

Permutation Tests

Livio Finos

University of Padova

Contents

1	Introduction	2
1.1	Introduction	2
1.2	Renewed interest toward permutation testing	2
1.3	The package <code>flip</code>	3
1.4	The Age vs Reaction Time Dataset	3
1.5	Measuring the dependence between two variables	4
2	Permutation approach to Hypothesis Testing	6
2.1	Permutation tests - in a nutshell	6
2.2	To sum up	11
2.3	A slightly more formal approach	13
2.4	A comparison (and relationships) with parametric linear model	16
2.5	Permutationally equivalent tests	17
3	Some special cases	19
3.1	Rank-correlation	19
3.2	The Two-independent-sample problem	20
3.3	Chi square and other categorical methods	22
3.4	ANOVA (C-sample)	24
3.5	Stratified permutations (discrete nuisances)	25
4	Multivariate Testing	26
4.1	Seeds data	26
4.2	Marginal vs Joint distribution	27
4.3	Rejection regions (and overall testing)	28

5	FWER control via Permutations tests	32
5.1	Permutation Bonferroni	32
5.2	Improved Bonferroni	32
5.3	Multiple testing using permutations	33
5.4	Correlation structure of p-values	33
5.5	Improved Holm: Westfall & Young	33
5.6	The general framework: Closed Testing	34
5.7	Conclusion	35
6	A case study: Pharmacokinetic Study of Carbidopa	36
6.1	A solution	36
7	(minimal) Bibliography	40

1 Introduction

1.1 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.
- Fisher exact test is a prototypical example, but
- the general approach has restricted applicability without the support of a computer.

1.2 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.

1.3 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')
```

Before we start

```
#clean the memory
rm (list=ls ())

# We customize the output of our graphs a little bit
par.old=par ()
par (cex.main=1.5, lwd=2, col="darkgrey", pch=20, cex=3)
# par (par.old)
palette (c ("#FF0000", "#00A08A", "#FFCC00", "#445577", "#45abff"))

# customize the output of knitr
knitr :: opts_chunk$set (fig.align="center")#, fig.width=6, fig.height=6)
```

1.4 The Age vs Reaction Time Dataset

The reaction time of these subjects was tested by having them grab a meter stick after it was released by the tester. The number of centimeters that the meter stick dropped before being caught is a direct measure of the person's response time.

The values of Age are in years. The Gender is coded as F for female and M for male. The values of Reaction.Time are in centimeters.

(data are fictitious)

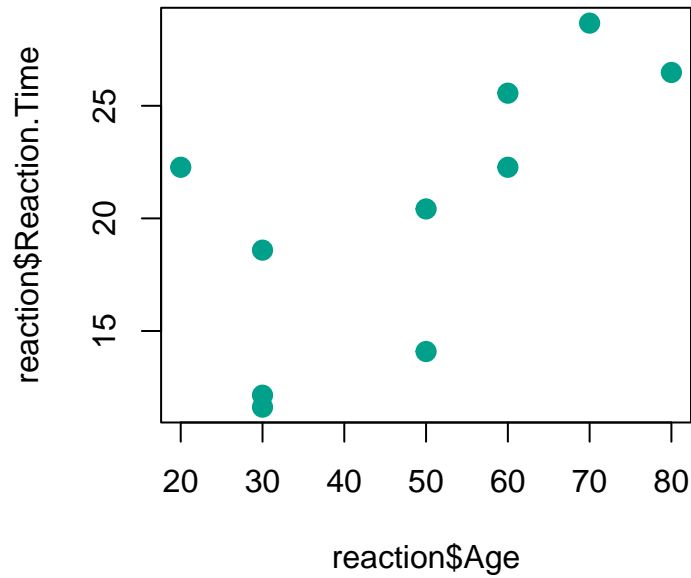
To read the data

```
data(reaction,package = "flip")
## or download it from: https://github.com/livioivil/flip/tree/master/data
## or
load("reaction.rda")
str (reaction)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    10 obs. of  3 variables:
## $ Age      : num  70 50 30 60 80 60 30 30 20 50
## $ Gender   : Factor w/ 2 levels "F","M": 1 1 2 1 2 1 2 2 1 2
## $ Reaction.Time: num  28.7 20.4 11.6 22.3 26.5 ...
```

We plot the data

```
plot(x=reaction$Age,y=reaction$Reaction.Time,pch=20,col=2,cex=2)
```



1.5 Measuring the dependence between two variables

we define:

- $X = Age$
- $Y = Reaction.Time$

We review some famous index to measure the (linear) dependence among two variables

1.5.1 Covariance and Variance

Covariance between X and Y :

$$\sigma_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}$$

- values between $-\infty$ and ∞
- $\sigma_{xy} \approx 0$: there is no dependency between X and Y
- $\sigma_{xy} > 0$ (< 0): there is a strong positive (negative) dependency between X and Y

Variance of X

$$\sigma_{xx} = \sigma_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Standard Deviation of X :

$$\sigma_{xx} = \sqrt{\sigma_{xx}} = \sigma_x$$

1.5.2 Correlation

With the Covariance it is difficult to understand when the relationship between X and Y is strong/weak. We note that

$$-\sigma_x\sigma_y \leq \sigma_{xy} \leq \sigma_x\sigma_y \text{ is equivalent to } -1 \leq \frac{\sigma_{xy}}{\sigma_x\sigma_y} \leq 1$$

Correlation between X and Y :

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x\sigma_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

- values between -1 and 1
- $\rho_{xy} \approx 0$: there is no dependency between X and Y
- $\rho_{xy} \approx 1(-1)$: there is a strong positive (negative) dependency between X and Y

1.5.3 Linear Trend, the least squares method

We describe the relationship between **Reaction.Time** and **Age** with a straight line.

$$E(\text{Reaction.Time}) \approx \beta_0 + \beta_1 \text{Age}$$

$$E(Y) = \beta_0 + \beta_1 X$$

Let's draw a line 'in the middle' of the data.

The **least-squares estimator**

We look for the one that passes more 'in the middle', the one that minimizes the sum of the squares of the residues:

$\hat{\beta}_0$ and $\hat{\beta}_1$ such that $\sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$ is minimum.

Estimates:

- Angular coefficient: $\hat{\beta}_1 = \frac{\sigma_{xy}}{\sigma_{xx}} = \rho_{xy} \frac{\sigma_y}{\sigma_x} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = 0.2064719$
- Intercept: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 10.3013483$
- Response (estimated y): $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$
- Residuals (from the estimated response): $y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) = y_i - \hat{y}_i$

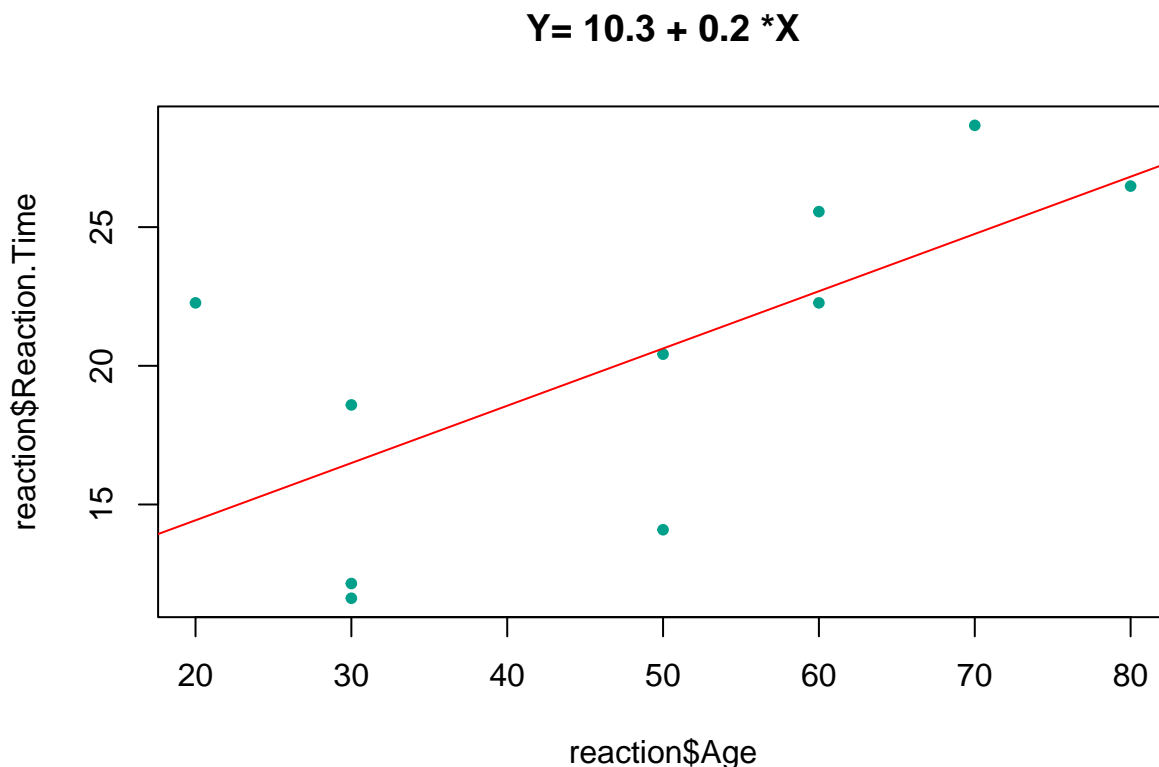
and therefore the least squares are the sum of the squared residuals: $\sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$

A graphical representation:

```
model=lm(Reaction.Time~Age,data=reaction)
coefficients(model)
```

```
## (Intercept)      Age
## 10.3013483    0.2064719
```

```
plot(reaction$Age,reaction$Reaction.Time,pch=20,col=2,cex=1)
coeff=round(coefficients(model),1)
title(paste("Y=",coeff[1],"+",coeff[2],"*X"))
abline(model,col=1)
```



2 Permutation approach to Hypothesis Testing

2.0.1 Some remarks

Let's note that all the measures above does not make any assumptions on the random process that generate them.

Let now assume that Y - and possibly X - is generated by a random variable.

Further minimal assumptions will be specified later.

The question: **Is there a relationship between Y and X ?**

We estimated $\hat{\beta}_1 = 0.2064719$

But the **true value** β_1 is really different from 0 (i.e. no relationship)?

Otherwise, is the difference from 0 due to the random sampling?

- **Null Hypothesis** $H_0 : \beta_1 = 0$ (the **true** β_1 , not its estimate $\hat{\beta}_1$!). There is no relationship between X and Y .
- **Alternative Hypothesis** $H_1 : \beta_1 > 0$ The relationship is positive.

Other possible specifications of $H_1 : \beta_1 < 0$ and, more commonly, $H_1 : \beta_1 \neq 0$.

2.1 Permutation tests - in a nutshell

As a toy example, let use a sub-set of the data:

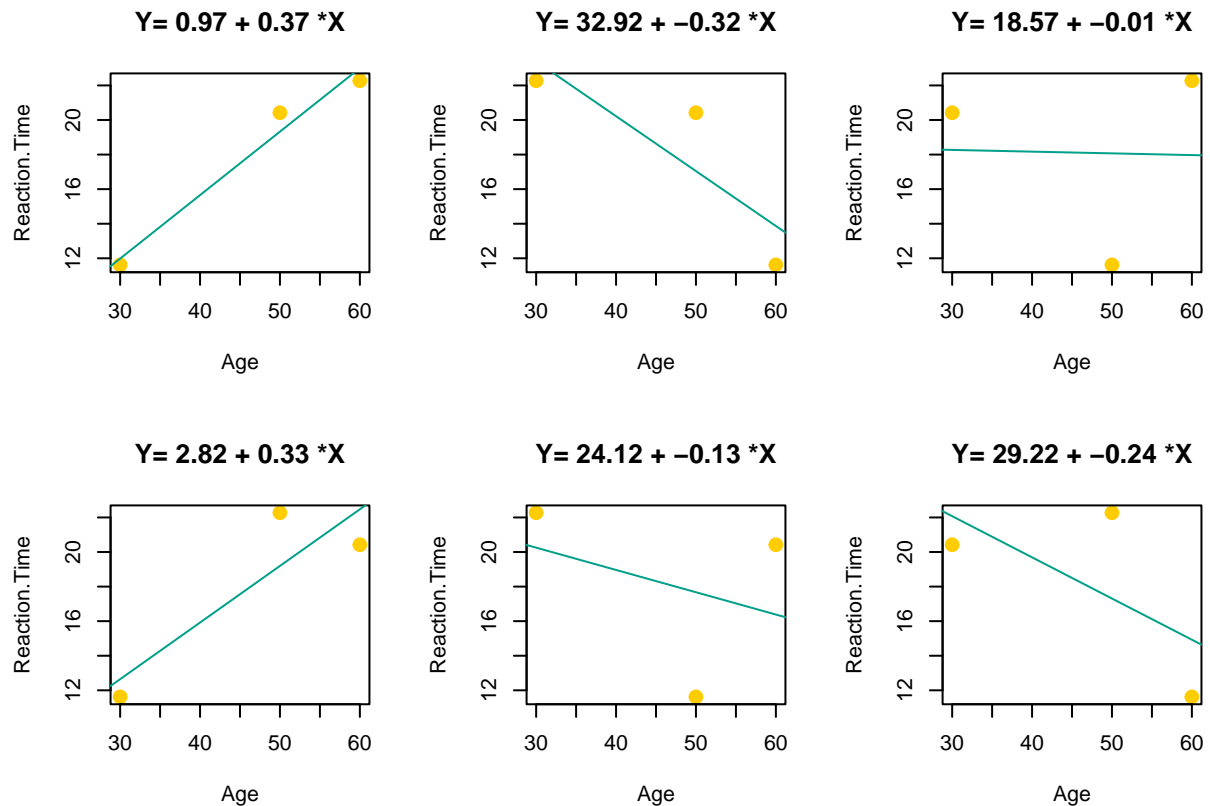
##	Age	Gender	Reaction.Time
## 2	50	F	20.42
## 3	30	M	11.62
## 4	60	F	22.27



- If H_0 is true: there is no linear relationship between X and Y
- Therefore, the trend observed on the data is due to chance.
- Any other match of x_i and y_i was equally likely to occur
- I can generate the datasets of other hypothetical experiments by exchanging the order of the observations in Y .
- How many equally likely datasets could I get with X and Y observed? $3! = 6$ possible datasets.

Remark: Here we only assume that y is a random variable. The only assumption here is the exchangeability of the observations: the joint density $f(y_1, \dots, y_n)$ does not change when the ordering of y_1, \dots, y_n is changed.

2.1.1 All potential datasets



2.1.1.1 In our data set We apply the same principle to the complete dataset...

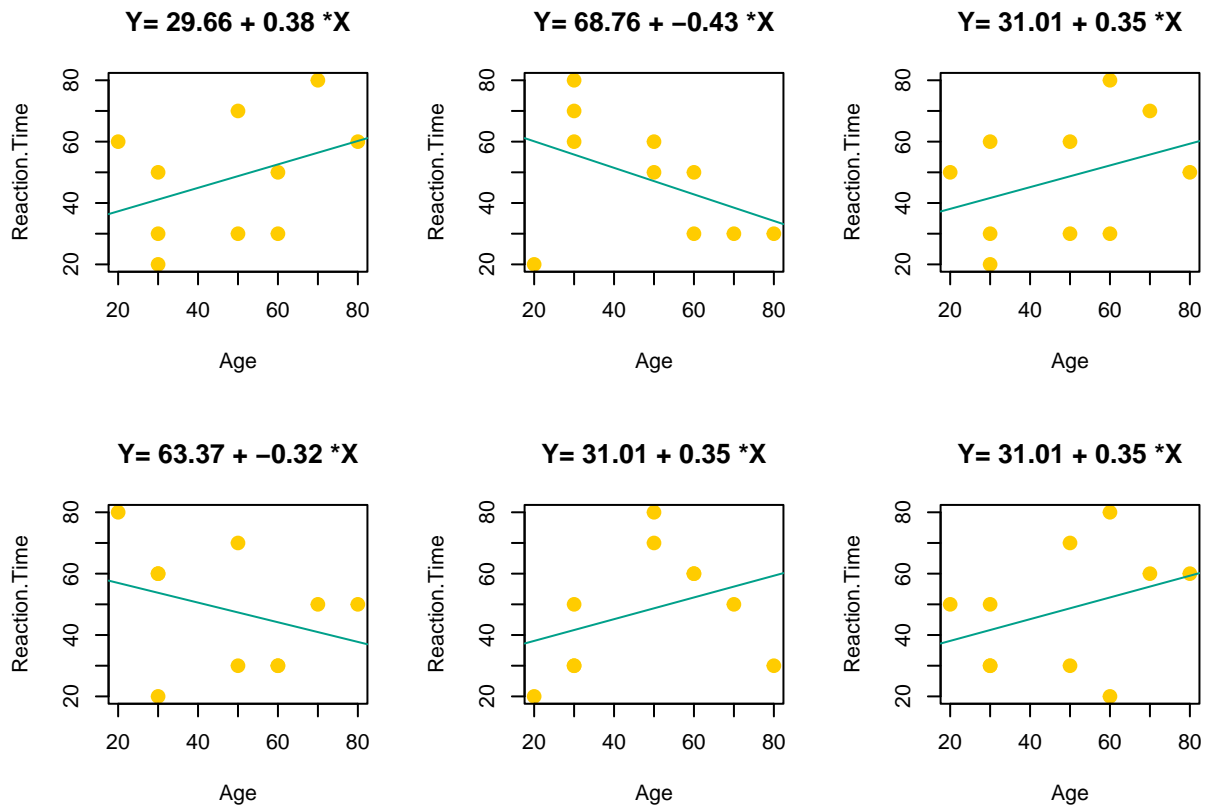
How many permutations of the vector y_1, \dots, y_n are possible? $n! = 10! = 3628800$.

big, perhaps not too big ... but what happen with, for example, $n = 20$? We got $20! = 2.432902e + 18$. This is too big, definitely!

We calculate a smaller (but sufficiently large) B of random permutations.

here some example

Age vs a permutations of Reaction.Time

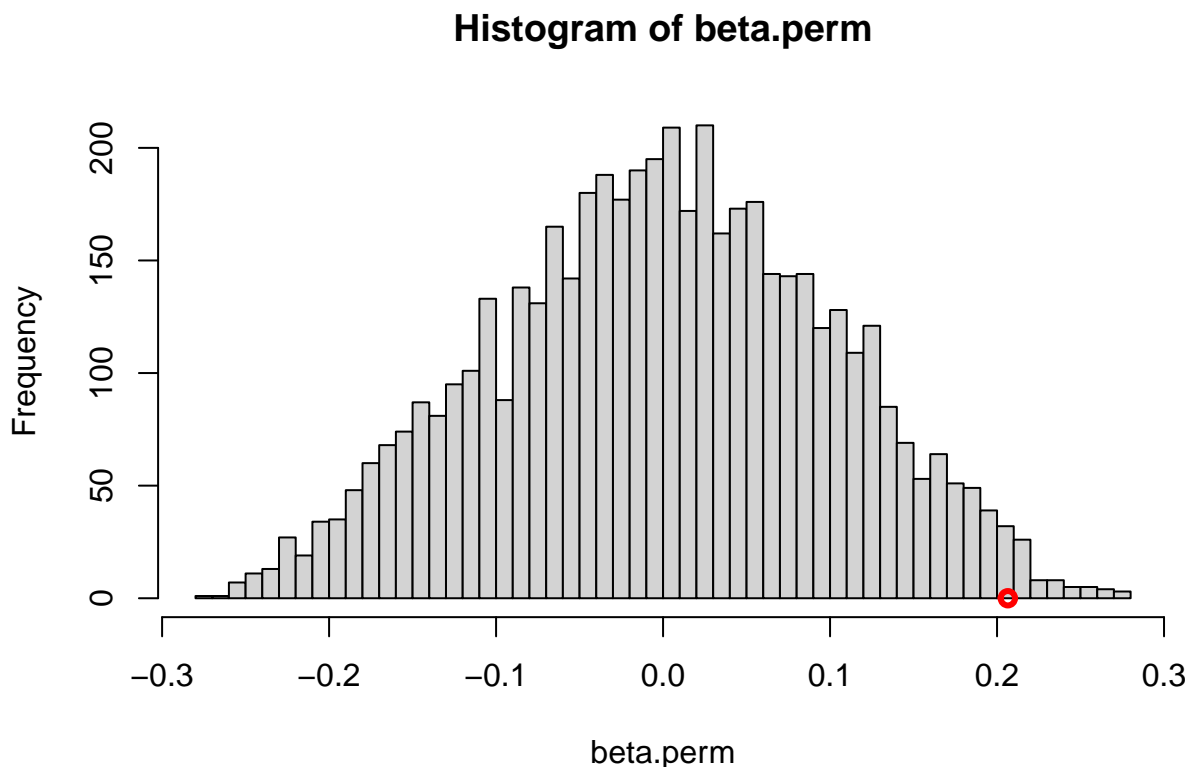


We repeat 5000 times and we look at the histogram of the $\hat{\beta}_1$

```
# beta_1 estimated on the observed data:
beta1=coefficients(lm(Reaction.Time~Age,data=reaction))[2]

# function that permutes the y values and calculates the coeff beta_1
my.beta.perm <- function(Y,X){
  model=lm(sample(Y)~X)
  coefficients(model)[2]
}

#replicate it B-1 times
beta.perm= replicate(B,my.beta.perm(reaction$Reaction.Time, reaction$Age ))
```



2.1.2 How likely WAS $\hat{\beta}_1^{obs}$?

(before the experiment!)

How likely was it to get a $\leq \hat{\beta}_1^{obs}$ value among the many possible values of $\hat{\beta}_1^{*b}$ (obtained by permuting data)?

Remarks:

- $\hat{\beta}_1^{*b} < \hat{\beta}_1^{obs}$ (closer to 0): less evidence against H_1 than $\hat{\beta}_1^{obs}$
- $\hat{\beta}_1^{*b} \geq \hat{\beta}_1^{obs}$: equal or more evidence towards H_1 than $\hat{\beta}_1^{obs}$

2.1.3 Calculation of the p-value

Over $B=5000$ permutations we got 4939 times a $\hat{\beta}_1^{*b} \leq \hat{\beta}_1^{obs}$.

The p-value (significance) is $p = \frac{\#(\hat{\beta}_1^{*b} \geq \hat{\beta}_1^{obs})}{B} = 0.0126$

($\hat{\beta}_1^{obs}$ counts as a random permutation)

2.1.4 Interpretation

The probability of $p = P(\hat{\beta}_1^* \geq \hat{\beta}_1 = 0.206 | H_0)$ is equal to $p = 0.0126$, i.e. very small.

So, it was unlikely to get a value like this **IF** H_0 is true.

Neyman-Pearson's approach has made common the use of a significance threshold for example $\alpha = .05$ (or $= .01$). When $p \leq \alpha$ rejects the hypothesis that there is no relationship between X and Y (H_0). If so, we are inclined to think that H_1 is true (there is a positive relationship).

- Type I error: False Positive
the true hypo is H_0 (null correlation), BUT we accept H_1 (correlation is positive)
- Type II error: False Negative
the true hypo is H_1 (positive correlation), BUT we do not reject H_0 (null correlation)

2.2 To sum up

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: $T = \hat{\beta}_1 = \hat{\sigma}_{xy} / \hat{\sigma}_{xx}$ is the (estimated) slope.
Higher the slope, higher the evidence for H_1 .

Type I error control

We want to guarantee not to get false relationships (a few false positives), better to be conservative. To make this, we want to bound the probability to make a false discovery:

$$P(p\text{-value} \leq \alpha | H_0) \leq \alpha$$

We built a machinery that in the long run (many replicates of the experiment) finds false correlations with probability α (e.g. $0.05 = 5\%$).

2.2.1 We make it in flip

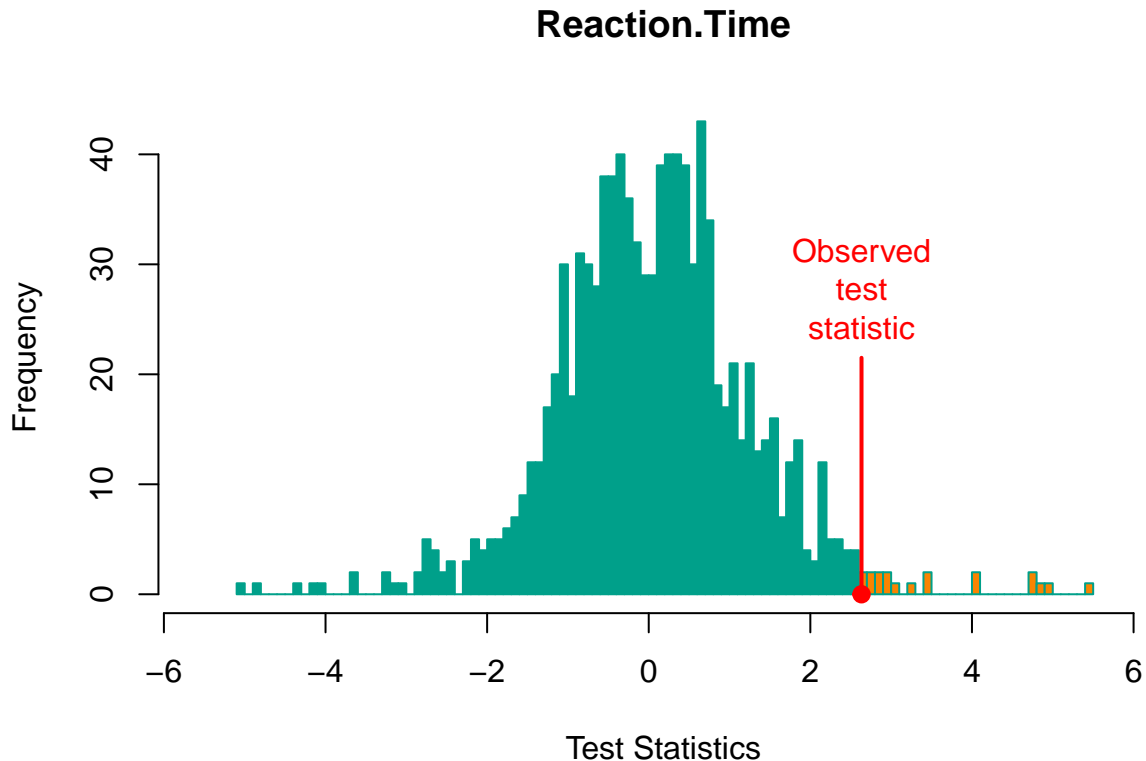
```
library(flip)

(res=flip(Reaction.Time~Age,data=reaction,tail=1))

##
##               Test  Stat tail p-value
## Reaction.Time    t 2.633    >  0.0180

## compare also with
# flip(Reaction.Time~Age,data=reaction,tail=1,statTest = "cor")
# flip(Reaction.Time~Age,data=reaction,tail=1,statTest = "coeff")

plot(res)
```



Type I error control

We want to guarantee not to get false relationships (a few false positives), better to be conservative. To make this, we want to bound the probability to make a false discovery:

$$P(p - value \leq \alpha | H_0) \leq \alpha$$

We built a machinery that in the long run (many replicates of the experiment) finds false correlations with probability α (e.g. $0.05 = 5\%$).

2.2.2 Composite alternatives (bilateral)

The hypothesis $H_1 : \beta_1 > 0$ (the relation is positive) must be justified with a priori knowledge.

More frequently, the Alternative hypothesis is appropriate: $H_1 : \beta_1 \neq 0$ (there is a relationship, I do not assume the direction)

I consider anomalous coefficients estimated as very small but also very large ('far from 0'). The p-value is

$$p = \frac{\#(|\hat{\beta}_1^{*b}| \geq |\hat{\beta}_1^{obs}|)}{B} = 0.0306$$

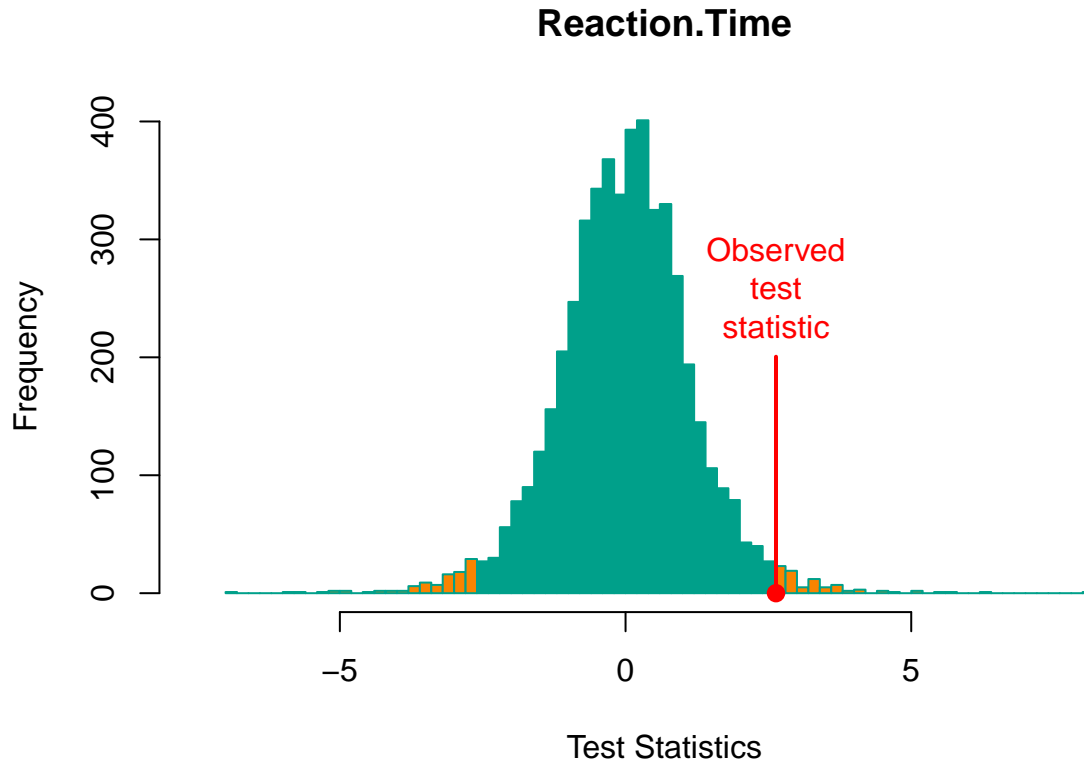
(remark: the observed test stat is included among the permuted one)

In flip:

```
library(flip)
(res=flip(Reaction.Time~Age,data=reaction,tail=0,perms=5000))
```

```
##
##           Test Stat tail p-value
## Reaction.Time      t 2.633   >< 0.0350
```

```
plot(res)
```



2.3 A slightly more formal approach

(see also Pesarin, 2001; Hemerik & Goeman, 2018)

Let Y be data taking values in a sample space \mathcal{Y} . Let Π be a – usually finite – set of transformations $\pi : \mathcal{Y} \rightarrow \mathcal{Y}$, such that Π is a **group** with respect to the operation of composition of transformations, that is:

- it contains identity,
- every element has an inverse in the group,
- closure: if $\pi_1, \pi_2 \in \Pi$: $\pi_1 \circ \pi_2 \in \Pi$

(e.g. Π set of all possible permutations)

Null Hypothesis

$$H_0 : Y \sim P \in \Omega_0$$

Randomization Hypothesis Under the null hypothesis, the distribution of Y is invariant under the transformations in Π ; that is, for every π in Π , πY and Y have the same distribution whenever Y has distribution P in Ω_0 .

(See also Lehmann, E. L., & Romano, J. P. (2006). Testing statistical hypotheses. Springer Science & Business Media.)

Test statistic $T(Y) : \mathbb{R}^n \rightarrow \mathbb{R}$

$T^{(k)}(Y)$ is the $\lceil (1 - \alpha)|\Pi| \rceil$ -th sorted value of $T(\pi Y)$

Define the test:

$$\phi(Y) = \begin{cases} 1 & \text{if } T(Y) > T^{(k)}(Y) \\ 0 & \text{if otherwise} \end{cases} \quad (1)$$

Theorem: Under H_0 , $E_P(\phi(Y)) = \alpha$, that is $P(T(Y) > T^{(k)}(Y)) \leq \alpha$.

Proof

For each $Y \in \mathcal{Y}$, define O_Y to be the orbit of Y , which is the set $\{\pi Y : \pi \in \Pi\} \subseteq \mathcal{Y}$.

Let $A = \{Y \in \mathcal{Y} : T(Y) > T^{(k)}(Y)\}$ be the set of elements of the sample space that lead to rejection. Suppose H_p holds. By the group structure, $\Pi\pi = \Pi$ for all $\pi \in \Pi$. Consequently, $T^{(k)}(\pi Y) = T^{(k)}(Y)$ for all $\pi \in \Pi$. Thus,

$$\#\{\pi \in \Pi : \pi Y \in A\} = \#\{\pi \in \Pi : T(\pi Y) > T^{(k)}(\pi Y)\} = \#\{\pi \in \Pi : T(\pi Y) > T^{(k)}(Y)\} \leq \alpha |O_Y|$$

.

Endow the space of orbits with the σ -algebra that it inherits from the σ -algebra on \mathcal{Y} . Analogously to the proof of Theorem 15.2.2 in Lehmann (2005), we obtain

$$P(Y \in A | O_Y) = \frac{|A|}{|O_Y|}.$$

By the argument above, this is bounded by α . Hence,

$$P(Y \in A) = \mathbb{E}\{P(Y \in A | O_Y)\} \leq \alpha$$

as was to be shown.

Alternative Proof

By construction, $\sum_{\pi \in \Pi} \phi(\pi Y) = |\Pi|\alpha$. Therefore $|\Pi|\alpha = E_P(\sum_{\pi \in \Pi} \phi(\pi Y)) = \sum_{\pi \in \Pi} E_P(\phi(\pi Y))$

Next, by the null hypothesis: $E_P(\phi(Y)) = E_P(\phi(\pi Y))$, so that $|\Pi|\alpha = \sum_{\pi \in \Pi} E_P(\phi(Y)) = |\Pi|E_P(\phi(Y))$ gives $E_P(\phi(Y)) = \alpha$

(See also Lehmann, E. L., & Romano, J. P. (2006). Testing statistical hypotheses. Springer Science & Business Media.)

2.3.1 More about permutation testing

Randomization Hypothesis and Group structure

Randomization Hypothesis applied to $\pi \in \Pi$ having a Group structure implies

$$(T(\pi_1 Y), \dots, T_1(\pi_{|\mathcal{O}|} Y)) = (T(\pi_1 \pi Y), \dots, T_1(\pi_{|\mathcal{O}|} \pi Y))$$

for each $\pi \in \Pi$.

Orbit of \mathcal{O} :

$$\mathcal{O} = \{\pi Y : \pi \in \Pi\} \subseteq \mathcal{Y}.$$

(losely) the set of all samples having the same likelihood under H_0 .

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

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

If we assume exchangeability of observations, then:

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

Remark about assumption of exchangeability: This means that, Under the Null Hypothesis, observations within subject are assumed to be exchangeable: e.g. $f(y_1, y_2) = f(y_2, y_1)$.

This assumption is always true as long as observations:

- are **identically distributed**,
- have the **same dependence**, e.g. the same correlation.

Parametric t -test and linear models assumes independence (more stringent than ‘same dependence’), and normality of the errors, i.e. more severe assumptions than permutation approach.

When normality is not met, the parametric approach only provides asymptotic control of the type I error, while permutation approach provides exactness.

An Intuition about the proof for an alternative proof of the control of the type I error

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

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

(due to group structure)

$$\begin{aligned} E(\phi(Y)|\mathbf{y} \in \mathcal{O}, H_0) &= \\ P(T(\mathbf{y}) > T^{(k)}|\mathbf{y} \in \mathcal{O}, H_0) &= \\ = \int_{T^{(k)}}^{+\infty} f(T(\mathbf{y}))dT(\mathbf{y}) &= \\ = \sum_{\mathbf{y} \in \mathcal{O}} I(T(\mathbf{y}) > T)/|\mathcal{O}| \leq \alpha \quad \forall \mathcal{O} \end{aligned}$$

And now $E(\phi(\mathbf{y})) = \int_P E(\phi(\mathbf{y})|\mathbf{y} \in \mathcal{O}, H_0)d\mathbf{y}$

2.3.2 Properties (see Pesarin, 2001)

The theorem above proves that the permutation tests have **exact control of the type I error**, i.e. $P(p\text{-value} \leq \alpha|H_0) = \alpha$ assuming $\alpha \in \{1/|\mathcal{O}|, 2/|\mathcal{O}|, \dots, 1\}$ - don't forget that the orbit \mathcal{O} is a finite set and the cumulative distribution of $T(\pi\mathbf{y})$ is a step function.

When α has different values, the test is (slightly) conservative (or one need to use randomized tests that are not discussed in this course).

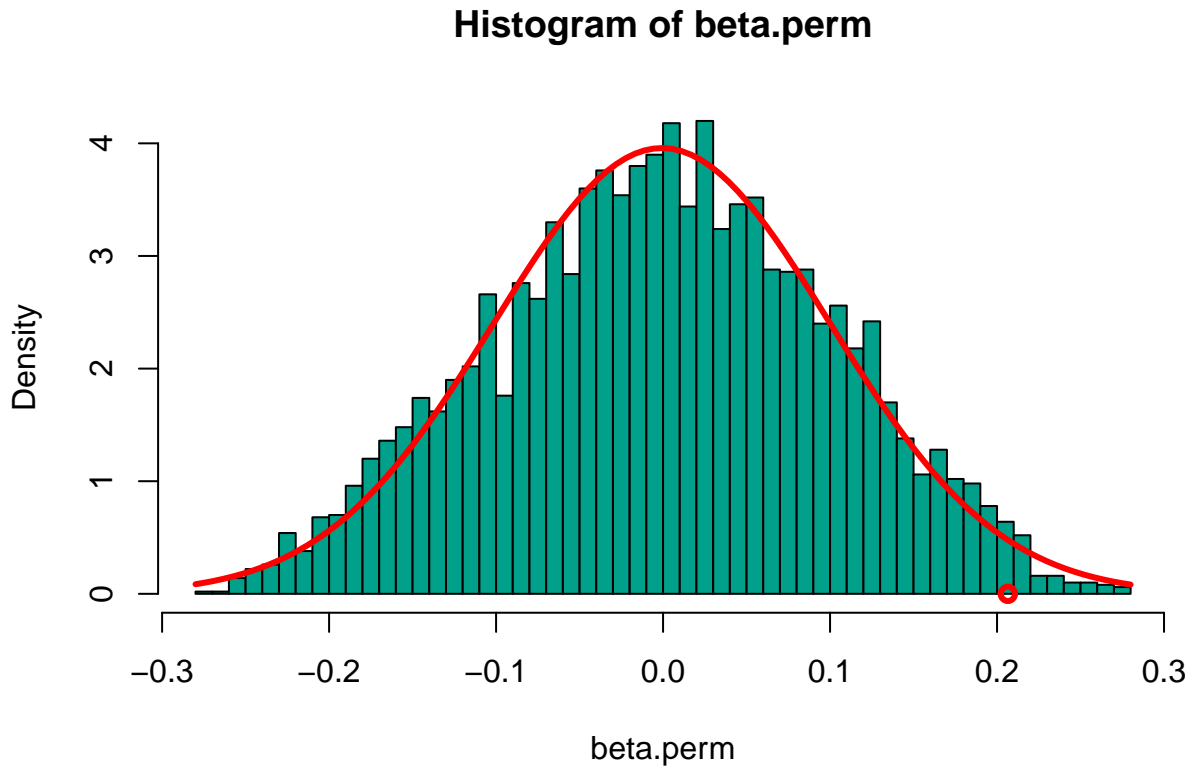
Further properties:

- The permutations tests are **Unbiased**: $P(p\text{-value} \leq \alpha|H_1) > \alpha$
- The test is **Consistent**: $P(p\text{-value} \leq \alpha|H_1) \rightarrow 1$ when $n \rightarrow \infty$
- The test converges to the parametric counterpart (when it exists)

2.4 A comparison (and relationships) with parametric linear model

We can see that the histogram of the statistical tests (calculated on the permuted data) is well described by a **Gaussian** (normal) curve.

```
hist(beta.perm,50,probability=TRUE,col=2)
curve(dnorm(x,mean(beta.perm),sd(beta.perm)),add=TRUE,col=1,lwd=3)
points(beta1,0,lwd=3,col=1)
```



2.4.1 The (simple) linear parametric model

We assume that the observed values are distributed around true values $\beta_0 + \beta_1 X$ according to a Gaussian law:

$Y = \text{linear part} + \text{normal error}$

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Assumptions of the linear model

- the $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ the relationship between X and Y is truly linear, less than the error term ε_i
- $\varepsilon_i \sim N(0, \sigma^2)$, $\forall i = 1, \dots, n$ errors have normal distribution with zero mean and common variance (homoscedasticity: same variance).

2.4.2 Hypothesis testing

If these assumptions are true,

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2 / \sum (x_i - \bar{x})^2)$$

We calculate the test statistic:

$$t = \frac{\hat{\beta}_1}{\text{std.dev } \hat{\beta}_1} = \frac{\hat{\beta}_1}{\sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / \sum (x_i - \bar{x})^2 / (n-2)}}$$

If $H_0 : \beta_1 = 0$, $t \sim t(n-2)$ is true

On reaction data and $H_1 : \beta_1 \neq 0$ (bilateral alternative)

```
model=lm (Reaction.Time ~ Age, data=reaction)
summary (model)

##
## Call:
## lm(formula = Reaction.Time ~ Age, data = reaction)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.535  -3.364  -0.272   2.676   7.839
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.30135     4.04407   2.547  0.0343 *
## Age          0.20647     0.07841   2.633  0.0300 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.678 on 8 degrees of freedom
## Multiple R-squared:  0.4643, Adjusted R-squared:  0.3973
## F-statistic: 6.934 on 1 and 8 DF,  p-value: 0.03003
```

Similar result, but much more assumptions!

2.4.3 Assumptions of a permutation test

What model do we assume in a permutation test?

Under the null hypo: $H_0 : f(y) = f(y|x) \forall x$

Under the alternative hypo no assumptions. in order to have power we hope that:

$H_1 : E(y|x) = \beta_0 + \beta_1 x$; with $\beta_1 \neq 0$ and for some x

that is:

$H_1 : E(yx) \neq E(x)E(y)$

No other assumptions on the distribution of $f(y|x)$ (normality, nor finite moments)

2.5 Permutationally equivalent tests

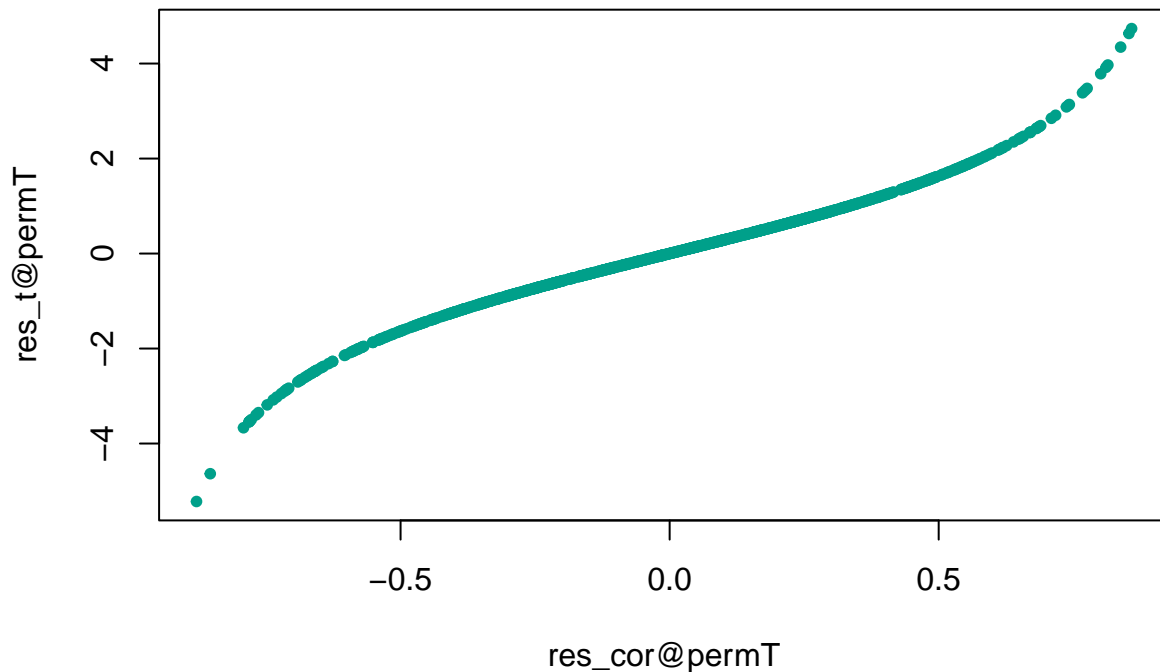
```
set.seed(1)
(res_cor=flip(Reaction.Time~Age,data=reaction,statTest = "cor"))

##
##              Test   Stat tail p-value
## Reaction.Time  cor 0.6814   >< 0.0410
```

```
set.seed(1)
(res_t=flip(Reaction.Time~Age,data=reaction,statTest = "t"))
```

```
##
##              Test  Stat tail p-value
## Reaction.Time    t 2.633   ><  0.0410
```

```
plot(res_cor@permT,res_t@permT,pch=20,col=2)
```



2.5.1 Conclusion

The permutation tests:

- Different from bootstrap methods. The former are extractions without reintegration, the latter with. The former have almost optimal properties and have (almost always) an exact control of the first type errors.
- They constitute a general approach and are applicable in many contexts. Very few assumptions.
- some dedicated R packages:
 - `coin` <http://cran.r-project.org/web/packages/coin/index.html>
 - `permuco` <https://cran.r-project.org/web/packages/permuco/index.html>
 - `flip` <http://cran.r-project.org/web/packages/flip/index.html> (the development version is on github <https://github.com/livioivil/flip>)
 - `flipscores` <http://cran.r-project.org/web/packages/flipscores/index.html> (the development version is on github <https://github.com/livioivil/flipscores>)
 - `multcomp` <https://cran.r-project.org/web/packages/multcomp/index.html>
 - `GFD` <https://cran.r-project.org/web/packages/GFD/index.html>

3 Some special cases

3.1 Rank-correlation

- n observations from y , we are interested on $F(y|x)$
 - we don't need y_1 and y_2 to be continuous, we don't even need to have finite moments (usual minimal assumption).
- Hypotheses
 - $H_0 : F(y|x) = F(y|x') \forall x, x'$
 - $H_1 : \exists x < x' : F(y|x) < F(y|x')$ or directional such as: $H_1 : \exists x, x' F(y_1) \neq F(y_2)$
- Test Statistic: rank-correlation

```
(res=flip(Reaction.Time~Age,data=reaction,perms = 5000,statTest = "rank"))

##
##                               Test   Stat tail p-value
## Reaction.Time Wilcoxon 2.189    ><  0.0204

# to see the rank correlation use the workaround:
(res=flip(rank(reaction$Reaction.Time)~rank(reaction$Age),perms = 5000,statTest = "cor"))

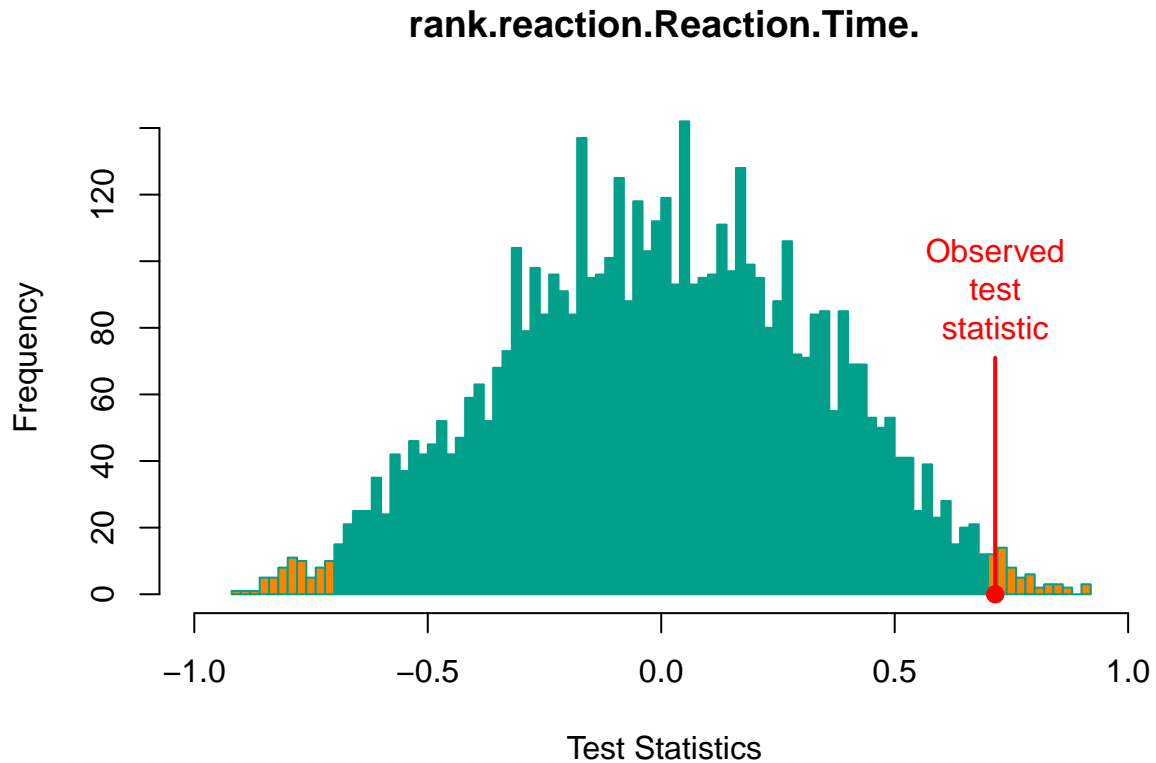
##
##                               Test   Stat tail p-value
## rank.reaction.Reaction.Time. cor 0.7153    ><  0.0222

(cor.test(reaction$Reaction.Time,reaction$Age,method="spe"))

## Warning in cor.test.default(reaction$Reaction.Time, reaction$Age, method =
## "spe"): Impossibile calcolare p-value esatti in presenza di ties

##
## Spearman's rank correlation rho
##
## data:  reaction$Reaction.Time and reaction$Age
## S = 46.983, p-value = 0.02005
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.715256

plot(res)
```



3.2 The Two-independent-sample problem

- Two samples:
 - n_1 observations from y_1
 - n_2 observations from y_2
 - we don't need y_1 and y_2 to be continuous, we don't even need to have second (nor higher order) finite moments, which is the usual minimal assumption.
- Hypotheses
 - $H_0 : F(y_1) = F(y_2)$
 - $H_1 : F(y_1) \neq F(y_2)$
(or directional such as: $H_1 : F(y_1) < F(y_2)$)
- Test Statistic:
 - Standardized mean difference (t-statistic)
 - Estimated slope coefficient (label of groups as dummy predictor)
 - other test statistic such as the (non standardized) mean difference are permutationally equivalent

```
data("seeds")
seeds=na.omit(seeds)

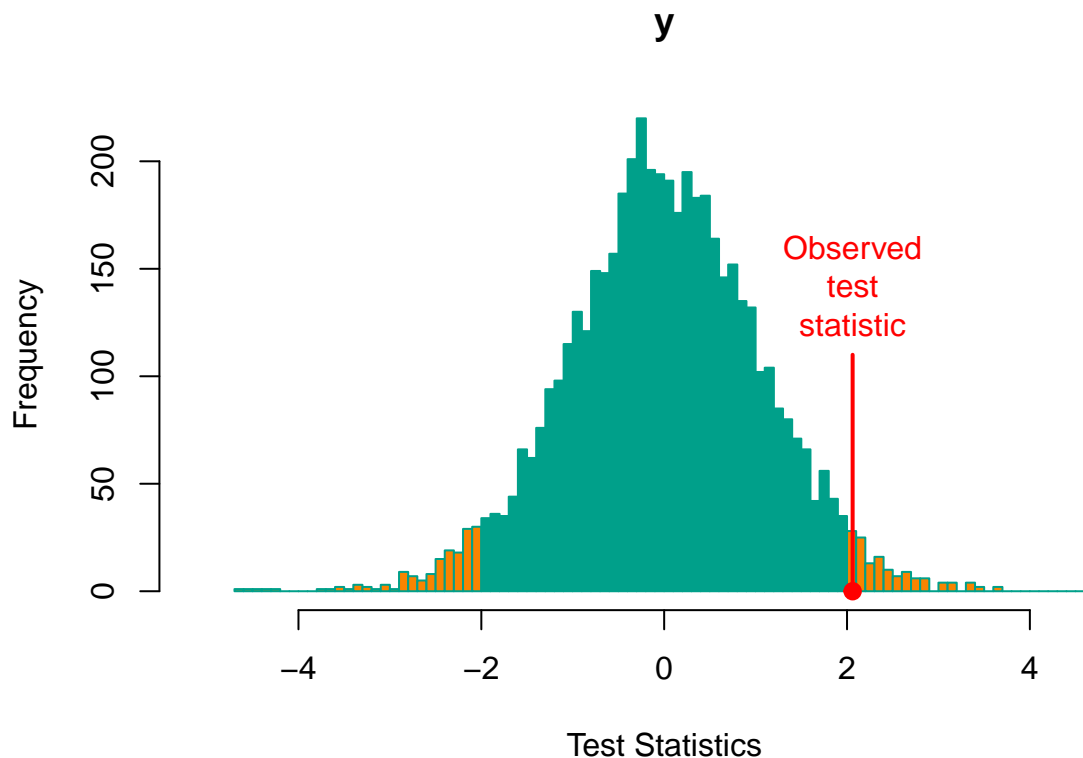
(res=flip(y~grp,data=seeds,perms = 5000))
```

```
##
## Test Stat tail p-value
## y t 2.061 >< 0.0526
```

```
(summary(lm(y~grp,data=seeds)))
```

```
##
## Call:
## lm(formula = y ~ grp, data = seeds)
##
## Residuals:
## Min 1Q Median 3Q Max
## -7.331 -2.931 -1.651 4.663 7.863
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.147 1.242 8.168 9e-09 ***
## grp 3.345 1.623 2.061 0.049 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.303 on 27 degrees of freedom
## Multiple R-squared: 0.136, Adjusted R-squared: 0.104
## F-statistic: 4.249 on 1 and 27 DF, p-value: 0.04903
```

```
plot(res)
```



3.2.1 Rank test

Can we use rank-based statistics?

Yes, equivalent to rank-tests, we just rely on exact distribution instead of asymptotic one (and we have no limitations with ties).

```
(res=flip(y~grp,data=seeds,statTest = "rank",perms=5000))
```

```
##
##          Test  Stat tail p-value
## y Wilcoxon 2.148   ><  0.0292
```

```
(wilcox.test(y~grp,data=seeds))
```

```
## Warning in wilcox.test.default(x = DATA[[1L]], y = DATA[[2L]], ...): non è
## possibile calcolare p-value esatto in presenza di ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: y by grp
## W = 53.5, p-value = 0.03353
## alternative hypothesis: true location shift is not equal to 0
```

3.3 Chi square and other categorical methods

```
data("seeds")
seeds$Germinated=!is.na(seeds$x)
seeds$Germinated=factor(seeds$Germinated)
seeds$grp=factor(seeds$grp)

table(seeds$grp,seeds$Germinated)
```

```
##
##      FALSE TRUE
## 0         8   12
## 1         3   17
```

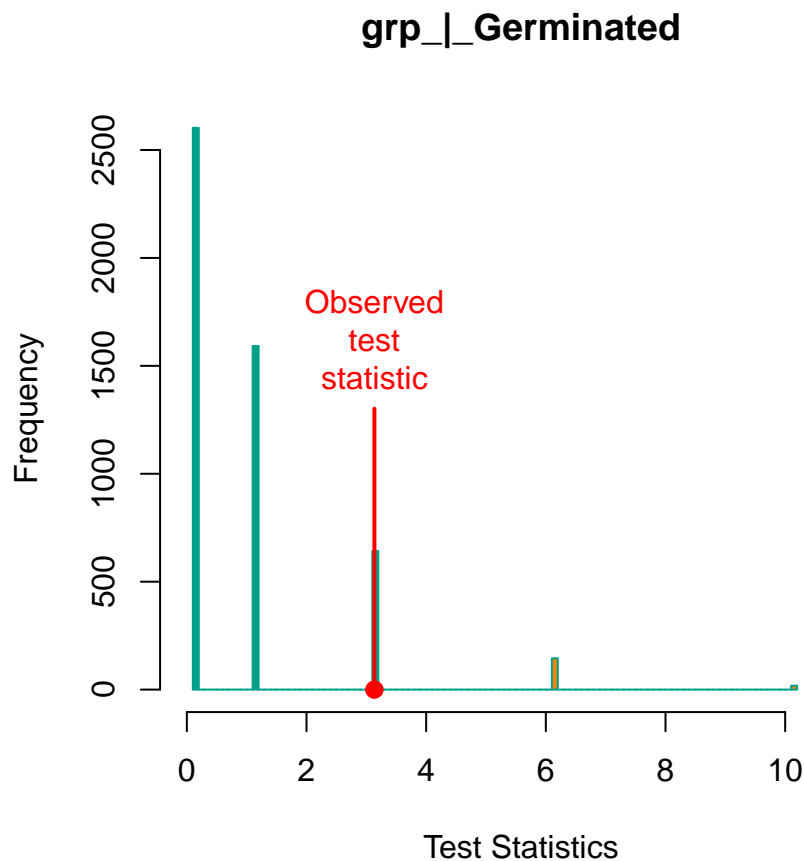
```
chisq.test(seeds$grp,seeds$Germinated)
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data: seeds$grp and seeds$Germinated
## X-squared = 2.0063, df = 1, p-value = 0.1567
```

```
(res=flip(Germinated~grp,data=seeds,statTest = "Chisq",perms=5000))
```

```
##
##               Test Stat tail p-value
## grp_|_Germinated Chi Squared 3.135    > 0.1610
```

```
plot(res)
```



... and the Fisher test:

```
fisher.test(seeds$grp,seeds$Germinated)$p.value
```

```
## [1] 0.1551874
```

```
(flip(Germinated~grp,data=seeds,perms=5000))
```

```
##
##               Test Stat tail p-value
## GerminatedFALSE t -1.798    >< 0.1596
## GerminatedTRUE  t  1.798    >< 0.1596
```

3.4 ANOVA (C-sample)

e.g. 3 groups of Age: young [18 – 35), middle age [35 – 60), old [60 – 100)

- C samples:
 - n_i observations from y_i ($i = 1, \dots, C$)
 - we don't need y_i to be continuous, we don't even need to have finite moments (usual minimal assumption)
- Hypotheses
 - $H_0 : F(y_i) = F(y_j) \forall (i, j)$
 - $H_1 : \exists (i, j) : F(y_i) \neq F(y_j)$
- Test Statistic:
 - F-statistic
 - R^2
 - other test statistic such as the (non standardized) mean difference are permutationally equivalent
 - Rank-based is also possible

```
reaction$AgeCateg=cut(reaction$Age,c(18,35,65,100),right = FALSE)

(res=flip(Reaction.Time~AgeCateg,data=reaction,perms = 5000,statTest = "ANOVA"))
```

```
##
##              Test Stat tail p-value
## Reaction.Time    F 4.02    > 0.0780
```

```
summary(lm(Reaction.Time~AgeCateg,data=reaction))
```

```
##
## Call:
## lm(formula = Reaction.Time ~ AgeCateg, data = reaction)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.495 -3.279  0.465  2.246  6.112
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      16.157      2.331   6.932 0.000225 ***
## AgeCateg[35,65)    4.428      3.296   1.343 0.221144
## AgeCateg[65,100)  11.418      4.037   2.828 0.025478 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.662 on 7 degrees of freedom
## Multiple R-squared:  0.5346, Adjusted R-squared:  0.4016
## F-statistic:  4.02 on 2 and 7 DF,  p-value: 0.06878
```


3.4.1 Stochastic Ordering

- Same assumptions of ANOVA
- Hypotheses
 - same null hypo $H_0 : F(y_i) = F(y_j) \forall (i, j)$
 - BUT $H_1 : \exists (i, j) : F(y_i) < F(y_j)$ (or $>$)

(more details on NPC later)

```
(res=flip(Reaction.Time~AgeCateg,data=reaction,perms = 5000,tail=1))
```

```
##
##                               Test   Stat tail p-value
## Reaction.Time_|_AgeCateg.[35,65).    t 0.1423    > 0.4336
## Reaction.Time_|_AgeCateg.[65,100).    t 2.2444    > 0.0220
```

```
npc(res)
```

```
##
##      comb.funct nVar  Stat p-value
## V1      Fisher    2 4.652 0.0202
```

3.5 Stratified permutations (discrete nuisances)

What if we want to test $x = \text{Age}$ also using $z = \text{Gender}$ as nuisance in the `reaction` data set?

Under the null hypothesis: $f(y|x, z) = f(y|x', z) = f(y|z) \forall (x, x')$

Therefore, even under the H_0 , it holds $f(y_i) = f(y_j)$ ONLY IF $z_i = z_j$ (obs i and j have the same gender).

Can e permute same as in the previous cases? NO. We permute the observations only within the strata defined by z .

Remark:

- we don't assume linear effect of the nuisance,
- we also allow heteroscedastic errors among strata.

(Test statistic remains the same)

```
(res=flip(Reaction.Time~Age,Strata=~Gender,data=reaction,perms=5000))
```

```
##
##               Test   Stat tail p-value
## Reaction.Time    t 2.633   >< 0.0718
```

Alternative model (more about NPC later):

```
(res=flip(Reaction.Time~Age*Gender,Strata=~Gender,data=reaction,perms=5000))
```

```
##
##               Test      Stat tail p-value
## Reaction.Time_|_Age      t  2.4826  ><  0.0750
## Reaction.Time_|_Age:Gender.M.      t -0.6518  ><  0.3394
```

```
npc(res)
```

```
##
##      comb.funct nVar  Stat p-value
## V1      Fisher    2 3.671 0.1406
```

4 Multivariate Testing

4.1 Seeds data

```
# install.packages("flip")
library(flip)
```

omit the NAs:

```
data(seeds,package = "flip")
seeds=na.omit(seeds)
seeds
```

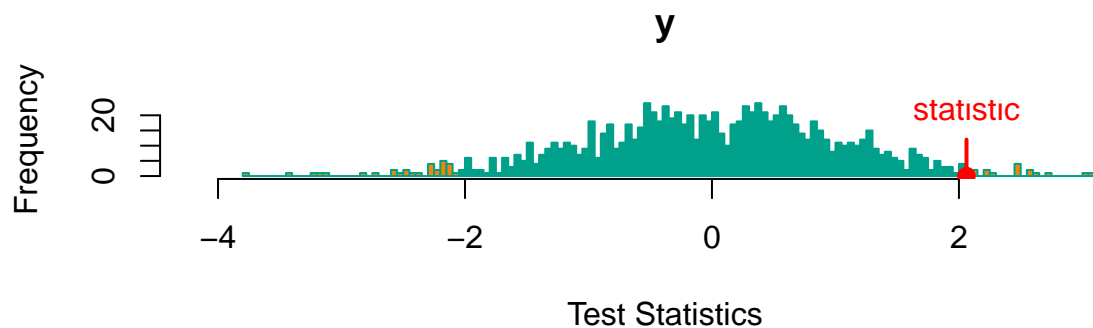
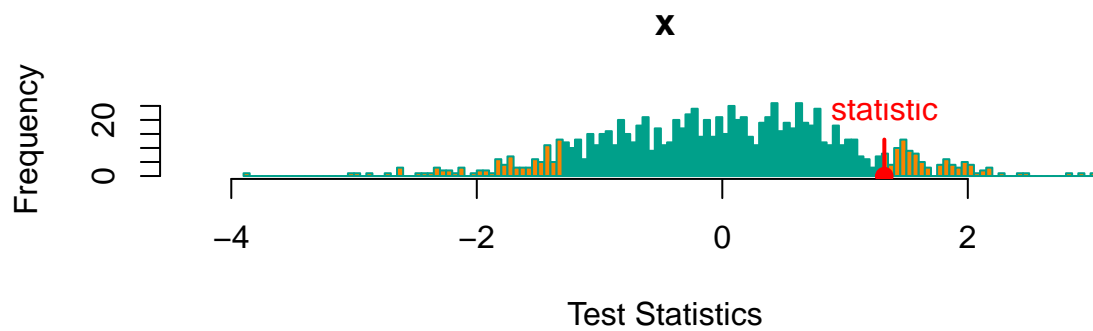
```
##      grp      x      y
## 9      0 6.03 12.54
## 10     0 4.20 14.81
## 11     0 4.49 16.71
## 12     0 2.00  7.53
## 13     0 2.84  7.02
## 14     0 3.88  8.09
## 15     0 2.04  5.76
## 16     0 5.48 18.01
## 17     0 2.31  8.81
## 18     0 1.90  8.17
## 19     0 1.75  6.62
## 20     0 3.02  7.69
## 24     1 3.31 18.49
## 25     1 6.56 19.20
## 26     1 3.16  9.85
## 27     1 4.07 15.83
## 28     1 2.09  6.16
## 29     1 6.72 17.58
## 30     1 3.93 19.29
## 31     1 2.56 10.77
## 32     1 8.30 18.31
## 33     1 4.21 10.56
## 34     1 1.86  9.48
## 35     1 3.09 12.54
## 36     1 5.09 18.35
```

```
## 37  1 4.08 11.84
## 38  1 3.63 11.44
## 39  1 2.61  7.66
## 40  1 5.21 12.00
```

4.2 Marginal vs Joint distribution

Use a permutation methods to test if there is any difference between the two groups in `grp` on the two variables `x` and `y`:

```
library(flip)
res=flip(.-grp,data=seeds,flipReturn =list(permP=TRUE,permT=TRUE))
hist(res)
```

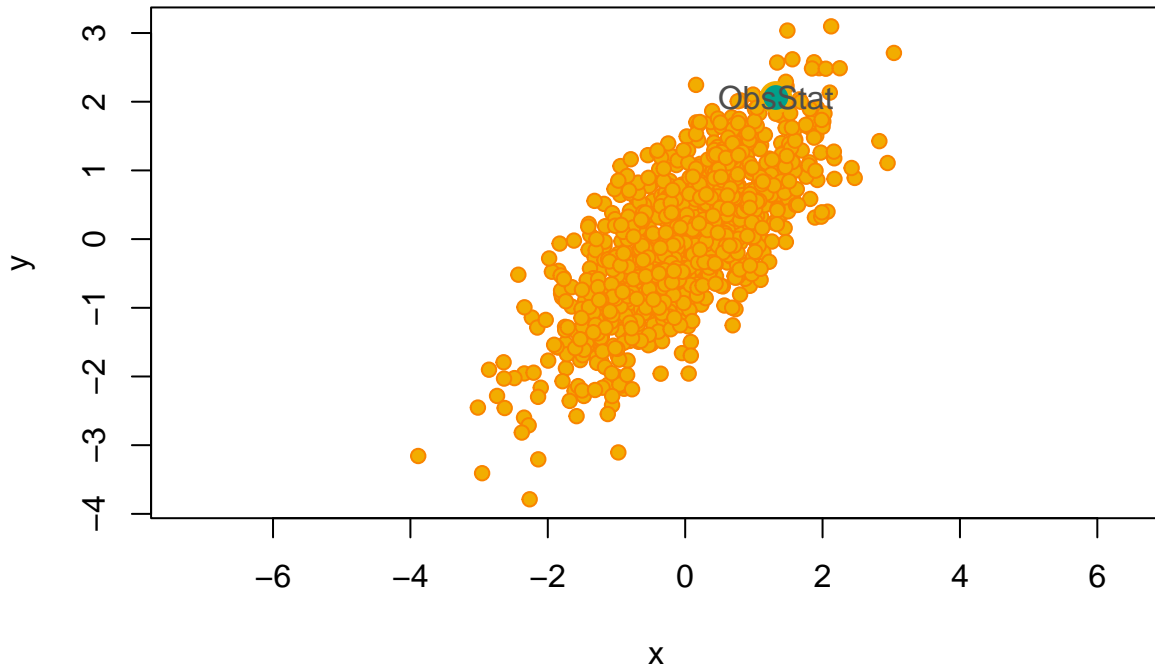


```
# flipReturn =list(permP=TRUE,permT=TRUE) is not really needed, we will need it later
```

We can perform the two tests for the two variables separately. But we don't get the overall p-value (is there any difference among the ANY of the two variables?)

```
plot(res)
```

Permutation Space



What will we see next:

- how to combine the two p-values (e.g. using the Fisher Combining Function) to test the global hypothesis
- How to use a closed testing procedure to adjust the 2 p-values: which variables?.

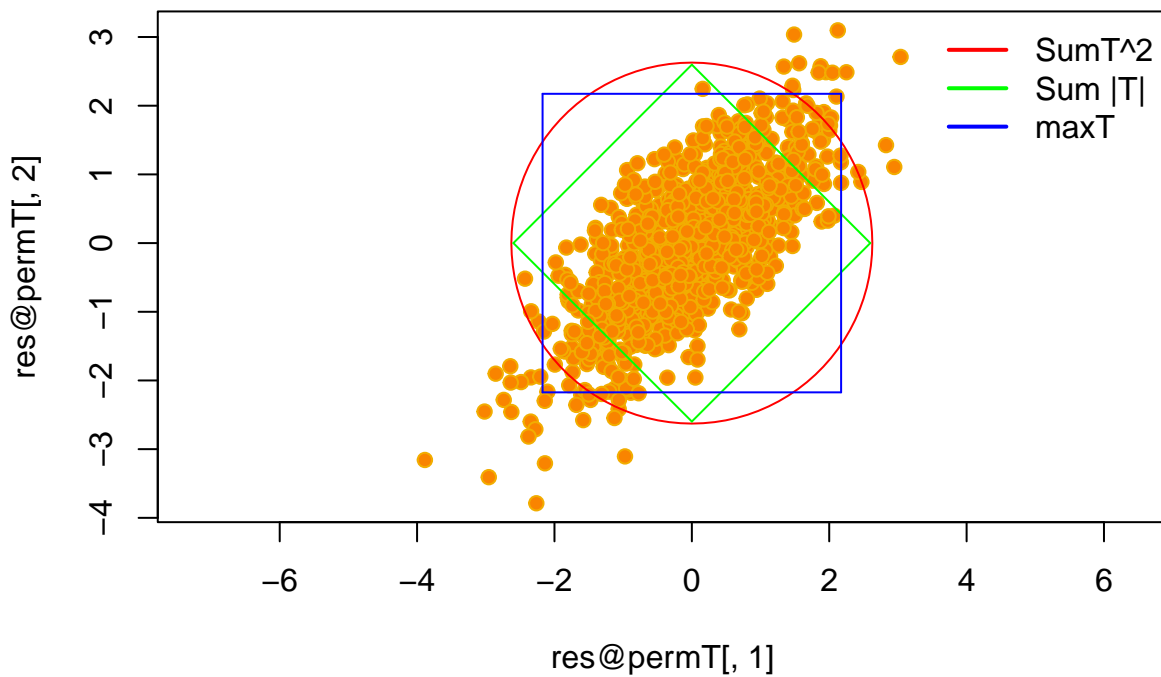
4.3 Rejection regions (and overall testing)

In the univariate setting it is easy to define what is *far from the null hypothesis* (i.e. usually from a value of 0 of the test statistic), while in the multivariate setting, there is not only one answer (nor a better one).

```
# install.packages("plotrix")
library("plotrix")
res.sumt2=npc(res,"sumT2",flipReturn=list(permP=TRUE,permT=TRUE))
limsumt2=res.sumt2@permT[which.min(abs(res.sumt2@permP-0.05))]
res.sumt=npc(res,"sumT",flipReturn=list(permP=TRUE,permT=TRUE))
limsumt=res.sumt@permT[which.min(abs(res.sumt@permP-0.05))]
res.maxt=npc(res,"maxT",flipReturn=list(permP=TRUE,permT=TRUE))
limmaxt=res.maxt@permT[which.min(abs(res.maxt@permP-0.05))]

plot(res@permT[,1],res@permT[,2],col="#F2AD00",bg="#F98400",pch=21,main="Some rejection region (alpha=.05)")
draw.circle(0, 0, limsumt2^.5,border="red")
segments(c(limsumt,-limsumt,limsumt,-limsumt),c(0,0,0,0),
         c(0,0,0,0),c(limsumt,-limsumt,-limsumt,limsumt),col="green")
segments(c(limmaxt,-limmaxt,-limmaxt,limmaxt),c(limmaxt,limmaxt,-limmaxt,-limmaxt),
         c(-limmaxt,-limmaxt,limmaxt,limmaxt),c(limmaxt,-limmaxt,-limmaxt,limmaxt),col="blue")
legend("topright",legend=c("SumT^2","Sum |T|","maxT"),col=c("red","green","blue"),bty="n",lwd=2)
```

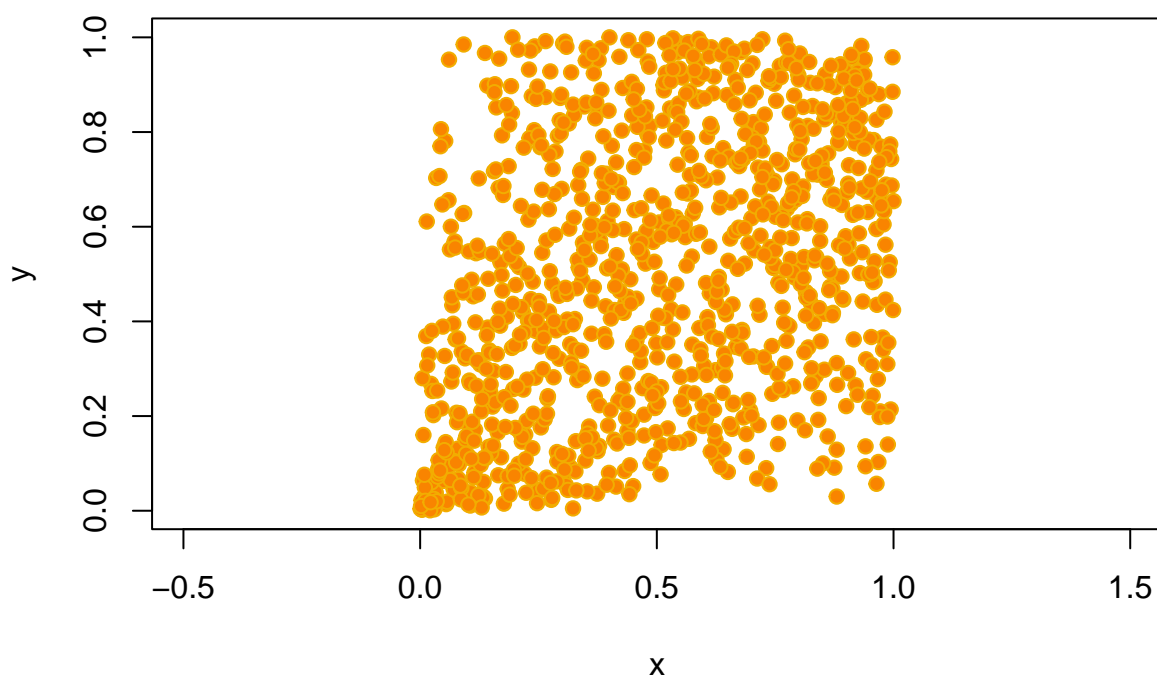
Some rejection region (alpha=.05)



REMARK We can derive the distribution of the p-value by computing the p-value for each test statistic (i.e. computed on observed and permuted data). We get then the multivariate distribution of the p-values, which looks something like this one:

```
plot(res@permP,col="#F2AD00",bg="#F98400",pch=21,main="Joint Distribution of p-values",asp=1)
```

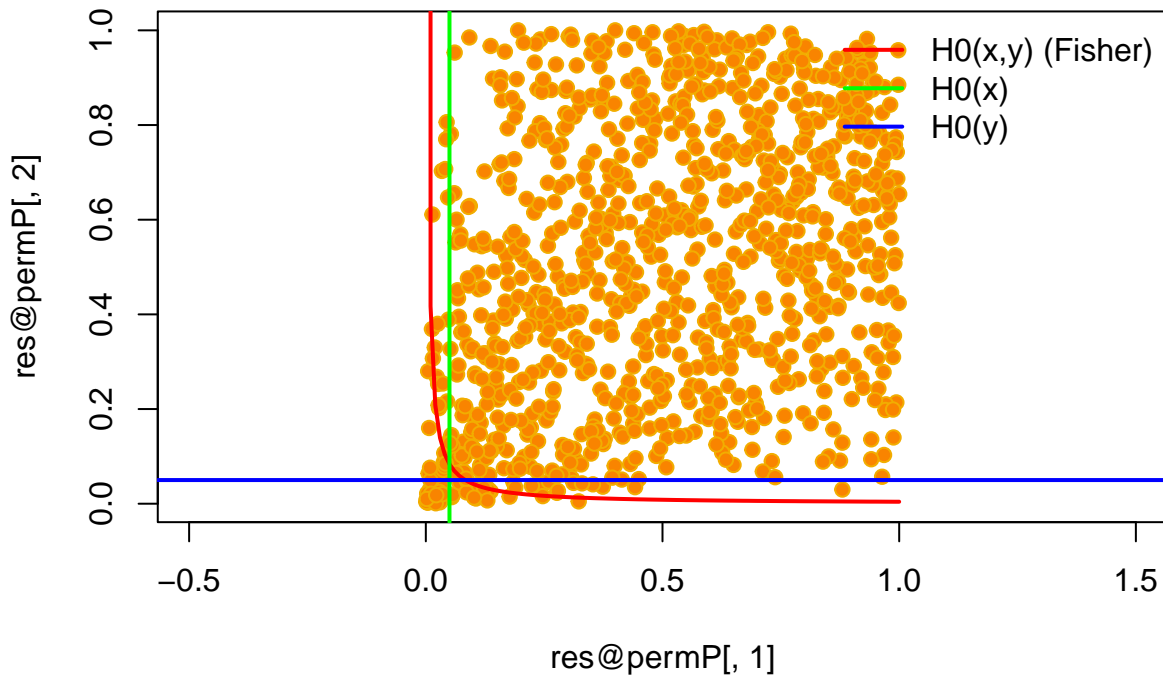
Joint Distribution of p-values



4.3.1 Fisher Combining Function

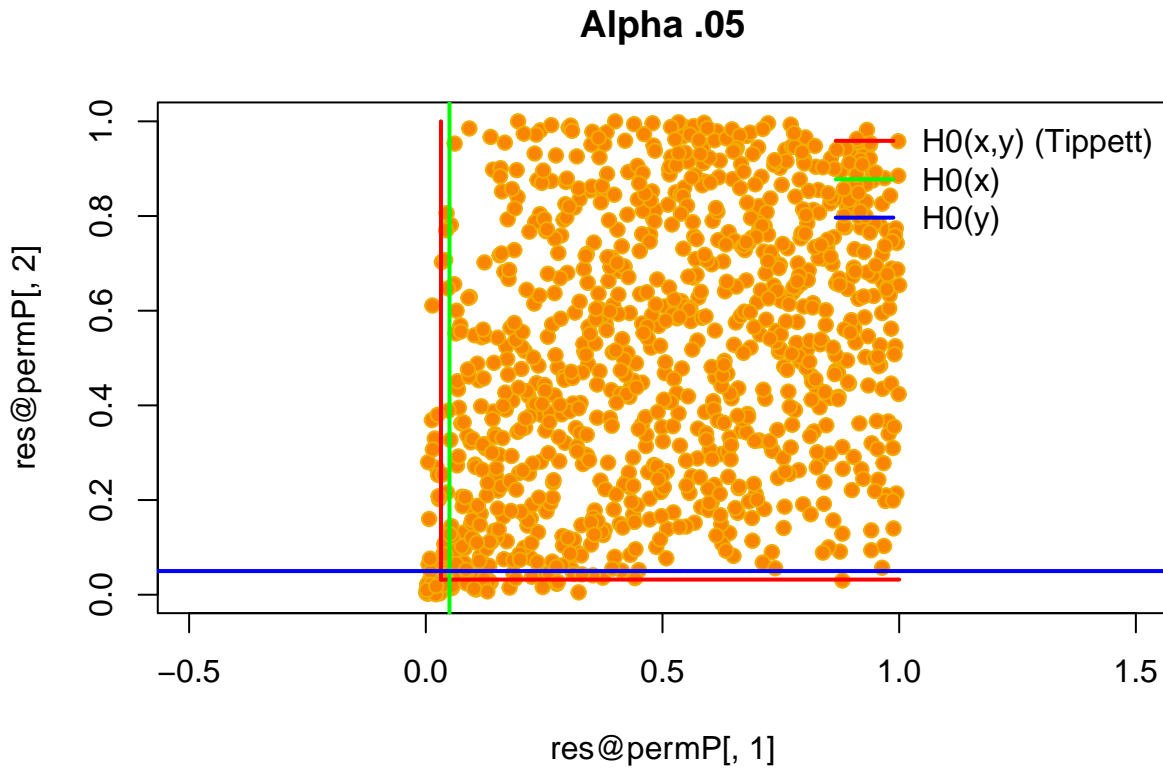
We inspect the rejection regions of the two univariate tests and the one of Fisher combination. The intersection of each univariate test with the Fisher region defines the rejection region of a closed testing - i.e. adjusted for multiple testing.

Alpha .05



4.3.2 Tippett (min-p) Combining Function

We inspect the rejection regions of the two univariate tests and the one of Fisher combination. The intersection of each univariate test with the Fisher region defines the rejection region of a closed testing - i.e. adjusted for multiple testing. This fall to be the same rejection region given by Wesfall & Young. Indeed, it is a closed testing with shortcut.



5 FWER control via Permutations tests

5.1 Permutation Bonferroni

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

5.2 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

5.3 Multiple testing using permutations

The single step min-P method

- Calculate the smallest p-value m for the real data
- Randomly permute the data
- Calculate new p-values for all tests based on permuted data
- Calculate the smallest p-value m^π for permuted data
- Repeat permutation many (say $k=1000$) times: m_1^π, \dots, m_k^π
- 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 $\tilde{\alpha}$

5.4 Correlation structure of p-values

Permutation

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

Consequence

- P-values of correlated tests (i.e. data) remain correlated in permutations
- Distribution of minimum p-value correctly takes correlations into account

When the gain relative to Bonferroni is the gain large?

- Negatively correlated p-values: typically no gain
- Independent p-values: minimal gain
- Positively correlated p-values: gain can be large

5.5 Improved Holm: Westfall & Young

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

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

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

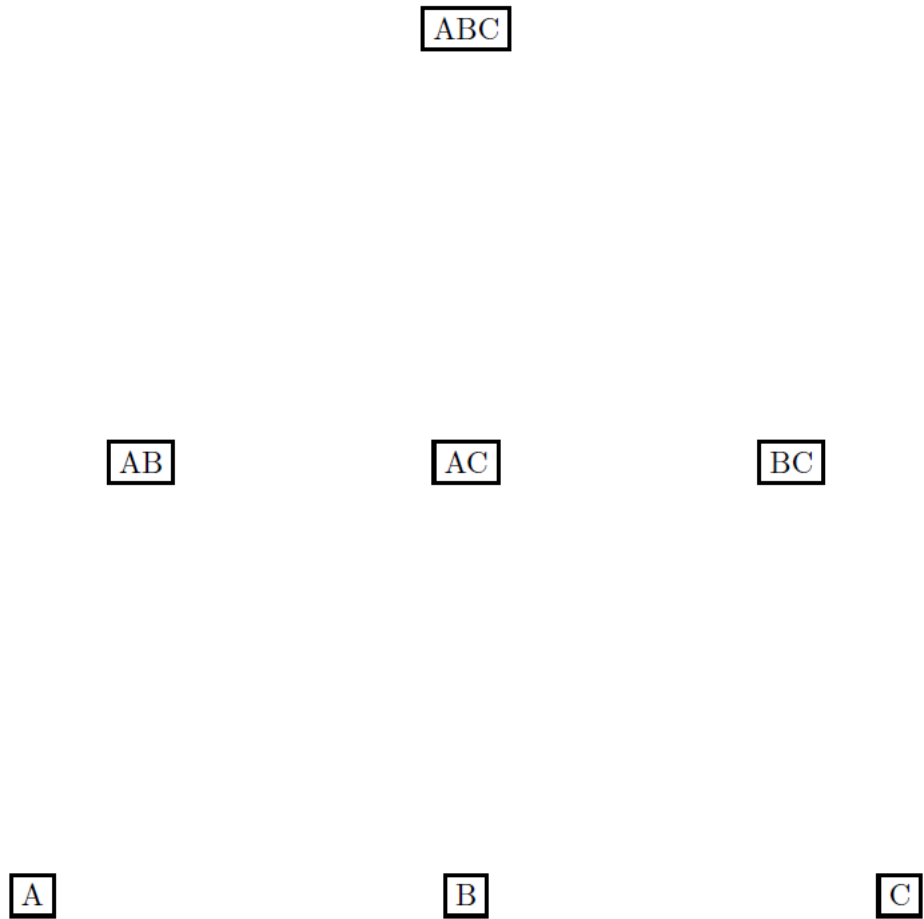
5.6 The general framework: Closed Testing

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

Test in each node: any multivariate permutation test

Westfall & Young is a special case of closed testing (i.e. each node of the closure set uses min-p/Tippett or max-T combining function)

5.6.1 Closure Set



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

In our data:

```
(res=flip.adjust(res,method = "Fisher"))
```

```
##
##   Test  Stat tail p-value Adjust:Fisher
## x    t 1.320   ><  0.1850           0.1850
## y    t 2.061   ><  0.0460           0.0750
```

```
(res=flip.adjust(res,method = "maxT"))
```

```
##
##   Test  Stat tail p-value Adjust:Fisher Adjust:maxT
## x    t 1.320   ><  0.1850           0.1850        0.1850
## y    t 2.061   ><  0.0460           0.0750        0.0640
```

5.7 Conclusion

Accounting for dependencies

Adjusted p-value become 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)

Real data

The variables of real data sets are often correlated

then permutations are (often) convenient

How? R: `library(flip); flip(); flip.adjust()`

6 A case study: Pharmacokinetic Study of Carbidopa

Description:

<http://webserv.jcu.edu/math//faculty/TShort/Bradstreet/part2/part2-table6.html>

As part of a pharmacokinetic study, 12 healthy male subjects were allocated randomly to a three period crossover design receiving one of three graded doses (25, 50, 100 mg) of Carbidopa q8h in each treatment period. A seven day washout period separated the treatment periods. The pharmacokinetic variables AUC, Cmax, and Tmax were calculated for each subject from plasma concentrations assayed from blood samples taken at 0, 0.5, 1, 1.5, 2, 3, 4, 5, 6, 7, and 8 hours postdosing following the second dose of carbidopa on the sixth day of each treatment period.

dataset:

<http://webserv.jcu.edu/math//faculty/TShort/Bradstreet/part2/Bradp2t6.txt>

Analyze the dataset without taking in account the Study Periods (which have been randomized in each subject, hence we can avoid to account for it in the analysis).

Research questions:

- Is there a dose response for AUC, Cmax, or Tmax? Overall?
- Can dose proportionality be established? (try to fit a linear model for each endpoint, then discuss the results)

6.1 A solution

We answer to both first and second question with a single analysis: we perform a linear model (accounting for individual variability) on log transformed end-points.

#Reading and make-up of the data

```
dati=read.table("http://webserv.jcu.edu/math//faculty/TShort/Bradstreet/part2/Bradp2t6.txt",skip = 1,header = 1)

dati=cbind(dati[,1],matrix(as.matrix(dati[,2:4]),nrow(dati)*3,4))
colnames(dati)=c("Sub","Dose","AUC","Cmax","Tmax")

dati=as.data.frame(dati)
str(dati)
```

```
## 'data.frame':   36 obs. of  5 variables:
```

```
## $ Sub : num 1 2 3 4 5 6 7 8 9 10 ...
## $ Dose: num 100 25 50 50 50 25 100 25 50 25 ...
## $ AUC : num 604 140 386 175 605 ...
## $ Cmax: num 137 44.4 86.6 46.4 194 44.9 318 29 119 58.4 ...
## $ Tmax: num 1.5 1 1.5 1.5 0.5 1 1 1 2 2 ...
```

```
# transform all responses with log-transformed,
# so that a linear relationship between time and end-point indicates proportionality
dati[,3:5]=log(dati[,3:5])
```

```
#Descriptives and plots:
summary(dati[, -1])
```

```
##      Dose      AUC      Cmax      Tmax
## Min.   : 25.00  Min.   :4.337  Min.   :3.219  Min.   : -0.6931
## 1st Qu.: 25.00  1st Qu.:5.156  1st Qu.:3.966  1st Qu.: 0.0000
## Median : 50.00  Median :5.886  Median :4.485  Median : 0.2027
## Mean   : 58.33  Mean   :5.873  Mean   :4.547  Mean   : 0.2474
## 3rd Qu.:100.00  3rd Qu.:6.539  3rd Qu.:5.280  3rd Qu.: 0.6931
## Max.   :100.00  Max.   :7.335  Max.   :5.989  Max.   : 1.0986
```

```
by(dati[,3:5], dati$Dose, summary)
```

```
## dati$Dose: 25
##      AUC      Cmax      Tmax
## Min.   :4.337  Min.   :3.219  Min.   : -0.6931
## 1st Qu.:4.803  1st Qu.:3.390  1st Qu.: 0.0000
## Median :4.972  Median :3.801  Median : 0.0000
## Mean   :5.051  Mean   :3.783  Mean   : 0.2071
## 3rd Qu.:5.289  3rd Qu.:4.022  3rd Qu.: 0.6931
## Max.   :5.818  Max.   :4.464  Max.   : 0.6931
## -----
## dati$Dose: 50
##      AUC      Cmax      Tmax
## Min.   :5.133  Min.   :3.837  Min.   : -0.6931
## 1st Qu.:5.670  1st Qu.:4.374  1st Qu.: 0.0000
## Median :5.886  Median :4.484  Median : 0.2027
## Mean   :5.815  Mean   :4.479  Mean   : 0.1689
## 3rd Qu.:5.967  3rd Qu.:4.625  3rd Qu.: 0.4055
## Max.   :6.405  Max.   :5.268  Max.   : 1.0986
## -----
## dati$Dose: 100
##      AUC      Cmax      Tmax
## Min.   :6.164  Min.   :4.920  Min.   :0.0000
## 1st Qu.:6.607  1st Qu.:5.229  1st Qu.:0.0000
## Median :6.782  Median :5.412  Median :0.4055
## Mean   :6.751  Mean   :5.378  Mean   :0.3662
## 3rd Qu.:6.922  3rd Qu.:5.515  3rd Qu.:0.6931
## Max.   :7.335  Max.   :5.989  Max.   :1.0986
```

```

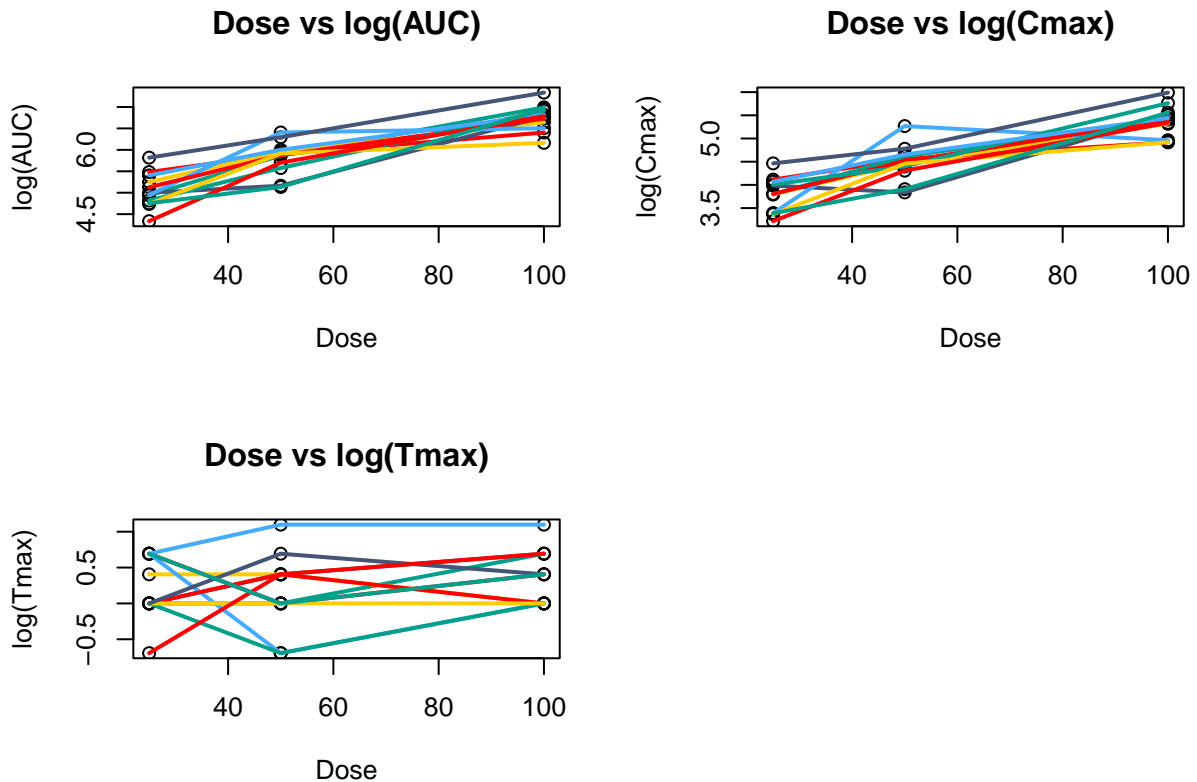
par(mfrow=c(2,2))
plot(dati$Dose,dati$AUC,ylab="log(AUC)",xlab="Dose",main="Dose vs log(AUC)")

r=sapply(unique(dati$Sub),function(s){
  d=subset(dati,Sub==s)
  d=d[order(d$Dose),]
  lines(d$Dose,(d$AUC),col=s,lwd=2)})

plot(dati$Dose,dati$Cmax,ylab="log(Cmax)",xlab="Dose",main="Dose vs log(Cmax)")
r=sapply(unique(dati$Sub),function(s){
  d=subset(dati,Sub==s)
  d=d[order(d$Dose),]
  lines(d$Dose,(d$Cmax),col=s,lwd=2)})

plot(dati$Dose,dati$Tmax,ylab="log(Tmax)",xlab="Dose",main="Dose vs log(Tmax)")
r=sapply(unique(dati$Sub),function(s){
  d=subset(dati,Sub==s)
  d=d[order(d$Dose),]
  lines(d$Dose,(d$Tmax),col=s,lwd=2)})

```



Now the analysis: A simple solution could be:

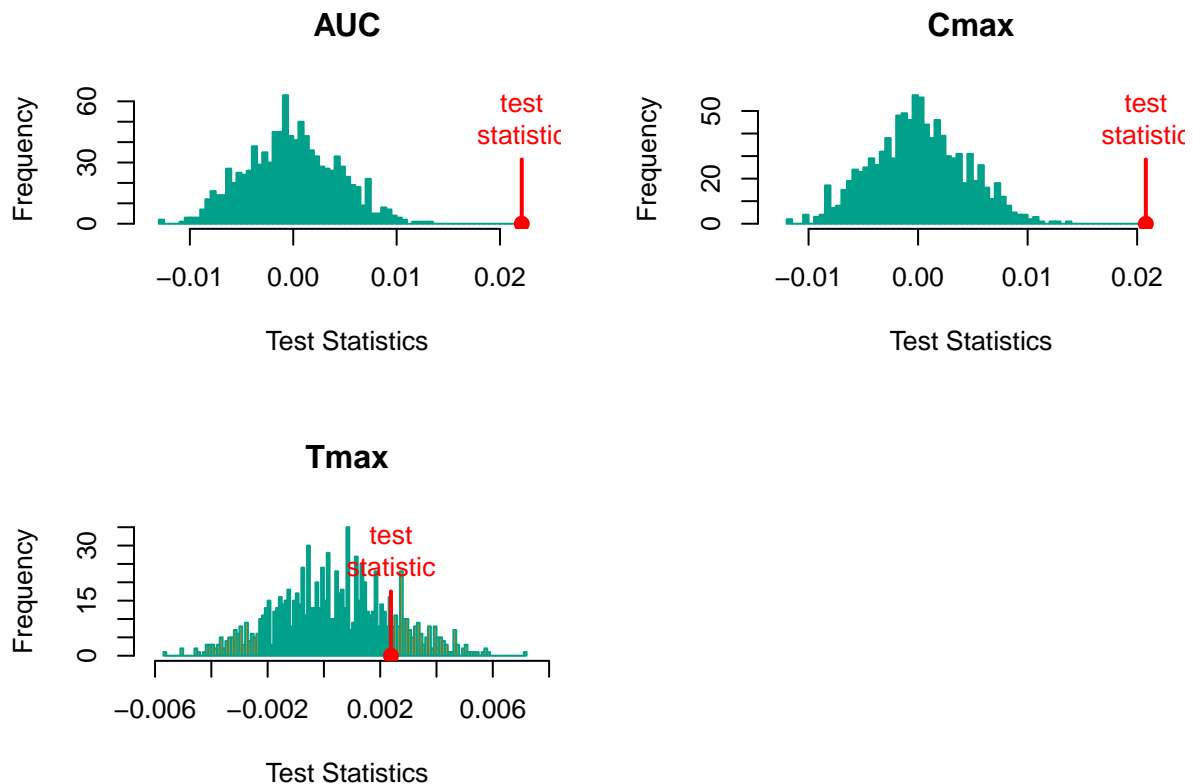
```

library(flip)
res=flip(.-Dose,data=dati,Strata=~Sub,statTest = "coeff")
summary(res)

```

```
## Call:
## flip(Y = . ~ Dose, data = dati, statTest = "coeff", Strata = ~Sub)
## 999 permutations.
##
##      Test   Stat tail p-value sig.
## AUC  coeff 0.0221  ><  0.0010 ***
## Cmax coeff 0.0208  ><  0.0010 ***
## Tmax coeff 0.0024  ><  0.2710
```

#here we ask for statTest = "coeff", i.e. estimated coefficient of a linear model
`hist(res)`



Multivariate:

- Overall

```
res=flip.adjust(res)
npc(res,"Fisher")
```

```
##
##   comb.funct nVar  Stat p-value
## V1      Fisher    3 15.12 0.0010
```

There is an effect of Dose, overall.

- By end-points (closed testing with max-t combining function). Try also different methods (e.g. `method="Fisher"`) and compare the results of `method="minP"` with the one of `method="Holm"`.

```
res=flip.adjust(res,method="holm")
res=flip.adjust(res,method="Fisher")
summary(res)
```

```
## Call:
## flip(Y = . ~ Dose, data = dati, statTest = "coeff", Strata = ~Sub)
## 999 permutations.
##
##      Test   Stat tail p-value Adjust:maxT Adjust:holm Adjust:Fisher sig.
## AUC  coeff 0.0221  ><  0.0010      0.0010      0.0030      0.0010 ***
## Cmax coeff 0.0208  ><  0.0010      0.0010      0.0030      0.0010 ***
## Tmax coeff 0.0024  ><  0.2710      0.2710      0.2710      0.2710
```

AUC and Cmax show a significant effect after correction for multiplicity, while Tmax does not.

7 (minimal) Bibliography

The Grounding Theory:

- Pesarin (2001) Multivariate Permutation Tests: With Applications in Biostatistics by Fortunato, Wiley, New York

An alternative approach to the Permutation testing:

- Hemerik J, Goeman J. Exact testing with random permutations. *Test (Madr)*. 2018;27(4):811-825. doi: 10.1007/s11749-017-0571-1. Epub 2017 Nov 30. PMID: 30930620; PMCID: PMC6405018.

A flexible approach to General Linear Model based on the sign-flip score test:

- Hemerik, Goeman and Finos (2020) Robust testing in generalized linear models by sign flipping score contributions. *Journal of the Royal Statistical Society Series B (Statistical Methodology)* 82(3). DOI: 10.1111/rssb.12369

Implemented in R package *flipscores*:

<https://cran.r-project.org/web/packages/flipscores/index.html>

better to use the github develop version:

<https://github.com/livioivil/flipscores>

A nice review of the regression model within the permutation framework:

- Anderson M. Winkler, Gerard R. Ridgway, Matthew A. Webster, Stephen M. Smith, Thomas E. Nichols (2014) Permutation inference for the general linear model, *NeuroImage*, Volume 92, Pages 381-397, ISSN 1053-8119 <https://doi.org/10.1016/j.neuroimage.2014.01.060>