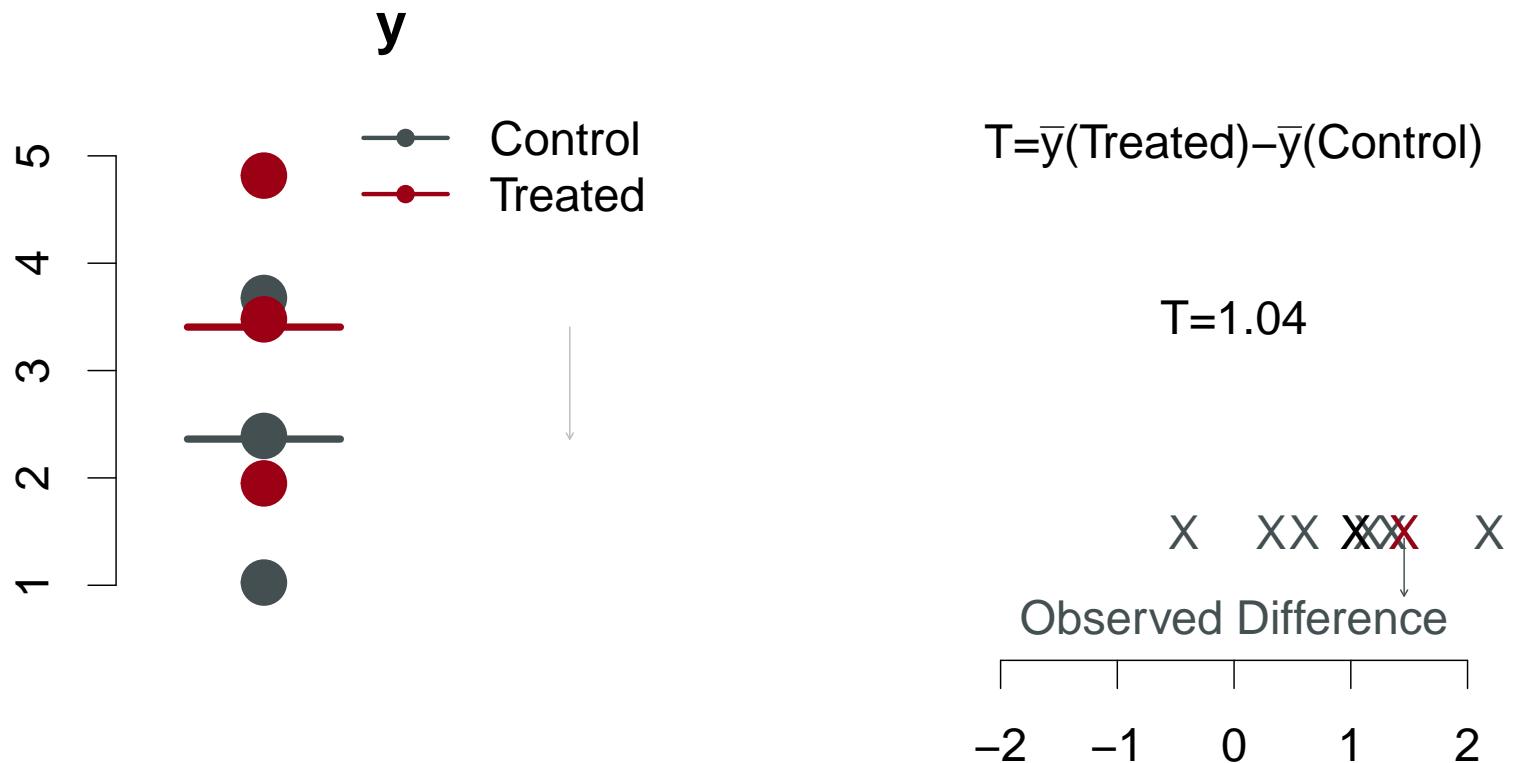
A Naive approach to Permutation Testing

...and go on with all hypothetical experiments...



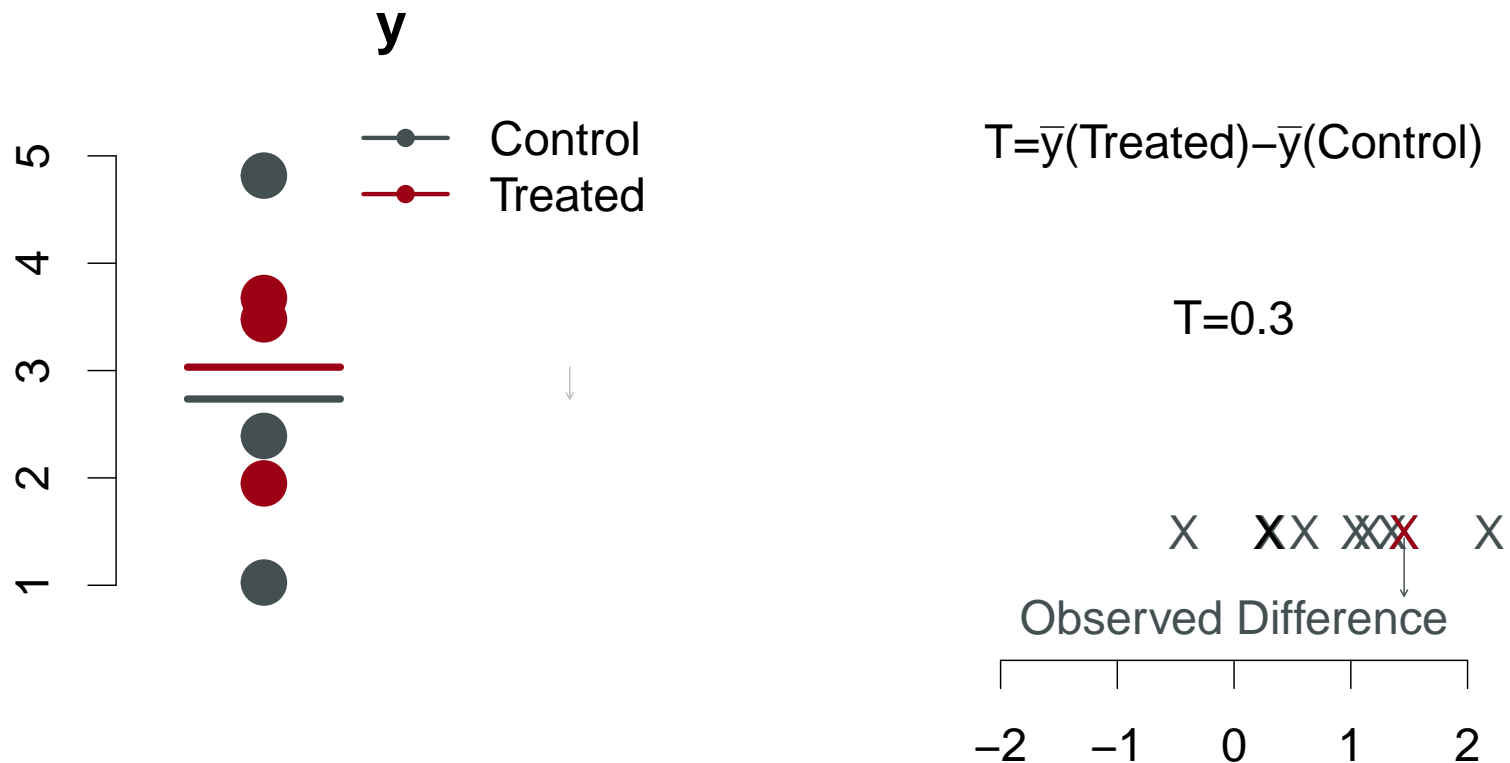
A Naive approach to Permutation Testing

...and go on with all hypothetical experiments...



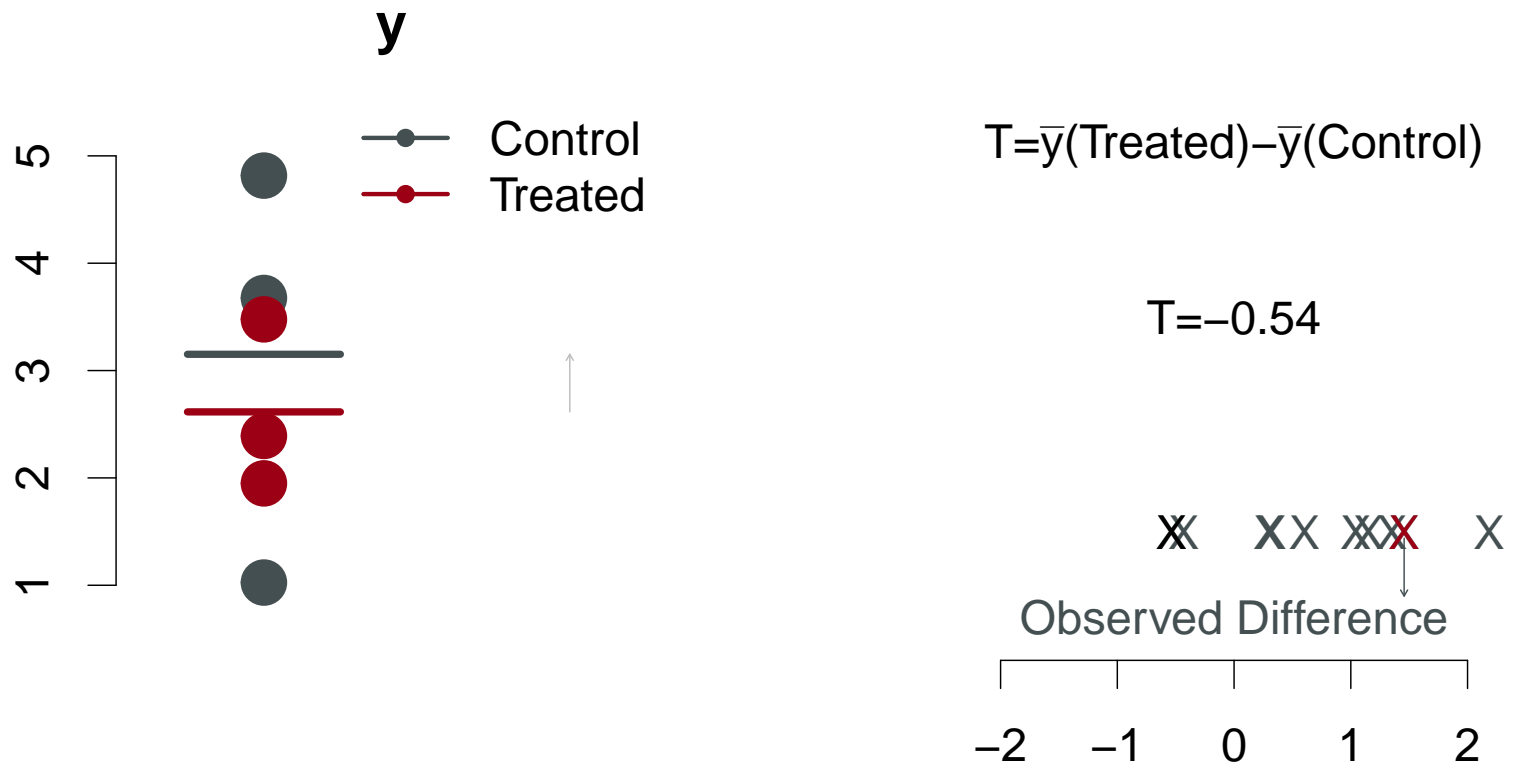
A Naive approach to Permutation Testing

...and go on with all hypothetical experiments...



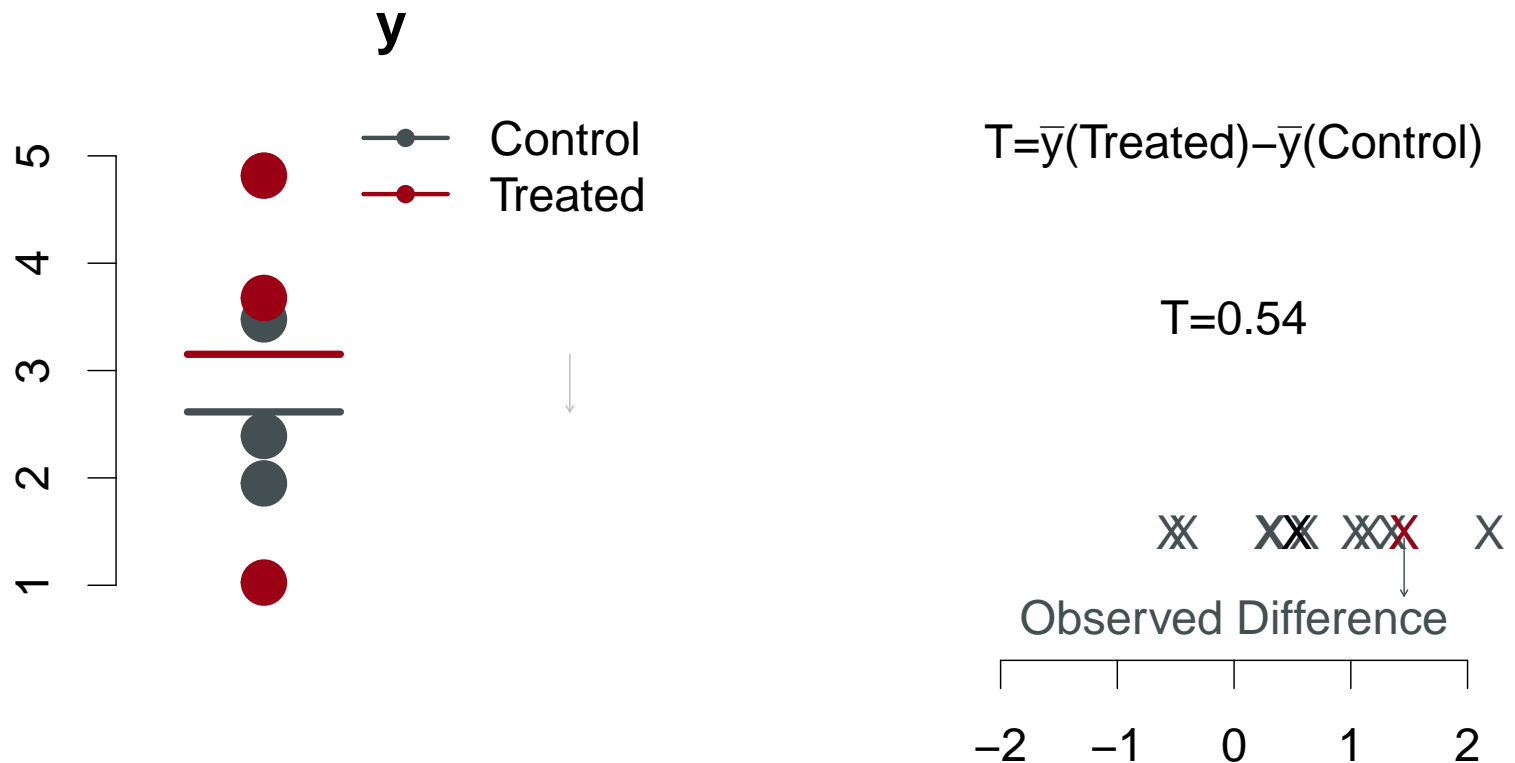
A Naive approach to Permutation Testing

...and go on with all hypothetical experiments...



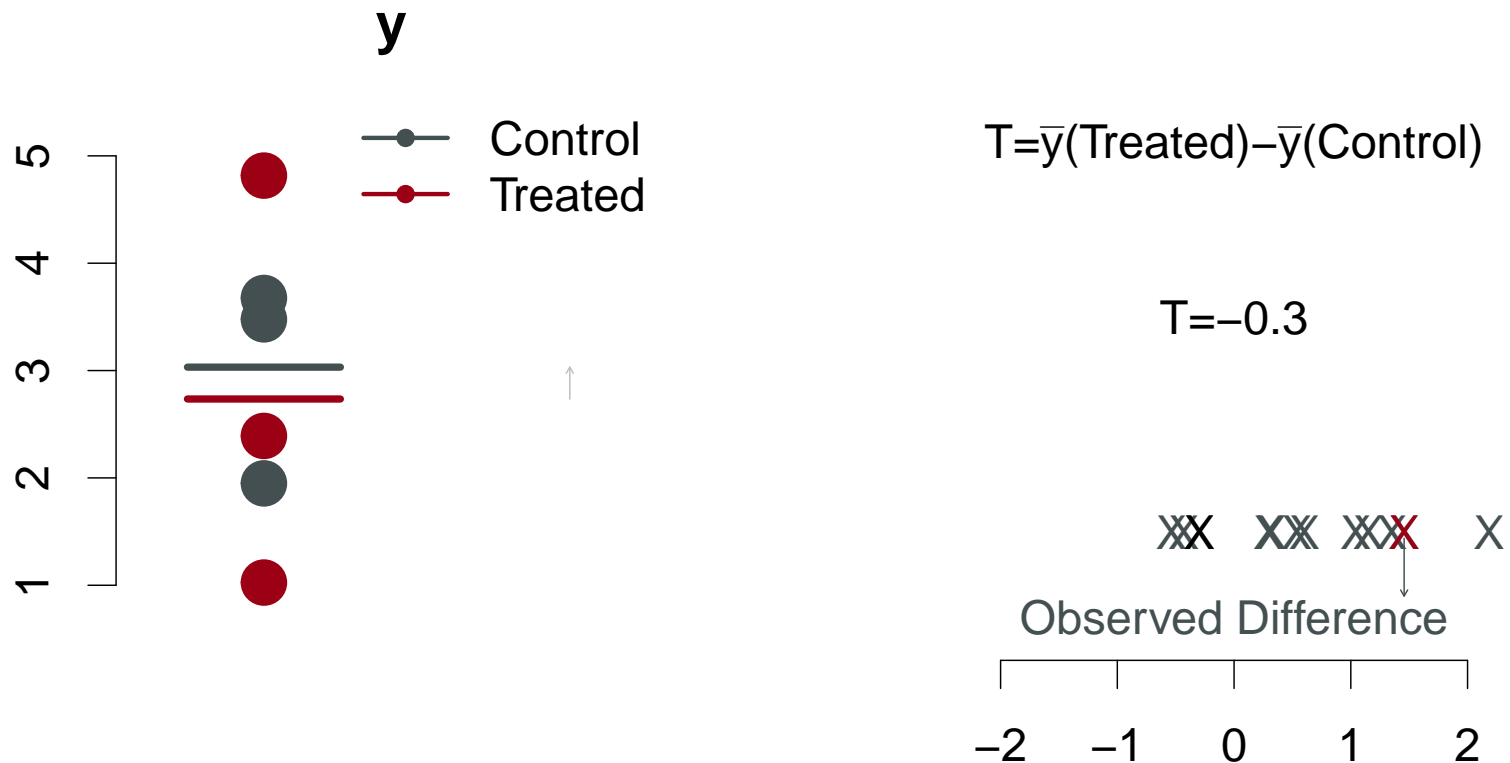
A Naive approach to Permutation Testing

...and go on with all hypothetical experiments...



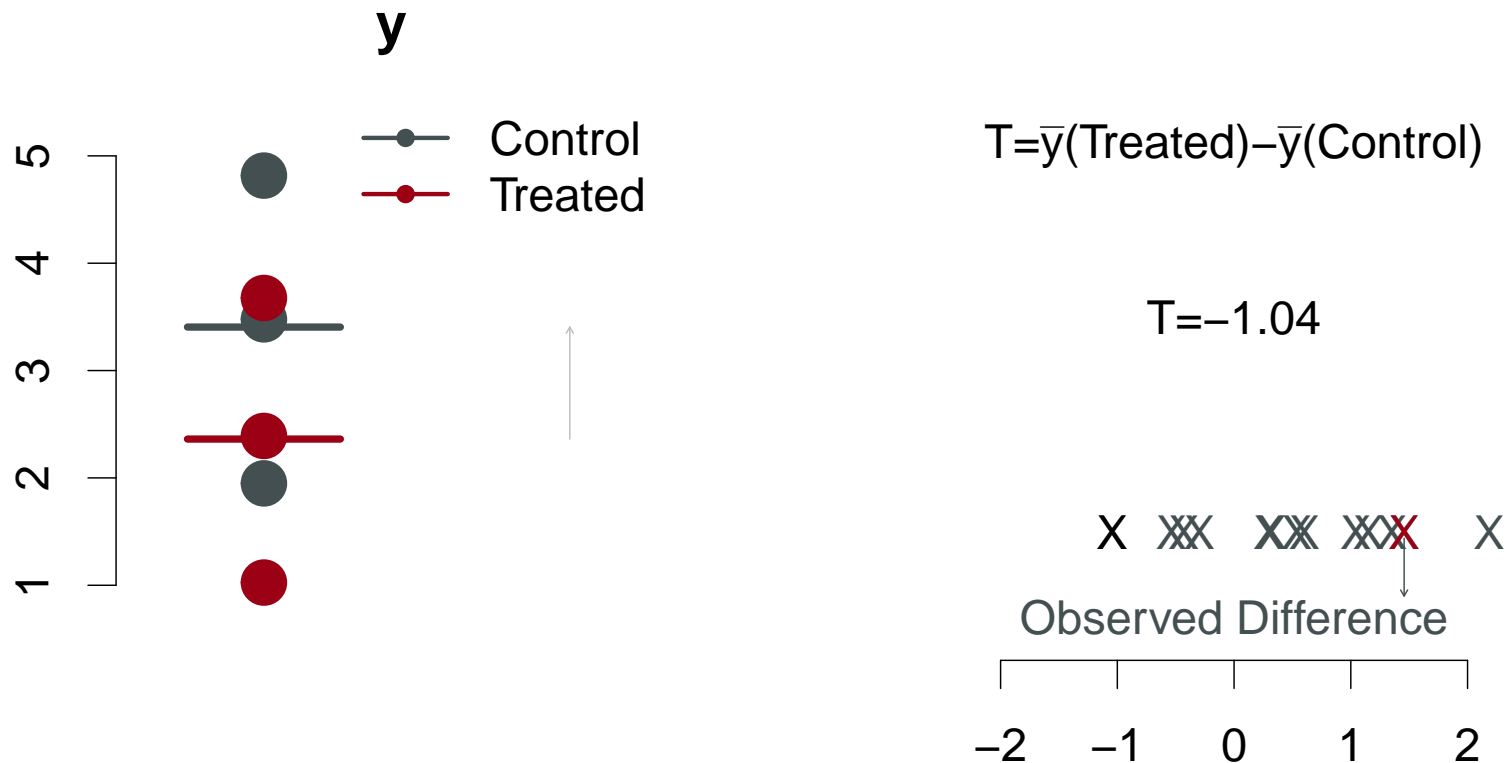
A Naive approach to Permutation Testing

...and go on with all hypothetical experiments...



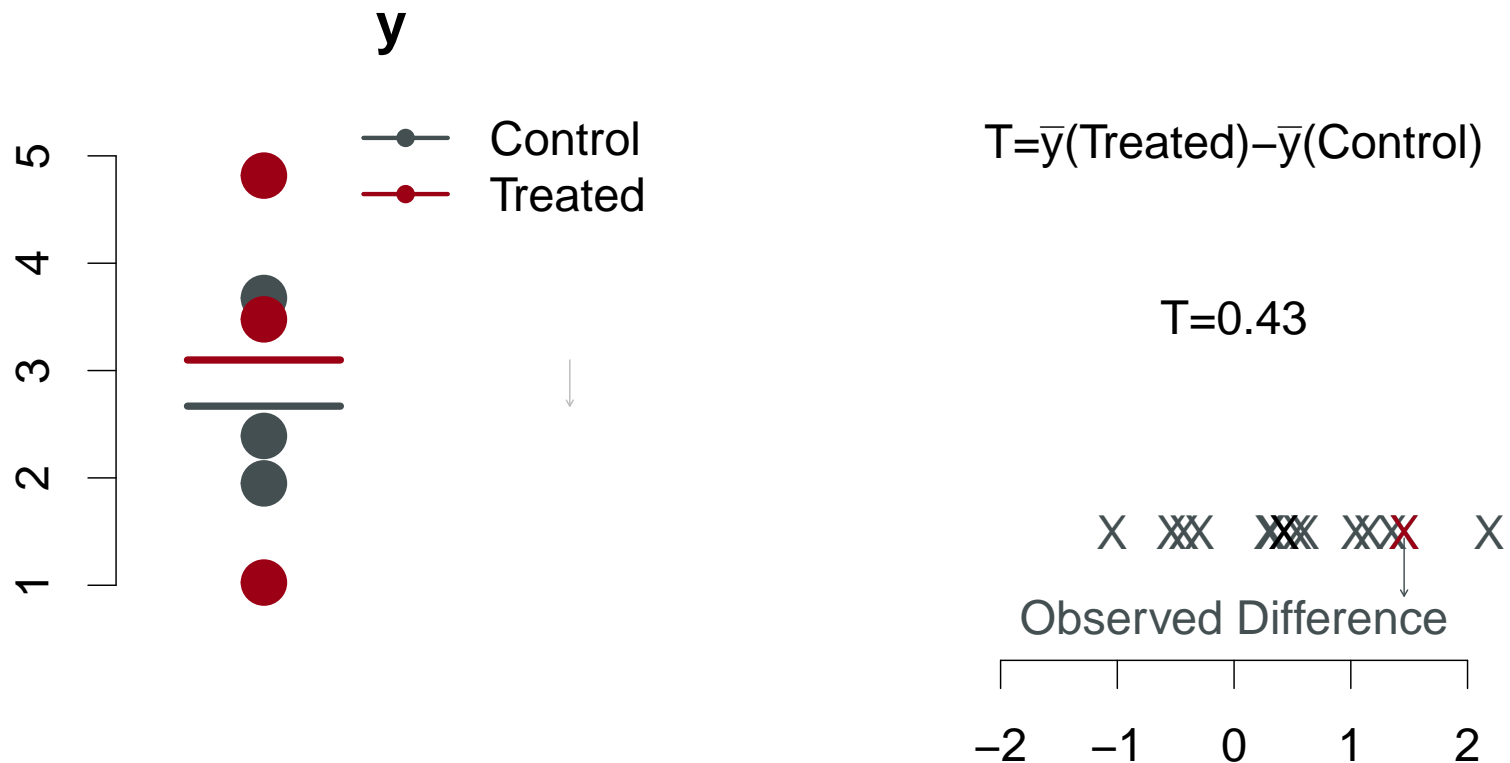
A Naive approach to Permutation Testing

...and go on with all hypothetical experiments...



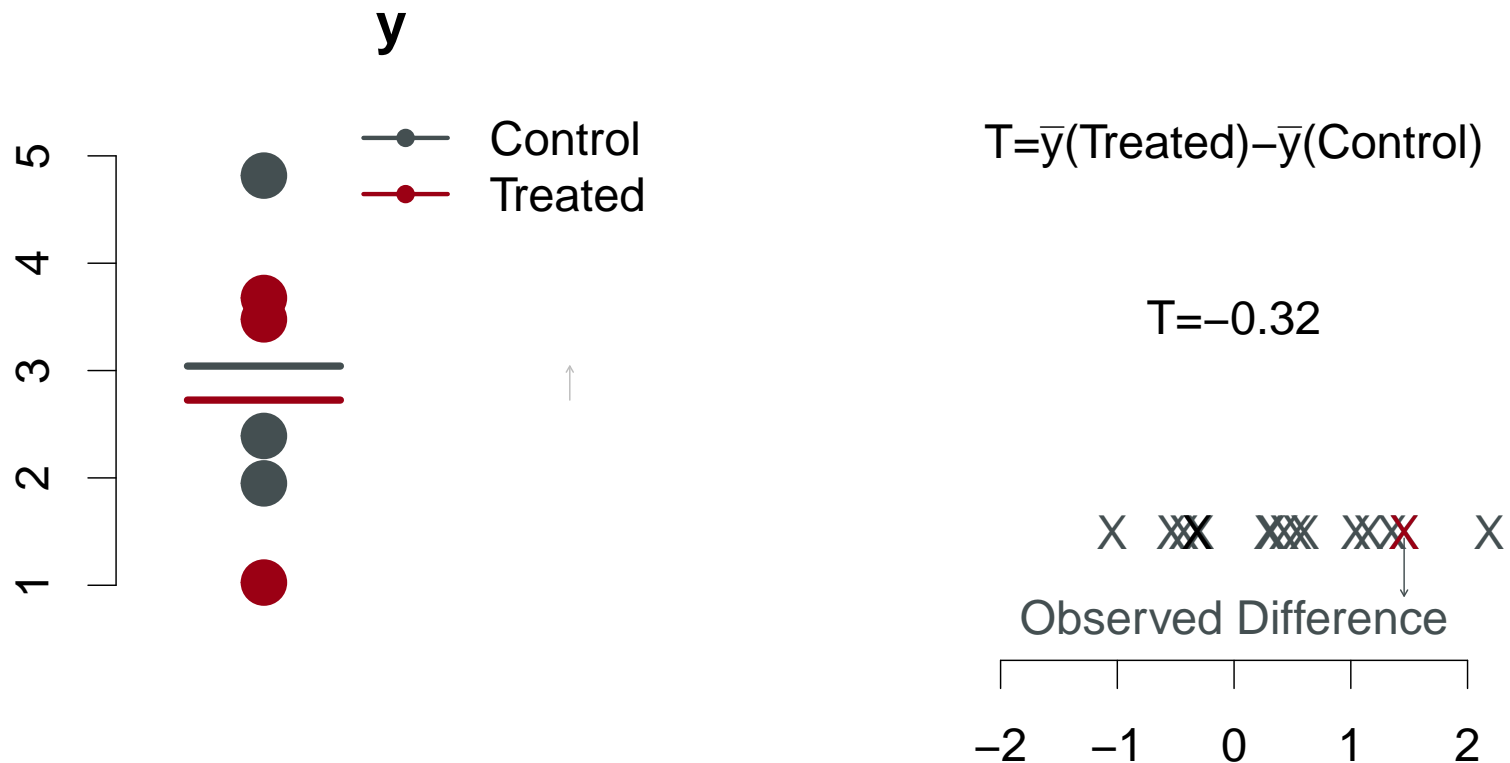
A Naive approach to Permutation Testing

...and go on with all hypothetical experiments...



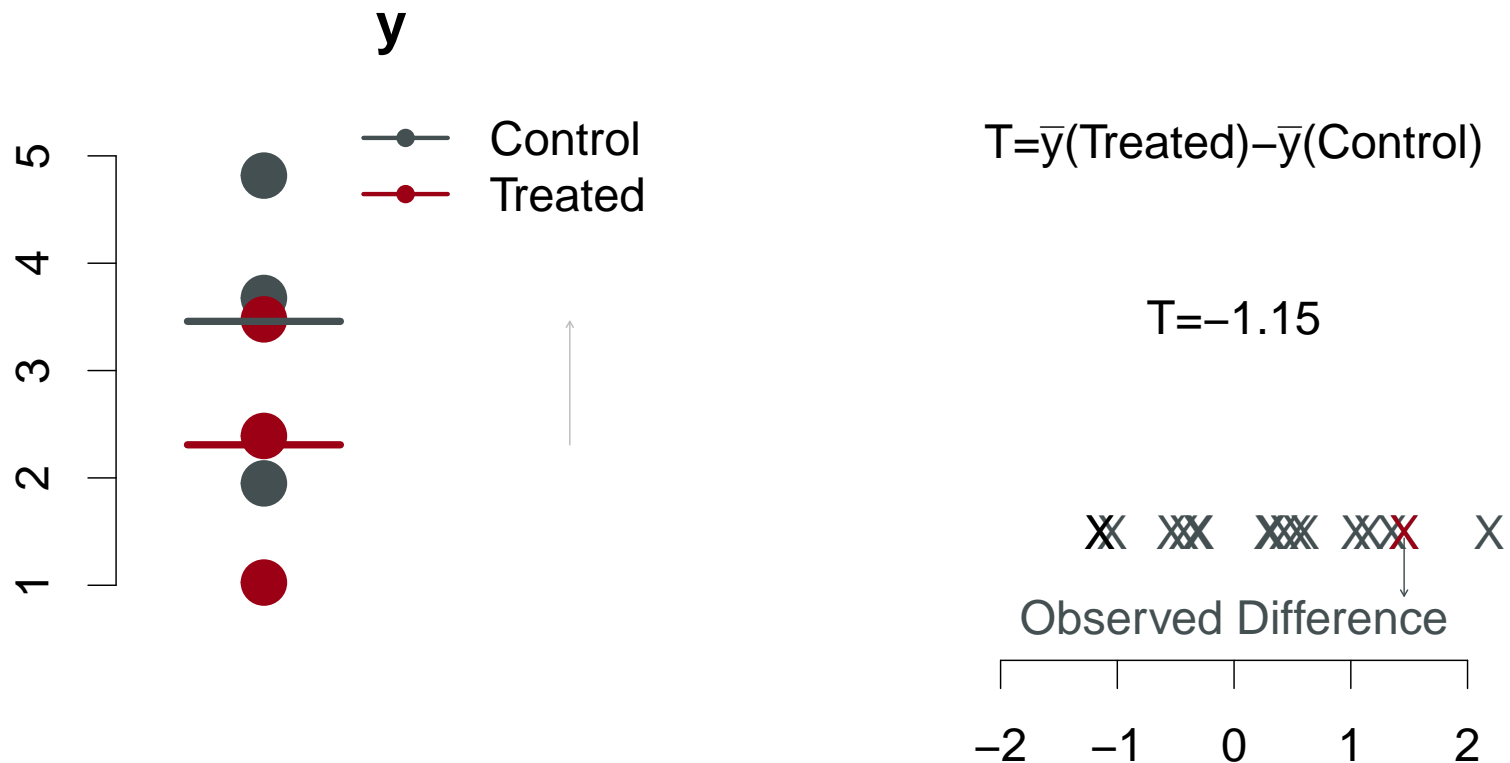
A Naive approach to Permutation Testing

...and go on with all hypothetical experiments...



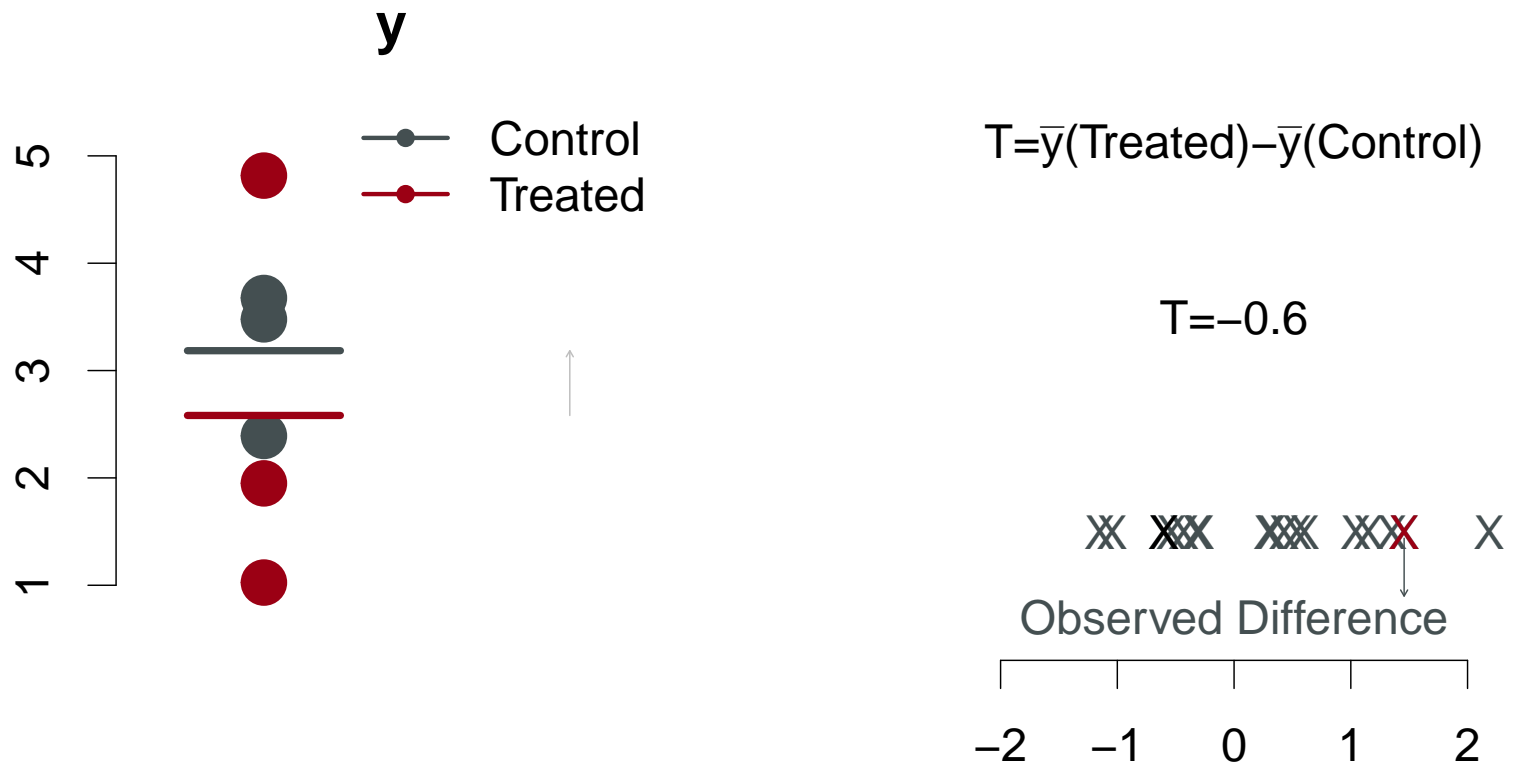
A Naive approach to Permutation Testing

...and go on with all hypothetical experiments...



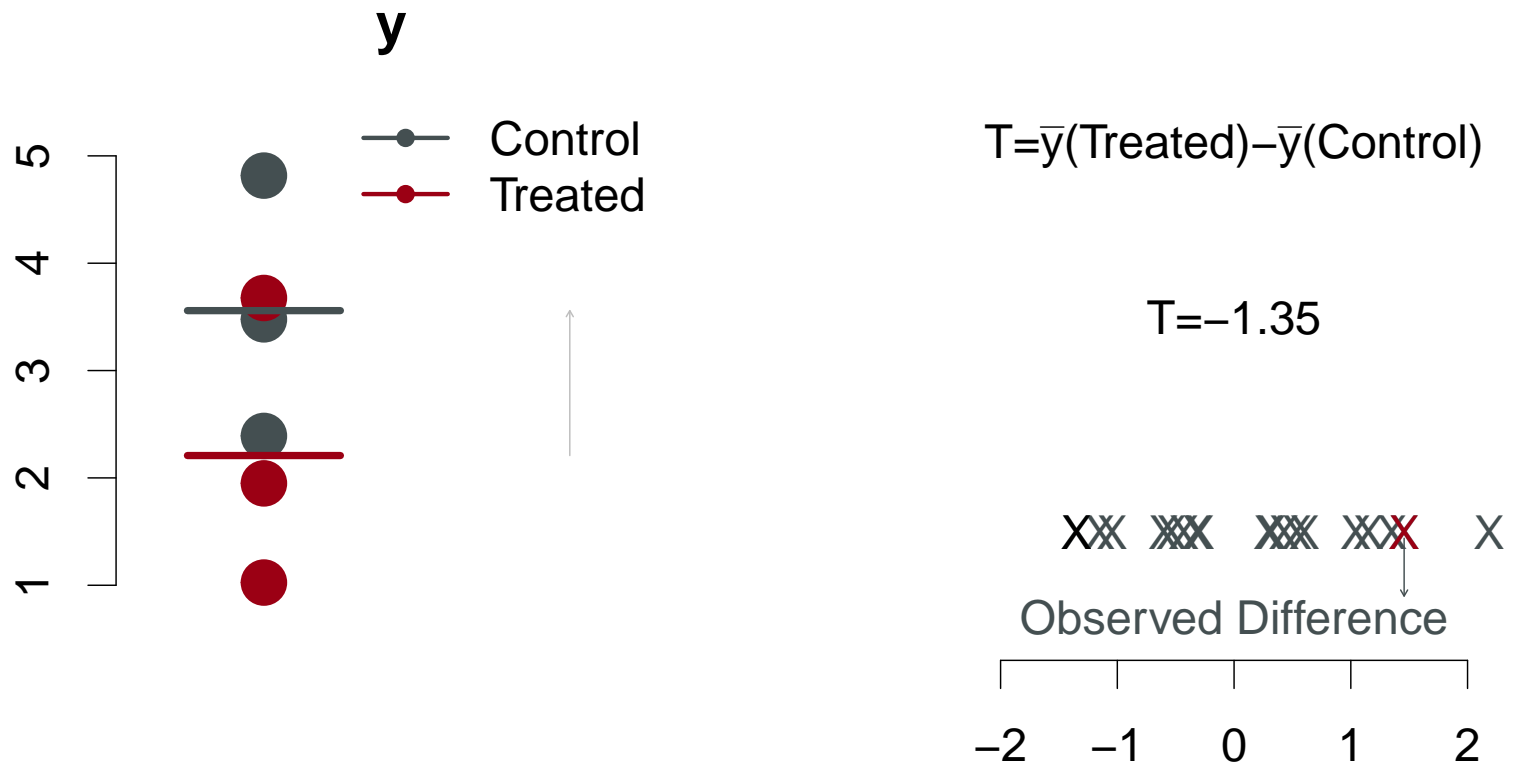
A Naive approach to Permutation Testing

...and go on with all hypothetical experiments...



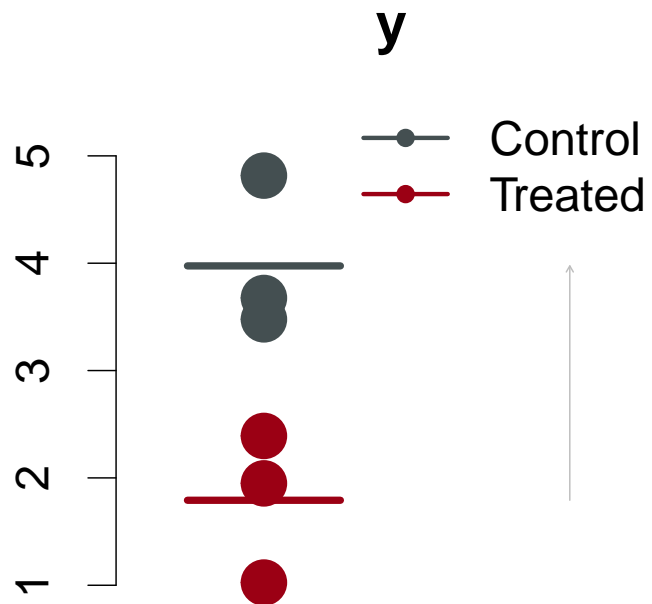
A Naive approach to Permutation Testing

...and go on with all hypothetical experiments...



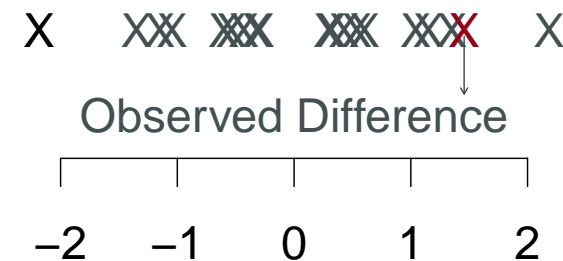
A Naive approach to Permutation Testing

...and go on with all hypothetical experiments...



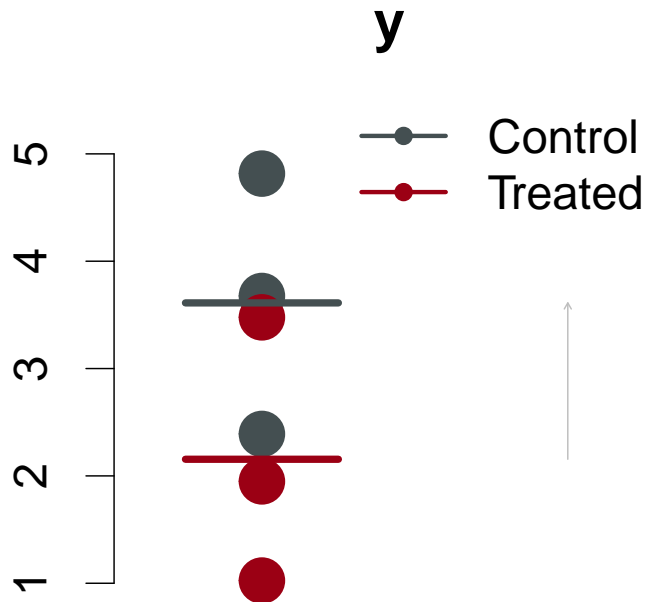
$$T = \bar{y}(\text{Treated}) - \bar{y}(\text{Control})$$

$$T = -2.18$$



A Naive approach to Permutation Testing

...and compute the p-value!



$$T = \bar{y}(\text{Treated}) - \bar{y}(\text{Control})$$
$$\text{p-value} = 2/20 = 0.10$$

$$T = -1.46$$



Summary

The Permutation Test:

- Conditioned to observed data (i.e. the distribution of the test statistic depends on the data).
- Under H_0 cases and controls have the same distribution (eg. they have the same probability to get high values),
- explore all possible experiments that we can observe with the data (ie. exchanging cases and controls),
- compute the p-value as the proportion of experiments providing equal or more evidence against H_0 with respect to observed data.

Outline

Introduction

A toy example

A Naive approach to Permutation Testing

Theory (very short)

Other cases

Multivariate Testing

A (slight) more formal approach

(see also Pesarin, 2001)

$\mathbf{y} = (y_1, y_2, \dots, y_n)$ the vector of observed data

Orbit: the set of all samples having the same likelihood under H_0 .

$$\mathcal{O} = \{\mathbf{y}^* : f(\mathbf{y}^*) = f(\mathbf{y})\}$$

(and $|\mathcal{O}|$ number of elements of \mathcal{O})

If we assume exchangeability of observations, then:

$$\mathcal{O} = \{\text{all permutations of the observed data } \mathbf{y}\} = \{\mathbf{y}^* : \pi^* \circ \mathbf{y}\}$$

($\pi^* \in \Pi$, Π set of all possible permutations)

A (slight) more formal approach

Remark: exchangeable observations: $f(y_1, y_2) = f(y_2, y_1)$.

It implies observations:

- are **identically distributed**
t-test and linear models assume normality, only asymptotic control of the type I error
- have the **same dependence**
t-test and linear models assume independence, which is just a special case, i.e. more stringent assumptions

A (slight) more formal approach

p-value: proportion of experiments providing equal or more evidence against H_0 with respect to observed data.

To compute it, we need the **Orbit** \mathcal{O} and a

A (slight) more formal approach

p-value: proportion of experiments providing equal or more evidence against H_0 with respect to observed data.

To compute it, we need the **Orbit** \mathcal{O} and a

Test statistic ($T: \mathbb{R}^n \rightarrow \mathbb{R}$) quantifies the evidence against H_0

- higher values provide more evidence against H_0
- compute a test statistic for each element of the Orbit \mathcal{O} , this induces an ordering on \mathcal{O} .

In our example: $T = \bar{y}(Treated) - \bar{y}(Control)$ is the difference in mean, higher the difference, higher the evidence for H_1 .

A (slight) more formal approach

$$f(\mathbf{y}^*|\mathcal{O}) = \frac{f(\mathbf{y}^* \cap \mathcal{O})}{f(\mathcal{O})} = \frac{f(\mathbf{y}^*)}{f(\mathcal{O})} = \frac{f(\mathbf{y}^*)}{f(\cup_{y \in \mathcal{O}} y)} = \frac{1}{|\mathcal{O}|} \quad \forall \mathbf{y}^* \in \mathcal{O}$$

i.e. each permutation is equally likely in the Orbit \mathcal{O} .

The **p-value**:

$$\begin{aligned} P(T(\mathbf{y}^*) \geq T(\mathbf{y}) | \mathbf{y}^* \in \mathcal{O}, H_0) &= \\ &= \int_{T(\mathbf{y})}^{+\infty} f(T(\mathbf{y}^*)) dT(\mathbf{y}^*) = \\ &= \sum_{\mathbf{y}^* \in \mathcal{O}} I(T(\mathbf{y}^*) \geq T(\mathbf{y})) / |\mathcal{O}| \quad \forall \mathcal{O} \end{aligned}$$

The package flip

It is on CRAN and on github (<https://github.com/livioivil/flip>)

To install the github version type (in R):

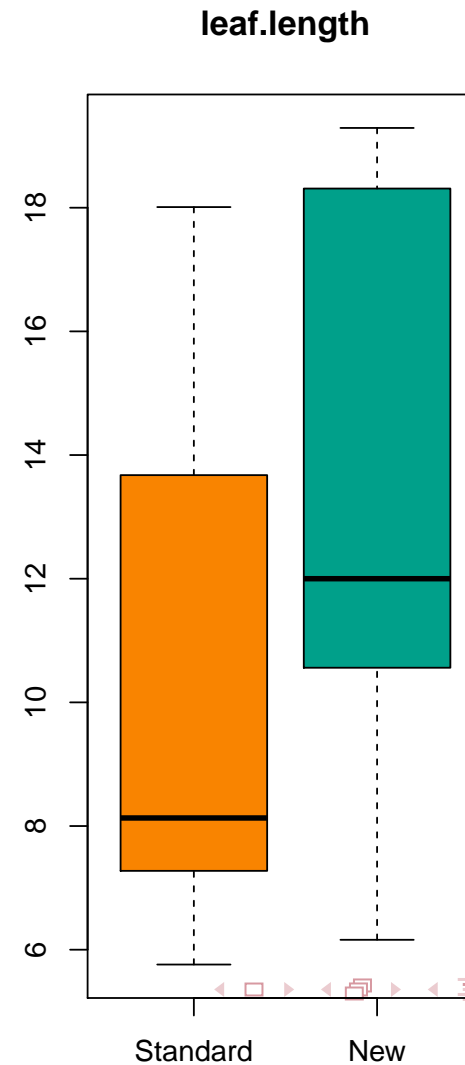
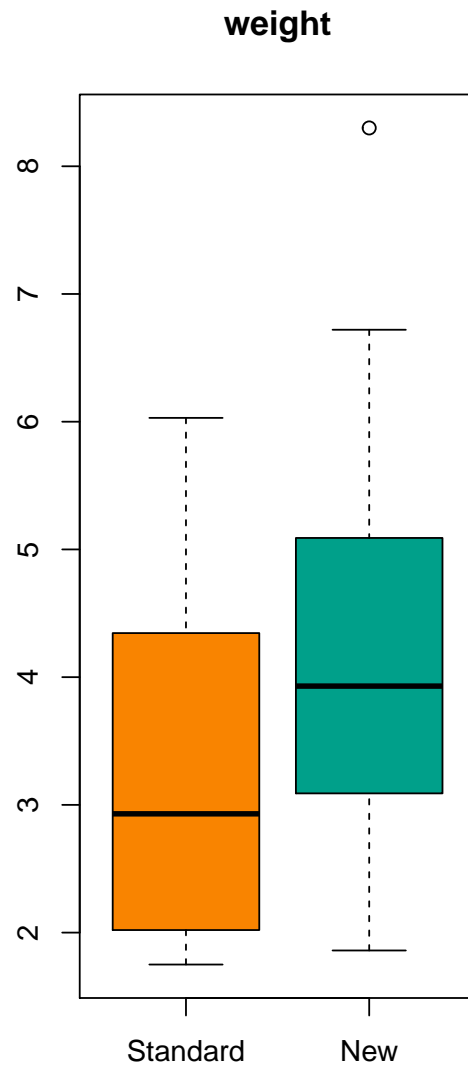
```
library(devtools)
```

```
install_github('livioivil/flip')
```

Seeds data (Pesarin, 2001)

Standard fertilizer ($grp = 0$) vs New fertilizer ($grp = 1$)

Total weight of the plant and average leaves length is recorded.



Hypothesis testing

About weight:

- $H_0(weight) : F(weight|grp = 0) = F(weight|grp = 1)$

vs

- $H_1(weight) : F(weight|grp = 0) > F(weight|grp = 1)$

And about leaf length:

- $H_0(leaf.len) : F(leaf.len|grp = 0) = F(leaf.len|grp = 1)$

vs

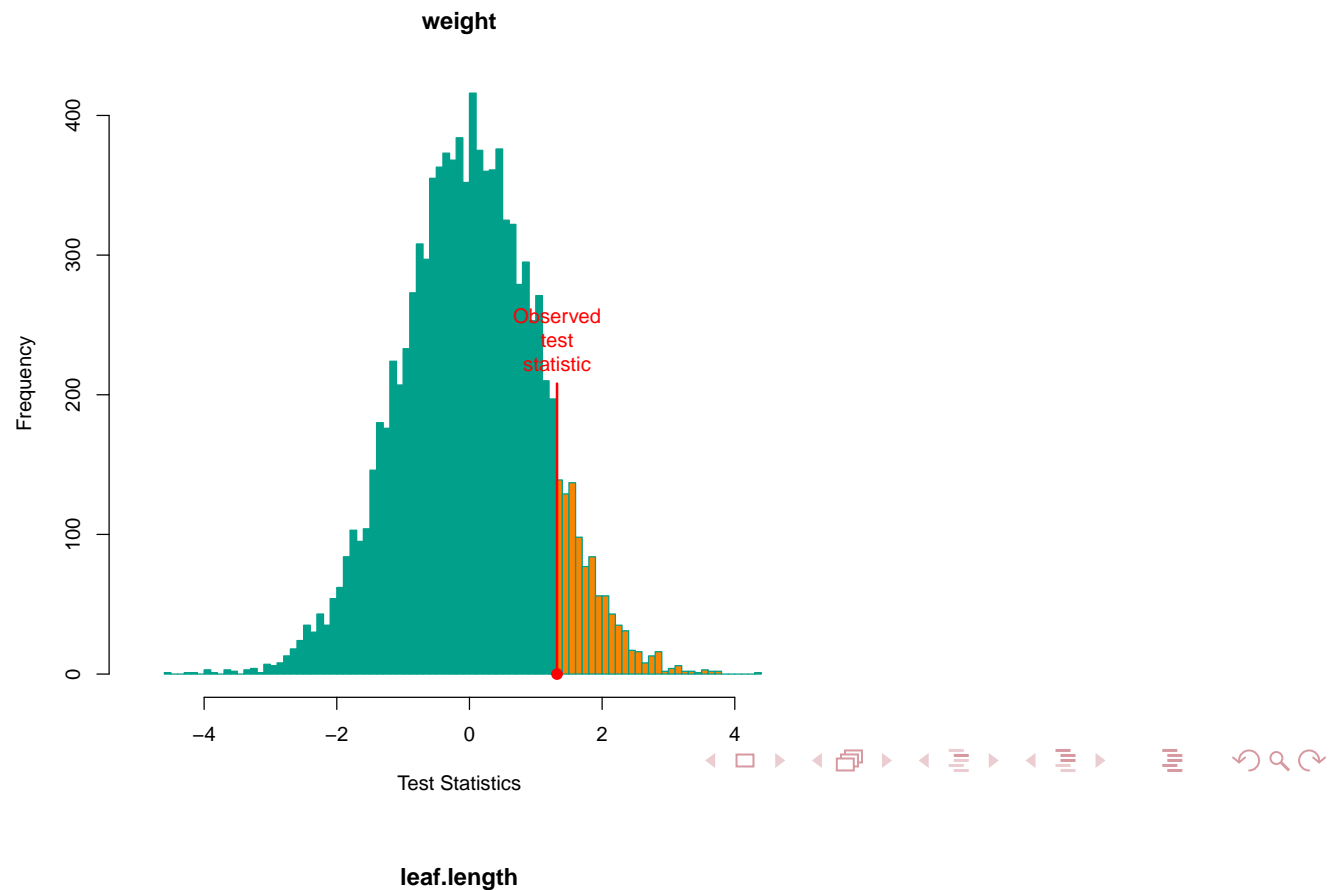
- $H_1(leaf.len) : F(leaf.len|grp = 0) > F(leaf.len|grp = 1)$

Hypothesis testing

```
res=flip(.~ grp, data=seeds, tail=1)
```

	Test	Stat	tail	p-value
weight	t	1.320	>	0.098
leaf.length	t	2.061	>	0.030

together with some visualization `hist(res)`



Two-tailed tests

About weight:

- $H_0(\text{weight}) : F(\text{weight}|\text{grp} = 0) = F(\text{weight}|\text{grp} = 1)$

vs

- $H_1(\text{weight}) : F(\text{weight}|\text{grp} = 0) \neq F(\text{weight}|\text{grp} = 1)$

And about leaf length:

- $H_0(\text{leaf.len}) : F(\text{leaf.len}|\text{grp} = 0) = F(\text{leaf.len}|\text{grp} = 1)$

vs

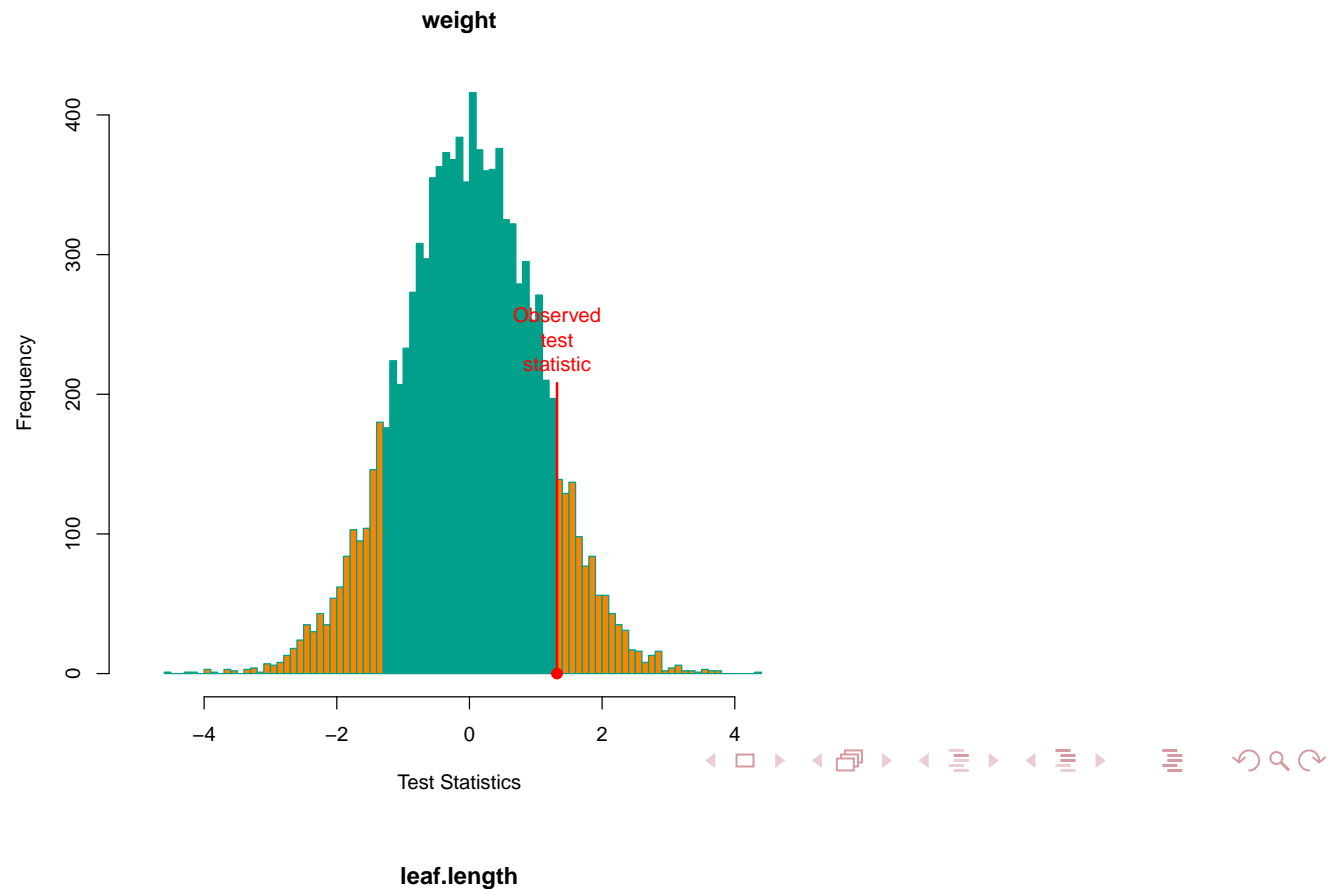
- $H_1(\text{leaf.len}) : F(\text{leaf.len}|\text{grp} = 0) \neq F(\text{leaf.len}|\text{grp} = 1)$

Two-tailed tests

```
res=flip(~ grp, data=seeds, tail=0)
```

	Test	Stat	tail	p-value
weight	t	1.320	><	0.202
leaf.length	t	2.061	><	0.049

Also very negative values provide evidence against H_0



Properties

- Exact control of the Type I Error:
 $P(p \leq \alpha | H_0) < \alpha \quad \forall \text{ (attainable)} \alpha$
- Consistency: $P(p \leq \alpha | H_1) \rightarrow 1$ when $n \rightarrow \infty$
- Converges to parametric counterpart (i.e. asymptotic optimality if the parametric test is optimal)

Remark

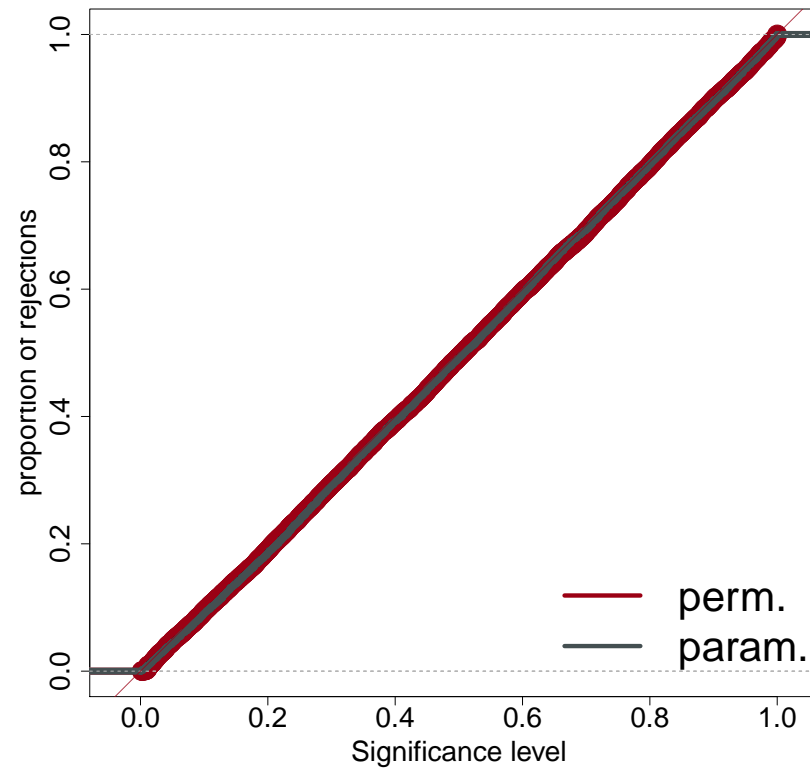
The number of possible permutations (size of the Orbit $|\mathcal{O}|$) is often huge, we can not compute the test statistic for all elements. Common use to sample from the Orbit (i.e. randomly permute the labels B times).

The properties remain the same.

Simulation: normal distribution

- Comparison of Two groups (labels A, B) of size 5
- $y_i \sim N(0, 1)$
 - $H_0 : f(y|grp = A) = f(y|grp = B)$
 - $H_0 : f(y|grp = A) \neq f(y|grp = B)$
(i.e. two-sided alternatives)
- 10000 replications
- 1000 random permutations for each test

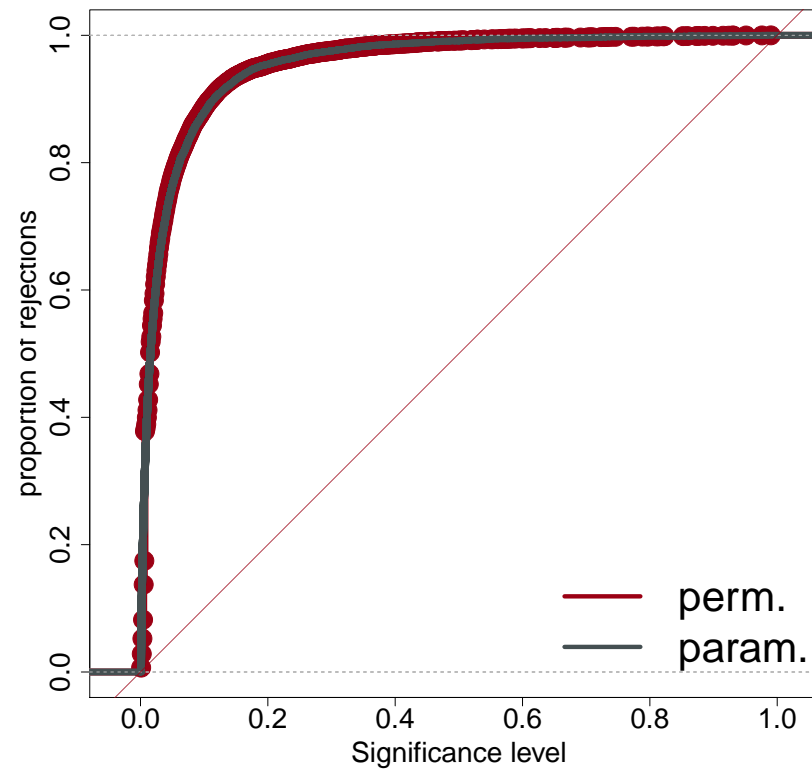
Simulation: H_0



Empirical Type I error	≤0.01	≤0.05	≤0.1	≤0.5	≤0.75
Permutation	0.00	0.05	0.09	0.49	0.74
Parametric t.test	0.01	0.04	0.09	0.49	0.75

Simulation: H_1

now $(y|grp = A) \sim N(0, 1)$, $(y|grp = B) \sim N(2, 1)$

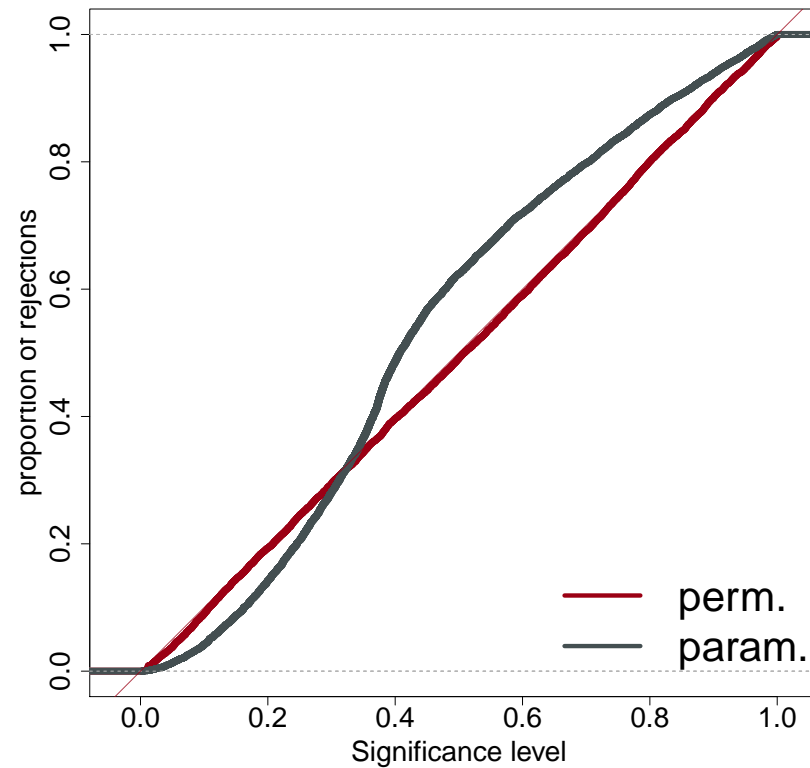


Empirical Power	≤ 0.01	≤ 0.05	≤ 0.1	≤ 0.5	≤ 0.75
Permutation	0.40	0.78	0.89	0.99	1.00
Parametric t.test	0.41	0.77	0.88	0.99	1.00

Simulation: Cauchy distribution

- Comparison of Two groups (labels A, B)
- $y_i \sim \text{Cauchy}$
 - $H_0 : f(y|grp = A) = f(y|grp = B)$
 - $H_0 : f(y|grp = A) \neq f(y|grp = B)$
(i.e. two-sided alternatives)
- 10000 replications
- 1000 random permutations for each test

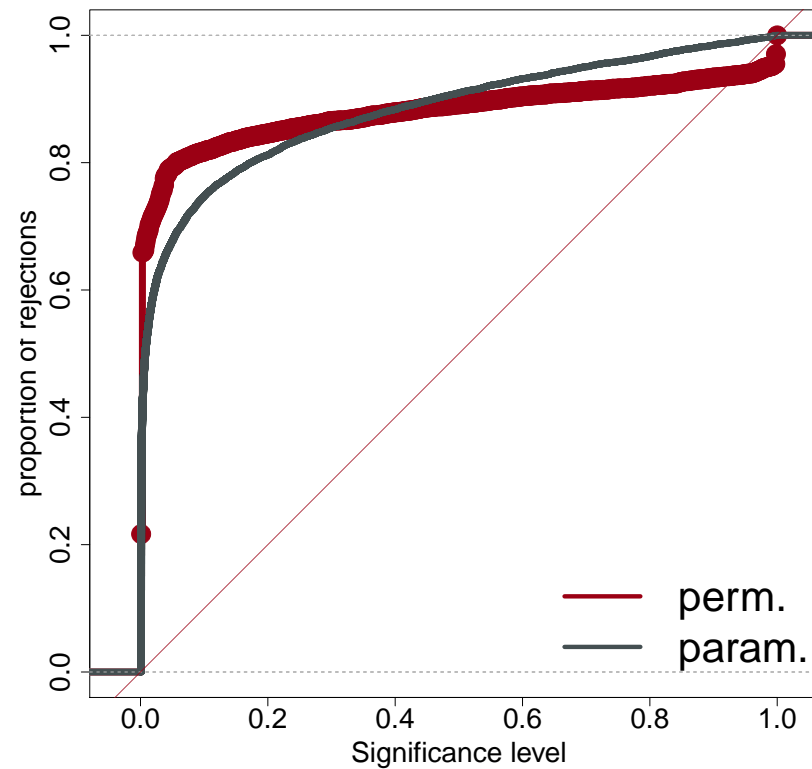
Simulation: H_0



Empirical Type I error	≤ 0.01	≤ 0.05	≤ 0.1	≤ 0.5	≤ 0.75
Permutation	0.00	0.04	0.09	0.49	0.74
Parametric t.test	0.00	0.01	0.04	0.62	0.84

Simulation: H_1

now $(y|grp = A) \sim Cauchy(0)$, $(y|grp = B) \sim Cauchy(10)$



Empirical Power	≤0.01	≤0.05	≤0.1	≤0.5	≤0.75
Permutation	0.66	0.78	0.81	0.89	0.92
Parametric t.test	0.53	0.68	0.75	0.91	0.96

Outline

Introduction

A toy example

A Naive approach to Permutation Testing

Theory (very short)

Other cases

Multivariate Testing

A very general approach

This approach (Orbit \mathcal{O} + Test statistic T) is very general.
It includes:

- ChiSquare test
- Fisher exact test
- McNemar test
- rank tests
- ANOVA tests
- linear models
- other models difficult to deal within the parametric framework
- ...

The case of contingency table

$x \setminus y$	0	1	2
A	4	3	0
B	1	2	4

```
> chisq.test(x,y)
```

Pearson's Chi-squared test

data: x and y

X-squared = 6, df = 2, p-value = 0.04979

Warning message:

In chisq.test(x, y) : Chi-squared approximation may
be incorrect

The case of contingency table

Use: `simulate.p.value = TRUE`

```
> chisq.test(x,y,simulate.p.value = TRUE)
```

Pearson's Chi-squared test (based on 2000 replicates)

data: x and y

X-squared = 6, df = NA, p-value = 0.09345

Same schema:

- Orbit \mathcal{O} : all possible permutation of \mathbf{y} (under H_0 \mathbf{x} and \mathbf{y} are independent)
- test statistic $T(\mathbf{y}^*)$: the χ^2 statistic computed (higher is better)
- p-value: proportion of T^* greater than the one computed on observed data \mathbf{y} : $p = \sum_{\mathbf{y}^* \in \mathcal{O}} I(T(\mathbf{y}^*) \geq T(\mathbf{y})) / |\mathcal{O}|$.

The case of experimental design with blocks (or within-subject)

For example: n lots/subjects, each with two treatments (A vs B)
we can assume a specific effect for each lot/subject:

$$y_{ij} \sim (\nu_i + \mu_j, \sigma_i) \quad i = 1, \dots, n, \quad j = A, B$$

$$H_0 : \mu_A = \mu_B \Leftrightarrow \mu_A - \mu_B = 0$$

Parametric approach:

- define: $z_i = y_{iB} - y_{iA} \sim (\mu_B - \mu_A, 2 \cdot \sigma_i) \quad i = 1, \dots, n$
- assume $\sigma_i = \sigma, \quad \forall i = 1, \dots, n$
- perform a 1-sample t-test (i.e. t-test for 2 paired samples)
- test is exact only if z_i is normal, it is approximated otherwise.

The case of experimental design with blocks (or within-subject)

What about permutation approach? how to define the Orbit \mathcal{O} ?

$$H_0 \Rightarrow \mu_A = \mu_B = \mu \Rightarrow y_{ij} \sim (\nu_i + \mu, \sigma_i) \quad i = 1, \dots, n, \quad j = A, B$$

$$\begin{aligned} & f(y_{1A}, \textcolor{red}{y}_{1B}, y_{2A}, \textcolor{red}{y}_{2B}, \dots, y_{nA}, \textcolor{red}{y}_{nB}) (\text{observed}) = \\ &= f(\textcolor{red}{y}_{1A}, y_{1B}, y_{2A}, \textcolor{red}{y}_{2B}, \dots, \textcolor{red}{y}_{nA}, y_{nB}) = \\ &= f(y_{1A}, \textcolor{red}{y}_{1B}, \textcolor{red}{y}_{2A}, y_{2B}, \dots, \textcolor{red}{y}_{nA}, y_{nB}) \dots \end{aligned}$$

- We exchange observations only within the same lot/subject!
- There are 2^n possible configurations: $|\mathcal{O}| = 2^n$
- we don't need to assume: $\sigma_i = \sigma, \quad \forall i = 1, \dots, n$
- (even we may allow non additive effect: $\nu_{ij} \neq \nu_i + \mu_j$)
- A part from \mathcal{O} , the procedure is the same.

Outline

Introduction

A toy example

A Naive approach to Permutation Testing

Theory (very short)

Other cases

Multivariate Testing

Multivariate hypotheses

Testing $H_0(\text{weight}) + \text{Testing } H_0(\text{leaf.length})$
is different from testing

- $H_0 : H_0(\text{weight}) \cap H_0(\text{leaf.length})$
(i.e. simultaneously true)

vs

- $H_1 : H_1(\text{weight}) \cup H_1(\text{leaf.length})$
(i.e. at least one null hypo is false)

Here test H_0 : the New is equal to the Standard in both variables
For 1-tailed alternative:

$p_{\text{weight}} = .098$ and $p_{\text{leaf.length}} = .030$.

Shall we reject the multivariate H_0 ?

Multivariate hypotheses

Common approaches:

- MANOVA test

- is OK but only for linear models (2 or more samples).
 - assumes multivariate normality.

- Does not allow for one-sided alternatives.

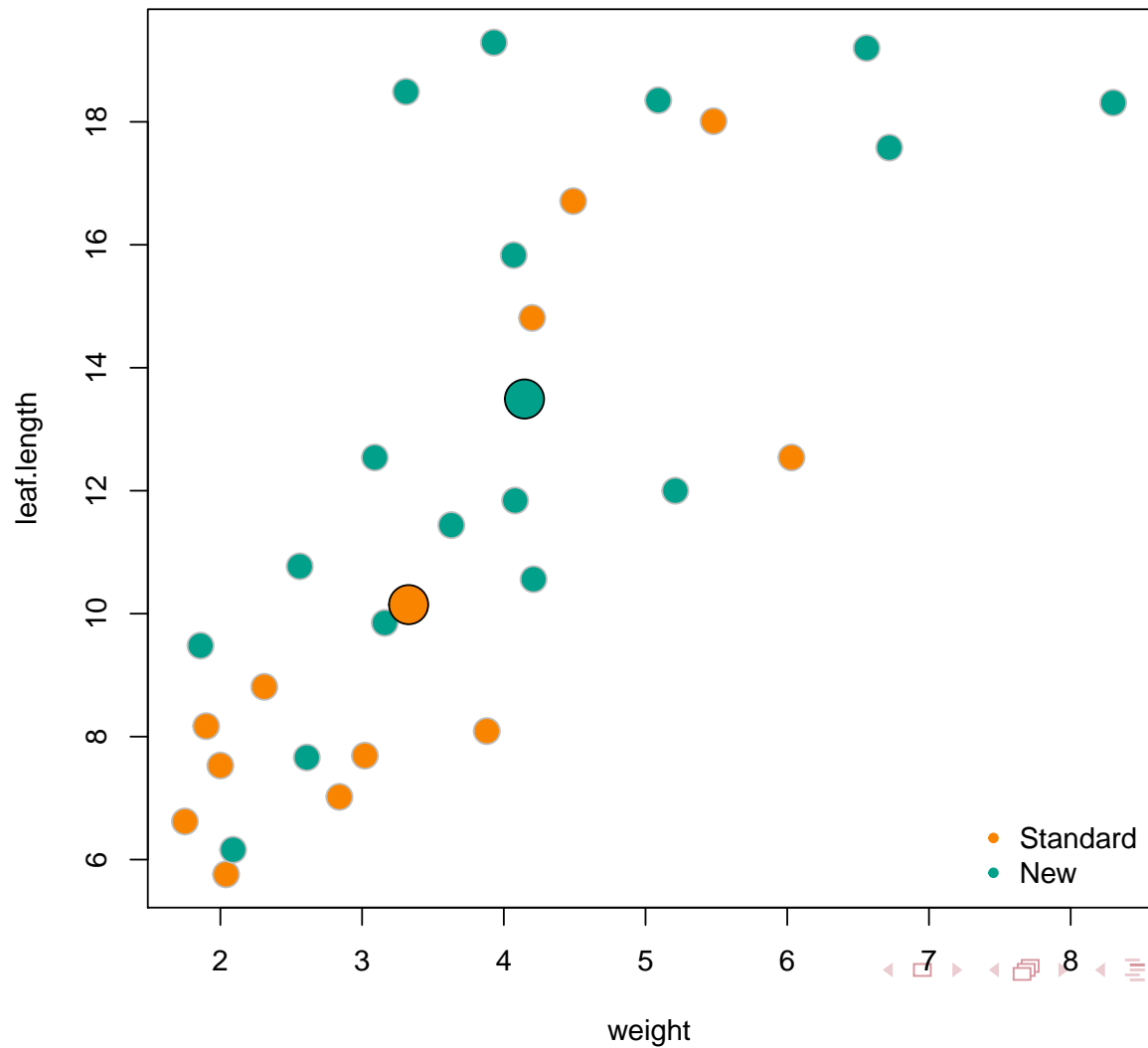
- Bonferroni correction ($p = \min(p_1, p_2) * 2$)

- very simple, always valid

- does not take in account dependences among data
(i.e. may be very conservative, i.e. high final p-value)

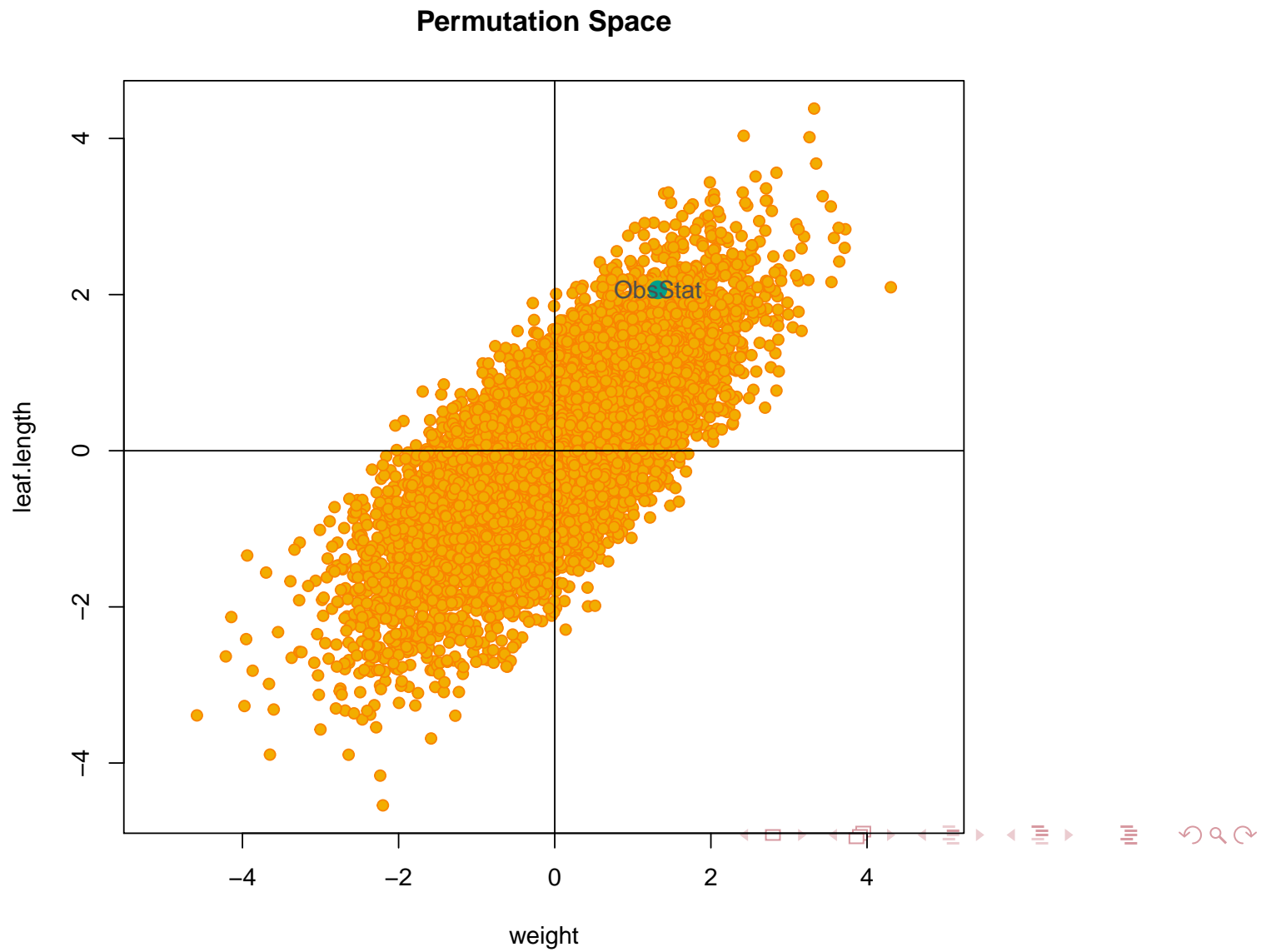
Joint distributions of the data

The two variables are dependent:



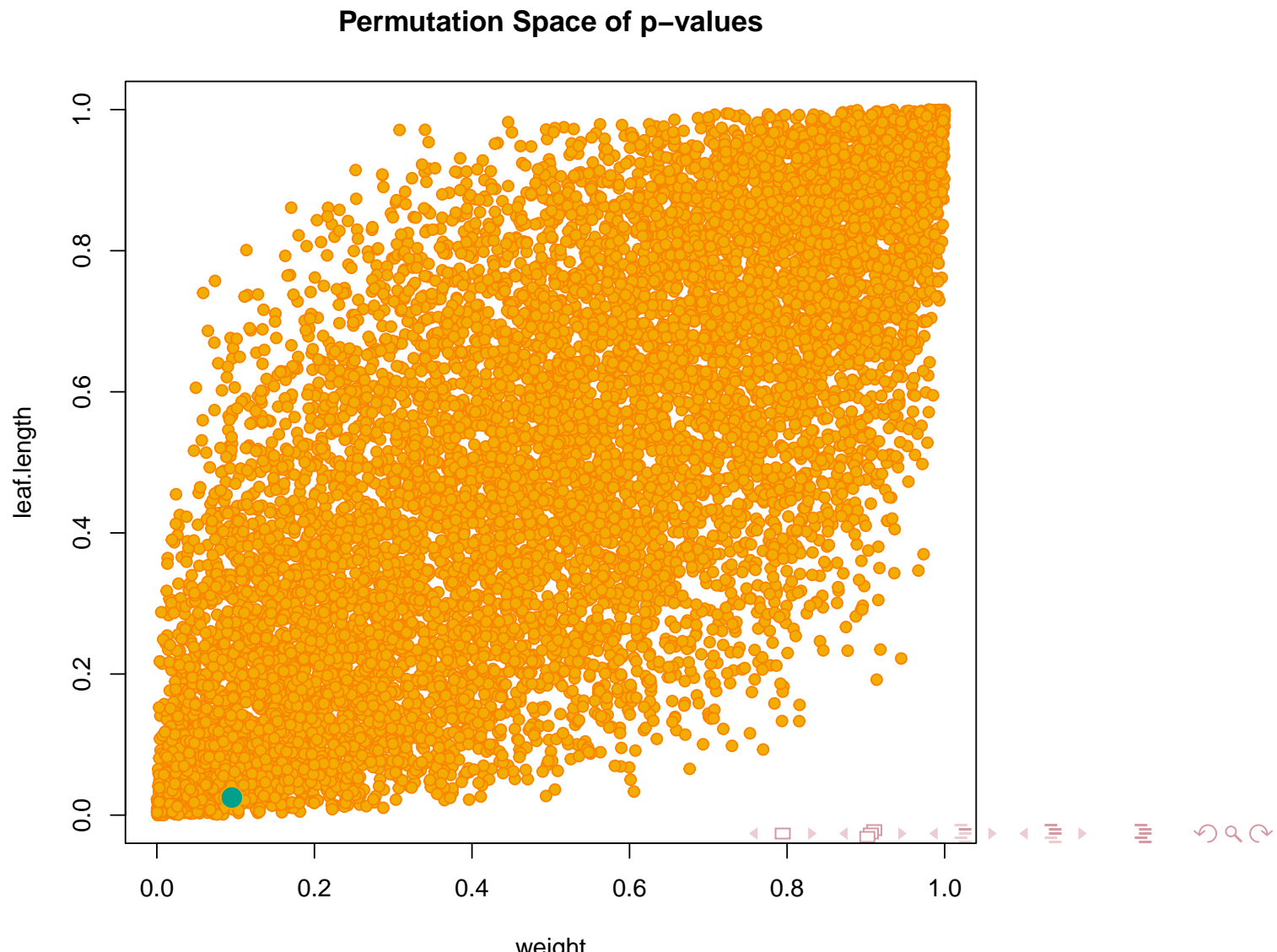
Joint distributions of test statistics

This dependence induce a dependence into the joint distribution of the test statistic



Joint distributions of the p-values

... and into the p-values joint distributions (i.e. compute the p-values for observed samples and all elements $\mathbf{y}^* \in \mathcal{O}$)



Nonparametric Combination methodology (Pesarin, 2001)

The Orbit is now defined in a multivariate framework:

$$\mathcal{O} = \{(\mathbf{y}_1, \mathbf{y}_2)^* : f((\mathbf{y}_1, \mathbf{y}_2)^*) = f((\mathbf{y}_1, \mathbf{y}_2))\}$$

(in practice: when you permute one observation in a variable, do the same in the other variables)

Also compute the p_1^*, p_2^* associated to each $\mathbf{y}_1^*, \mathbf{y}_2^*$.

Nonparametric Combination methodology (Pesarin, 2001)

The Orbit is now defined in a multivariate framework:

$$\mathcal{O} = \{(\mathbf{y}_1, \mathbf{y}_2)^* : f((\mathbf{y}_1, \mathbf{y}_2)^*) = f((\mathbf{y}_1, \mathbf{y}_2))\}$$

(in practice: when you permute one observation in a variable, do the same in the other variables)

Also compute the p_1^*, p_2^* associated to each $\mathbf{y}_1^*, \mathbf{y}_2^*$.

Define a Combining Function $\psi(p_1, \dots, p_m)$ having the following properties:

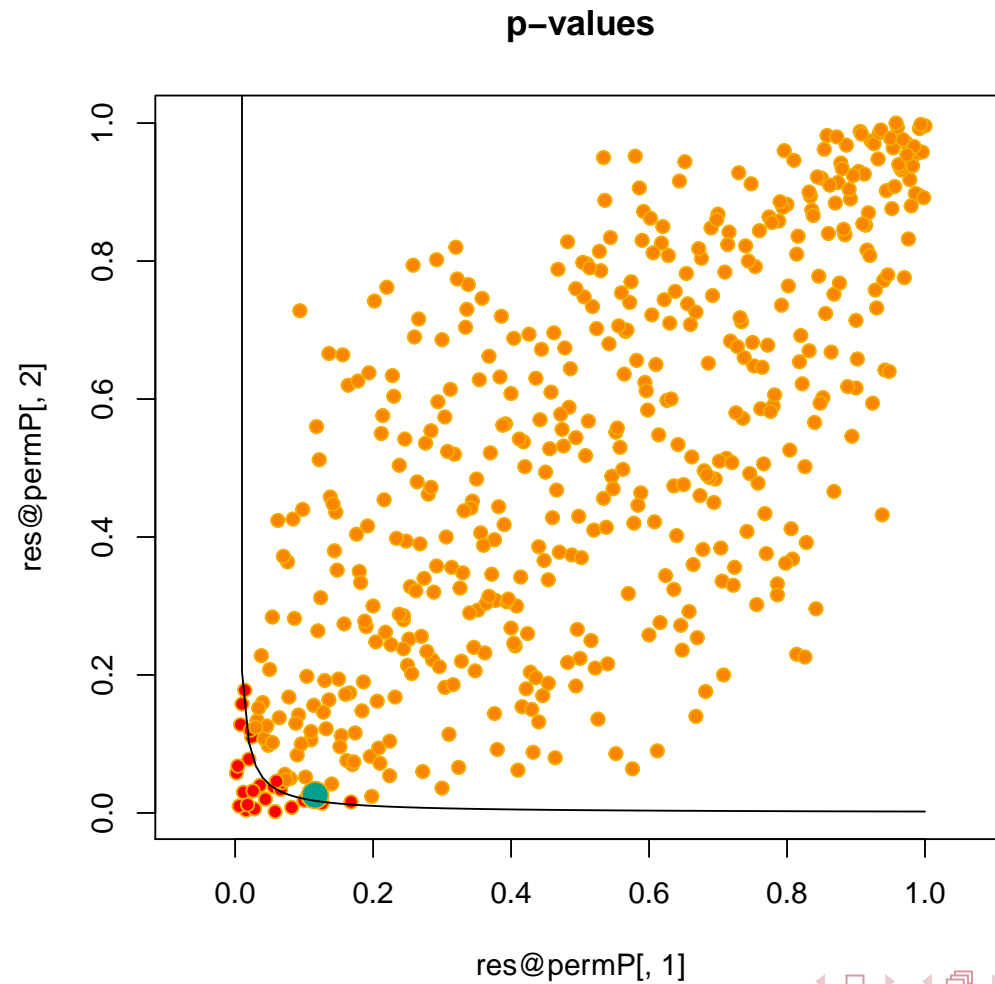
- i is non-increasing in each argument: $p_k < p'_k$ implies $\psi(\dots, p_k, \dots) \geq \psi(\dots, p'_k, \dots)$;
- ii attains its supremum ψ° if at least one argument attains 0;
- iii $\alpha > 0$ implies the critical value is such that $T_{\psi\alpha} < \psi^\circ$, i.e. no concentration of points at ψ° under H_0 .

Apply $T((p_1, p_2)^*) = \psi((p_1, p_2)^*)$ and compute the p_{global} .

Fisher combining function

$$\psi = -2 \cdot (\log(p_1) + \log(p_2)) = 6.029$$

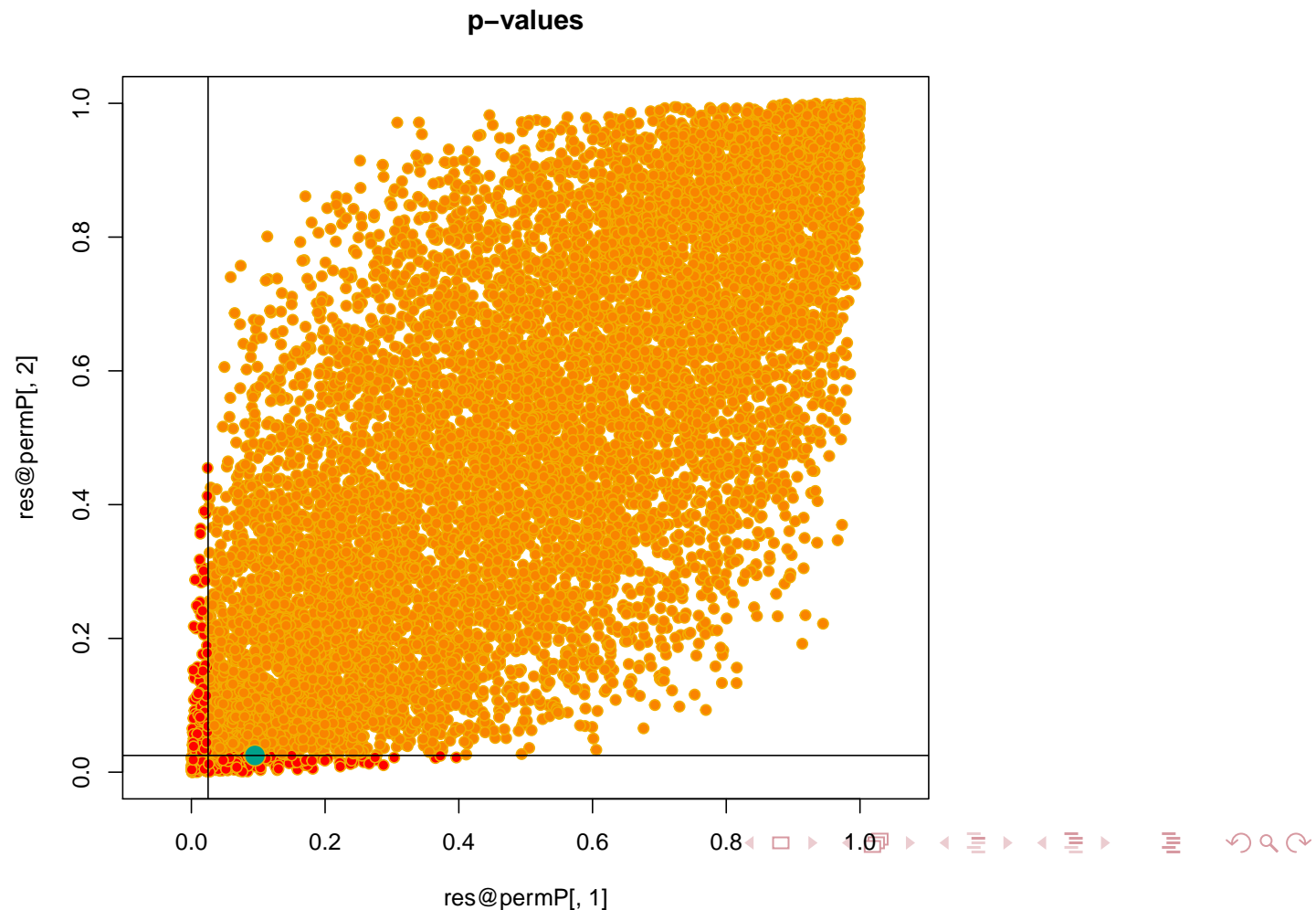
$$p_{global} = 0.0400$$



Tippett (min-p) combining function

$$\psi = 1 - \min(p_1, p_2) = 1 - 0.0280$$

$p_{global} = 0.0460$ ($\leq 2 \cdot 0.0280 = 0.0560$ similar to Bonferroni, but more powerfull)



Multiple testing

- Here we only have global p-values (in analogy to Manova, F-tests etc).

Multiple testing

- Here we only have global p-values (in analogy to Manova, F-tests etc).
- If the test is significant at least one among $H_0(weight)$ or $H_0(leaf.len)$ is false.
- We don't know which one! (weak control of the FamilyWise Error)

Multiple testing

- Here we only have global p-values (in analogy to Manova, F-tests etc).
- If the test is significant at least one among $H_0(weight)$ or $H_0(leaf.len)$ is false.
- We don't know which one! (weak control of the FamilyWise Error)
- How to cast them into multiple testing procedures?

Multiple testing

- Here we only have global p-values (in analogy to Manova, F-tests etc).
- If the test is significant at least one among $H_0(weight)$ or $H_0(leaf.len)$ is false.
- We don't know which one! (weak control of the FamilyWise Error)
- How to cast them into multiple testing procedures?
- (strong) control of the FWER is easy!

Closed Testing¹

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

Closure Set

ABC

AB

AC

BC

A

B

C

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

¹R Marcus, E Peritz, KR Gabriel (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 63: 655-660.

Westfall & Young min-p (and max-t) ²

Close testing infeasible when with many hypos ($2^m - 1$ tests).

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

²Resampling-based multiple testing: Examples and methods for p-value adjustment, volume 279. Wiley-Interscience, 1993.

Westfall & Young min-p (and max-t) ²

Close testing infeasible when with many hypos ($2^m - 1$ tests).

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 $t_{ABC} \geq t_\alpha$

²Resampling-based multiple testing: Examples and methods for p-value adjustment, volume 279. Wiley-Interscience, 1993.

Take Home Message

Permutation approach:

- very general uni/multi-variate approach
- few assumptions on the data-generating process
- natural approach in **randomized experimental design**
- good inferential properties
(in most of the cases: exact control of the type I error, consistency, asymptotic optimality)
- very convenient for multiplicity control methods, since it deals easily with dependent tests.

Warnings:

- more complex experimental design can be dealt, but with caution (Pesarin, 2001)
- multiple (generalized) linear models need some care
(Solari, Finos & Goeman, 2014 and other work in progress)

Software 'R': libraries 'coin', 'flip': `'flip()'`; `flip.adjust()`