

# Permutation-based multiple testing with R - theory 1/2

Livio Finos

`livio.finos@unipd.it`

University of Padua

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



# Outline

- 1 Introduction
- 2 A toy example
- 3 A Naive approach to Permutation Testing
- 4 Theory (very short)
- 5 Other cases
- 6 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

- 1 Introduction
- 2 A toy example**
- 3 A Naive approach to Permutation Testing
- 4 Theory (very short)
- 5 Other cases
- 6 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

- 1 Introduction
- 2 A toy example
- 3 A Naive approach to Permutation Testing**
- 4 Theory (very short)
- 5 Other cases
- 6 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:**

$$f(1.025, 1.949, 3.477, \mathbf{2.391}, \mathbf{3.676}, \mathbf{4.816}) =$$



# 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:**

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



# 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:**

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



# 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:**

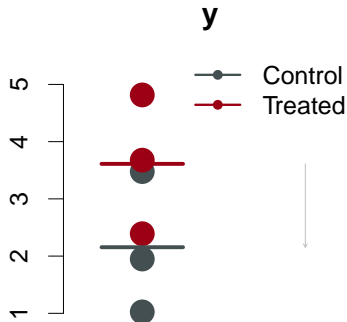
$$\begin{aligned} & f(1.025, 1.949, 3.477, \mathbf{2.391}, \mathbf{3.676}, \mathbf{4.816}) = \\ = & f(1.025, \mathbf{1.949}, 3.477, \mathbf{2.391}, 3.676, \mathbf{4.816}) = \\ = & f(\mathbf{1.025}, \mathbf{1.949}, 3.477, 2.391, \mathbf{3.676}, 4.816) = \\ = & \dots \end{aligned}$$

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



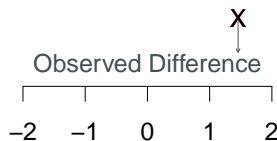
# A Naive approach to Permutation Testing

Compute the difference in mean of the two samples



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

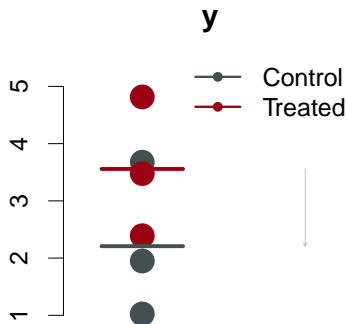
$$T = 1.46$$





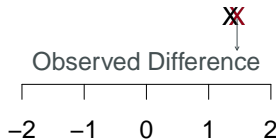
# A Naive approach to Permutation Testing

Compute the same difference on another hypothetical experiment



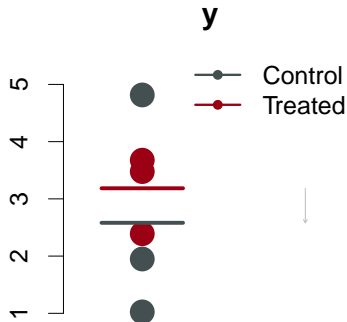
$$T = \bar{y}(\text{Treated}) - \bar{y}(\text{Control})$$

$$T = 1.35$$



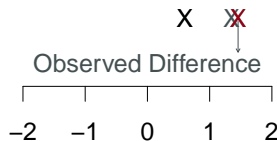
# A Naive approach to Permutation Testing

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



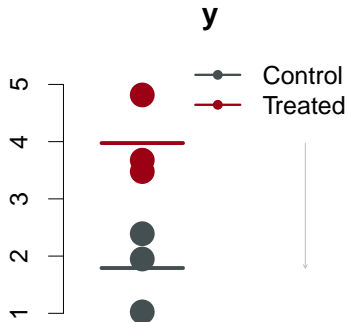
$$T = \bar{y}(\text{Treated}) - \bar{y}(\text{Control})$$

$$T = 0.6$$



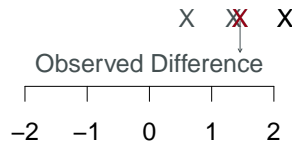
# A Naive approach to Permutation Testing

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



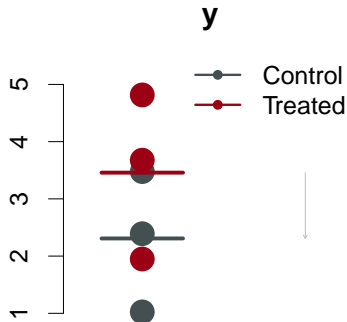
$$T = \bar{y}(\text{Treated}) - \bar{y}(\text{Control})$$

$$T = 2.18$$



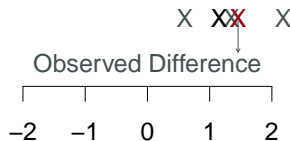
# A Naive approach to Permutation Testing

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



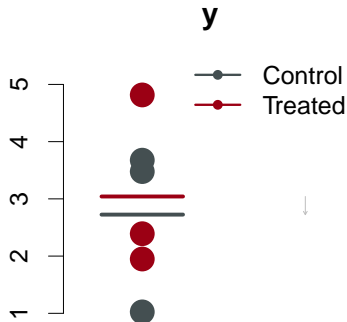
$$T = \bar{y}(\text{Treated}) - \bar{y}(\text{Control})$$

$$T = 1.15$$



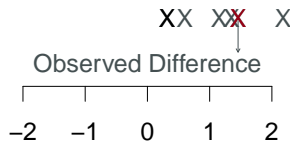
# A Naive approach to Permutation Testing

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



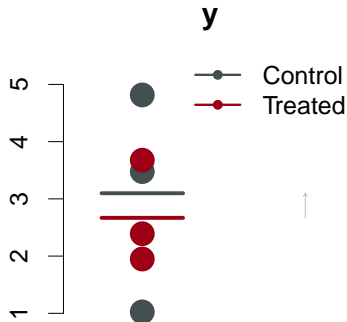
$$T = \bar{y}(\text{Treated}) - \bar{y}(\text{Control})$$

$$T = 0.32$$



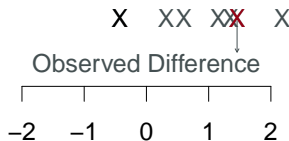
# A Naive approach to Permutation Testing

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



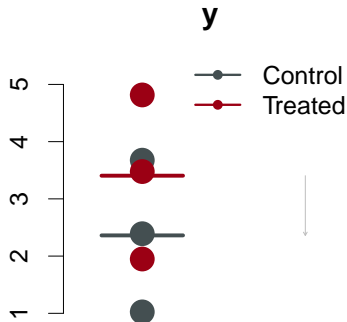
$$T = \bar{y}(\text{Treated}) - \bar{y}(\text{Control})$$

$$T = -0.43$$



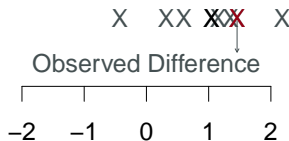
# A Naive approach to Permutation Testing

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



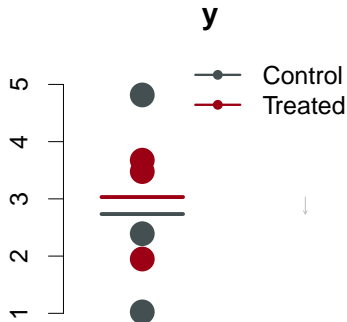
$$T = \bar{y}(\text{Treated}) - \bar{y}(\text{Control})$$

$$T = 1.04$$



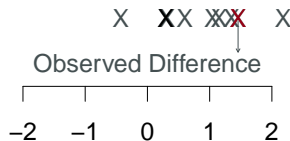
# A Naive approach to Permutation Testing

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



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

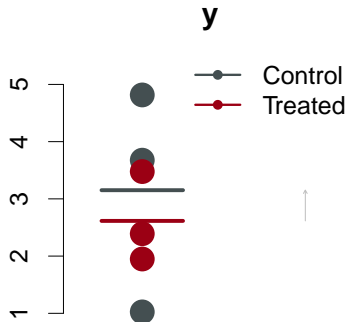
$$T = 0.3$$





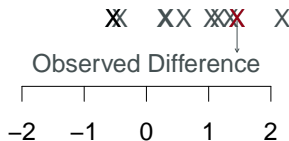
# A Naive approach to Permutation Testing

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



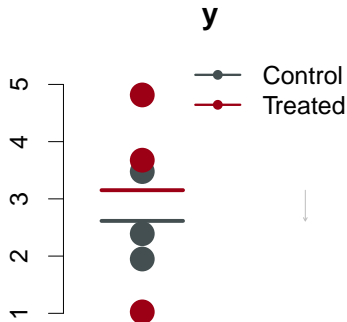
$$T = \bar{y}(\text{Treated}) - \bar{y}(\text{Control})$$

$$T = -0.54$$



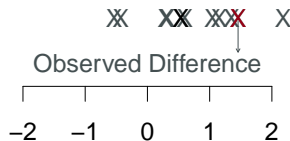
# A Naive approach to Permutation Testing

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



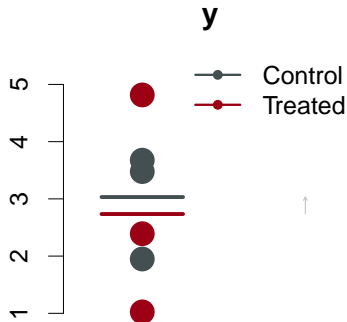
$$T = \bar{y}(\text{Treated}) - \bar{y}(\text{Control})$$

$$T = 0.54$$



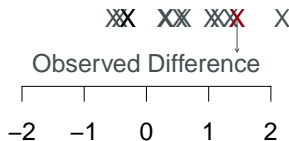
# A Naive approach to Permutation Testing

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



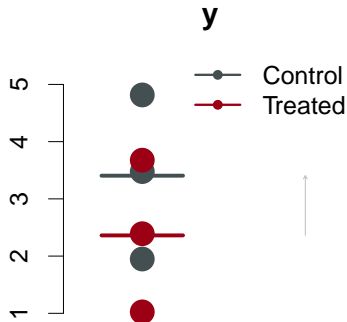
$$T = \bar{y}(\text{Treated}) - \bar{y}(\text{Control})$$

$$T = -0.3$$



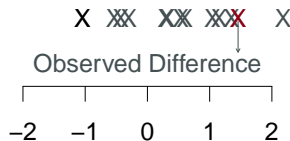
# A Naive approach to Permutation Testing

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



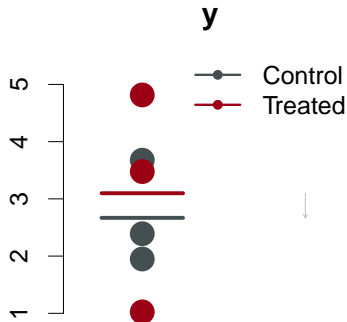
$$T = \bar{y}(\text{Treated}) - \bar{y}(\text{Control})$$

$$T = -1.04$$



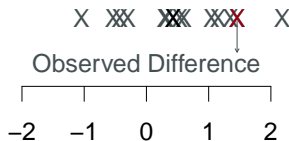
# A Naive approach to Permutation Testing

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



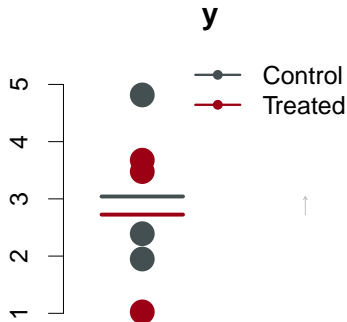
$$T = \bar{y}(\text{Treated}) - \bar{y}(\text{Control})$$

$$T = 0.43$$



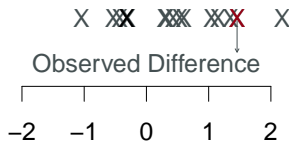
# A Naive approach to Permutation Testing

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



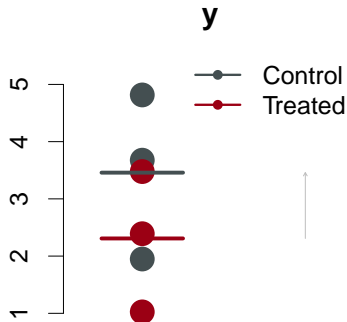
$$T = \bar{y}(\text{Treated}) - \bar{y}(\text{Control})$$

$$T = -0.32$$



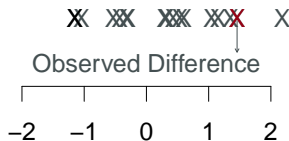
# A Naive approach to Permutation Testing

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



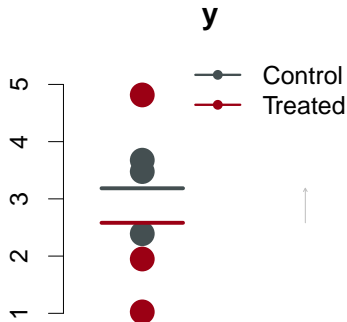
$$T = \bar{y}(\text{Treated}) - \bar{y}(\text{Control})$$

$$T = -1.15$$



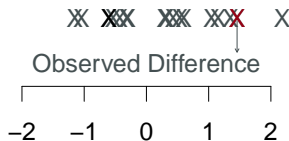
# A Naive approach to Permutation Testing

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



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

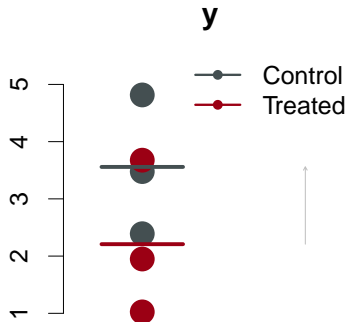
$$T = -0.6$$





# A Naive approach to Permutation Testing

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



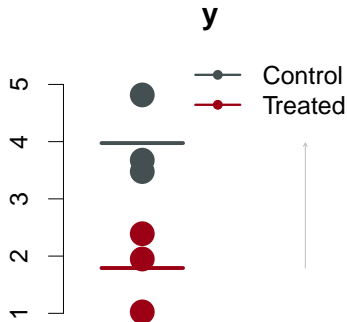
$$T = \bar{y}(\text{Treated}) - \bar{y}(\text{Control})$$

$$T = -1.35$$



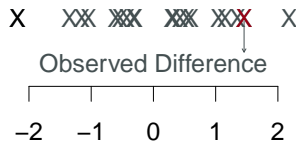
# A Naive approach to Permutation Testing

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



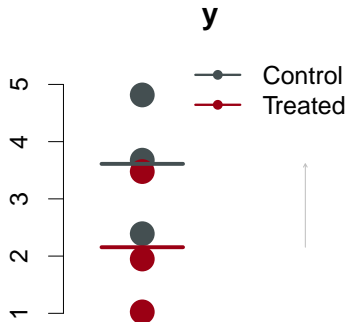
$$T = \bar{y}(\text{Treated}) - \bar{y}(\text{Control})$$

$$T = -2.18$$



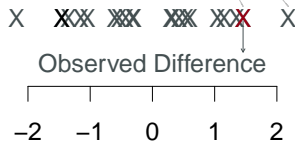
# A Naive approach to Permutation Testing

...and compute the p-value!



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

$$T = -1.46$$



# 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

- 1 Introduction
- 2 A toy example
- 3 A Naive approach to Permutation Testing
- 4 Theory (very short)**
- 5 Other cases
- 6 Multivariate Testing



# A (slight) more formal approach

(see also Pesarin, 2001)

$= (y_1, y_2, \dots, y_n)$  the vector of observed data

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

$$\mathbf{G} = \{*: f(*) = f()\}$$

(and  $|\mathbf{G}|$  number of elements of  $\mathbf{G}$ )

If we assume exchangeability of observations, then:

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

( $\pi^* \in \Pi$ ,  $\Pi$  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 G** 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  $\mathbf{G}$**  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  $\mathbf{G}$ , this induces an ordering on  $\mathbf{G}$ .

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

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

i.e. each permutation is equally likely in the Orbit  $\mathbf{G}$ .

The **p-value**:

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



# The package flip

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

To install the github version type (in R):

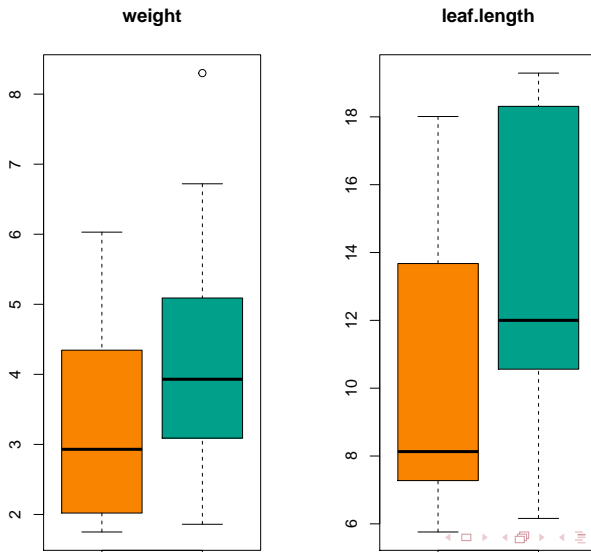
```
library(devtools)  
install_github('livioivil/flip')
```



# 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:

- $H_0(\text{weight}) : F(\text{weight}|\text{grp} = 0) = F(\text{weight}|\text{grp} = 1)$   
vs
- $H_1(\text{weight}) : F(\text{weight}|\text{grp} = 0) > F(\text{weight}|\text{grp} = 1)$

And about leaf length:

- $H_0(\text{leaf.len}) : F(\text{leaf.len}|\text{grp} = 0) = F(\text{leaf.len}|\text{grp} = 1)$   
vs
- $H_1(\text{leaf.len}) : F(\text{leaf.len}|\text{grp} = 0) > F(\text{leaf.len}|\text{grp} = 1)$

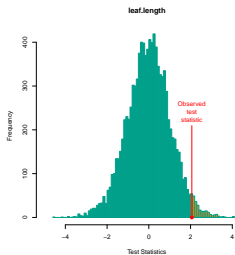
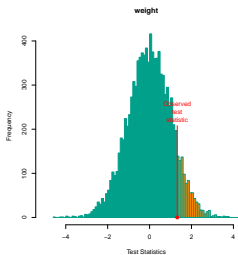


# Hypothesis testing

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

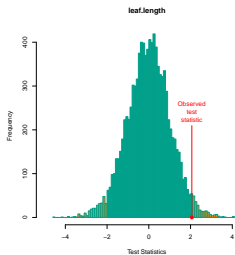
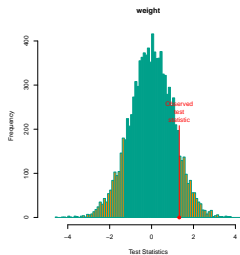


# Two-tailed tests

```
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  $|\mathbf{G}|$ ) 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.

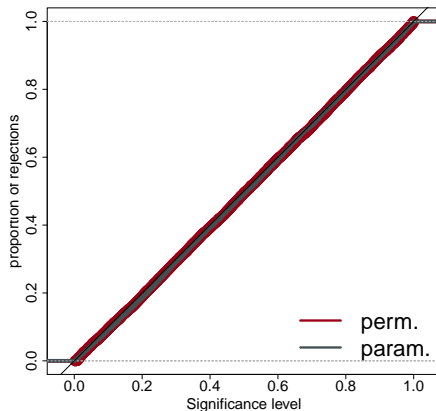


# Simulation: normal distribution

- 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: $H_0$

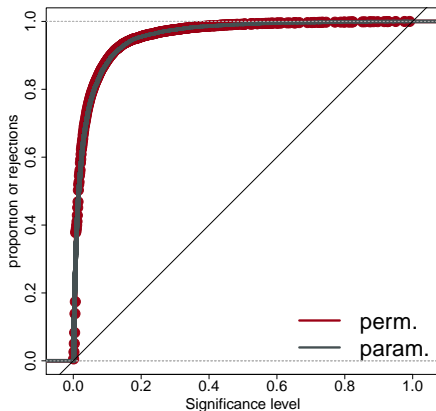


Empirical Type I error	$\leq 0.01$	$\leq 0.05$	$\leq 0.1$	$\leq 0.5$	$\leq 0.75$
Permutation	0.00	0.05	0.09	0.49	0.74
Parametric t.test	0.01	0.04	0.09	0.49	0.75



## Simulation: $H_1$

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



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

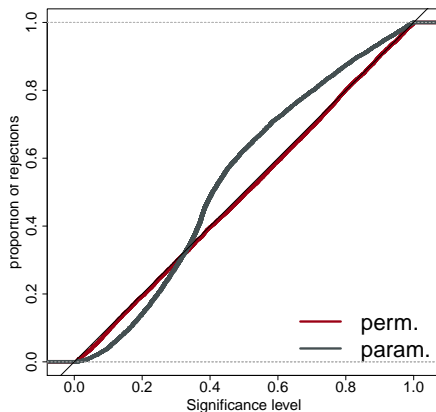


# Simulation: Cauchy distribution

- 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: $H_0$

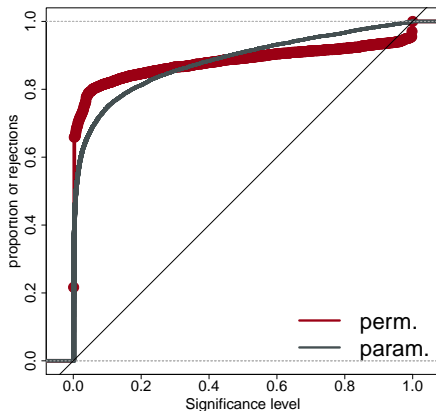


Empirical Type I error	$\leq 0.01$	$\leq 0.05$	$\leq 0.1$	$\leq 0.5$	$\leq 0.75$
Permutation	0.00	0.04	0.09	0.49	0.74
Parametric t.test	0.00	0.01	0.04	0.62	0.84



## Simulation: $H_1$

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



Empirical Power	$\leq 0.01$	$\leq 0.05$	$\leq 0.1$	$\leq 0.5$	$\leq 0.75$
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

- 1 Introduction
- 2 A toy example
- 3 A Naive approach to Permutation Testing
- 4 Theory (very short)
- 5 Other cases**
- 6 Multivariate Testing





# A very general approach

This approach (Orbit **G** + 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

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



# 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  $\mathbf{G}$ ?

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

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

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



# Outline

- 1 Introduction
- 2 A toy example
- 3 A Naive approach to Permutation Testing
- 4 Theory (very short)
- 5 Other cases
- 6 Multivariate Testing**



# Multivarite hypotheses

Testing  $H_0(\text{weight})$  + 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:

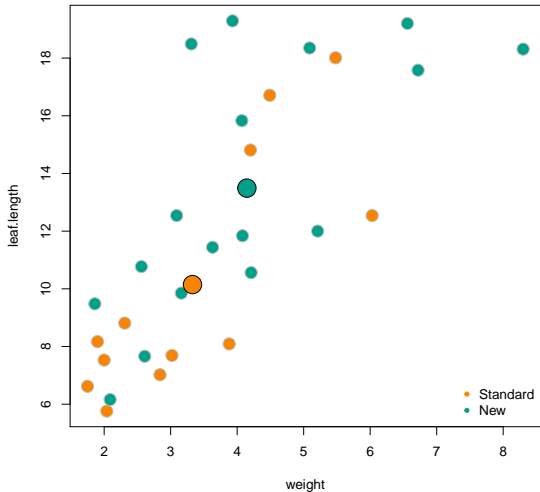
- MANOVA test
  - is OK but only for linear models (2 or more samples).
  - assumes multivariate normality.
  - Does not allow for one-sided alternatives.
- Bonferroni correction ( $p = \min(p_1, p_2) * 2$ )
  - very simple, always valid
  - does not take in account dependences among data
  - (i.e. may be very conservative, i.e. high final p-value)





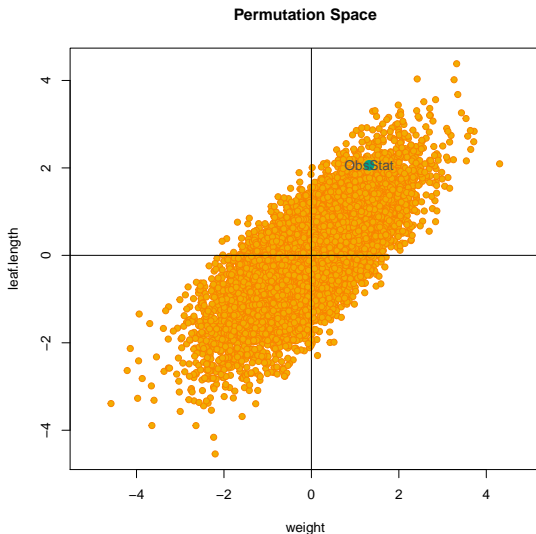
# Joint distributions of the data

The two variables are dependent:



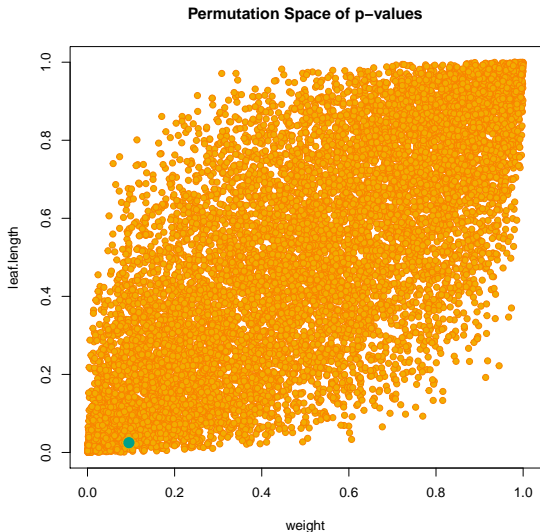
# Joint distributions of test statistics

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  $* \in \mathbf{G}$  )



# Nonparametric Combination methodology (Pesarin, 2001)

The Orbit is now defined in a multivariate framework:

$$\mathbf{G} = \{(\mathbf{1}, \mathbf{2})^* : f((\mathbf{1}, \mathbf{2})^*) = f((\mathbf{1}, \mathbf{2}))\}$$

(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  $\mathbf{1}, \mathbf{2}^*$ .



# Nonparametric Combination methodology (Pesarin, 2001)

The Orbit is now defined in a multivariate framework:

$$\mathbf{G} = \{((1,2)^* : f((1,2)^*) = f((1,2)))\}$$

(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  $_{1,2}^{*,*}$ .

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  $\psi^\circ$  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  $\psi^\circ$  under  $H_0$ .

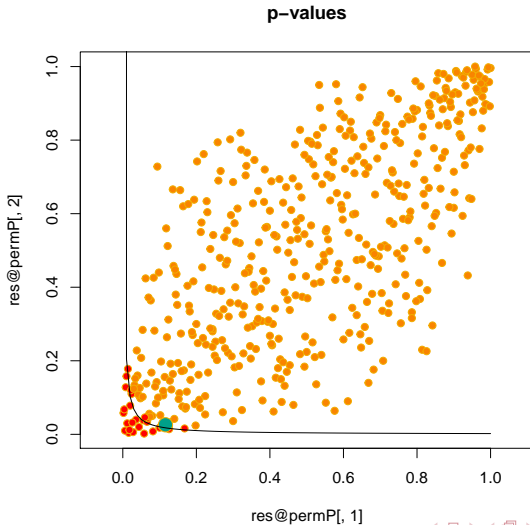
Apply  $T((p_1, p_2)^*) = \psi((p_1, p_2)^*)$  and compute the  $p_{global}$ .



# Fisher combining function

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

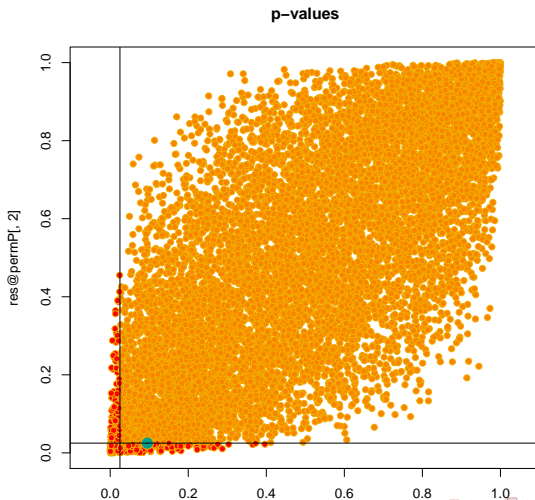
$$p_{\text{global}} = 0.0400$$



# Tippett (min-p) combining function

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

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



# The need for Multiple Testing

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





# The need for 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)



# The need for 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  $H_0(\text{weight})$  or  $H_0(\text{leaf.len})$  is false.
- We don't know which one! (weak control of the FamilyWise Error)
- How to cast them into multiple testing procedures?



# The need for 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  $H_0(\text{weight})$  or  $H_0(\text{leaf.len})$  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!



# 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; Hemerik, Goeman & Finos, 2018)

Software 'R': libraries 'coin', 'flip', 'permuco'

