Permutation Tests

Livio Finos

University of Padova

Contents

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 actives 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.

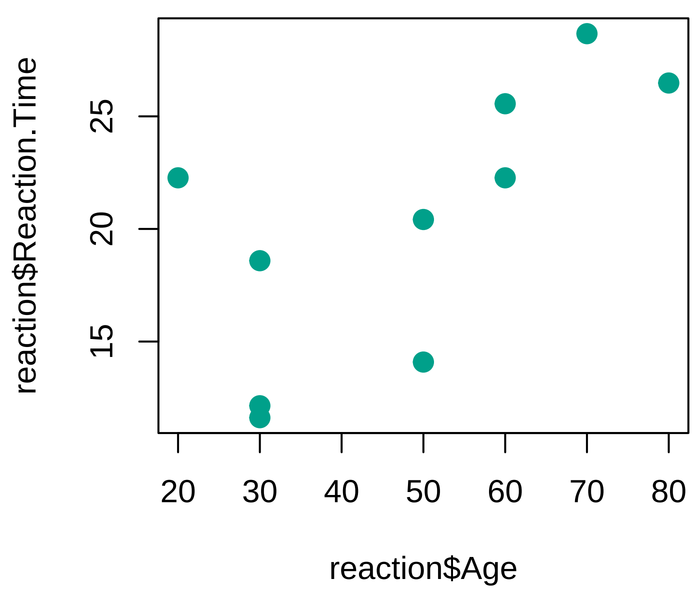
(data are fictitious)

To read the data

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

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
- $\sigma_{xy} >> (<<) 0$: 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

$-\sigma_x\sigma_y \leq \sigma_{xy} \leq \sigma_x\sigma_y$ is quivalent 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.

$E(Reaction.Time) \approx \beta_0 + \beta_1 Age$
$E(Y) = \beta_0 + \beta_1 X$

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): $y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) = y_i - \hat{y}_i$

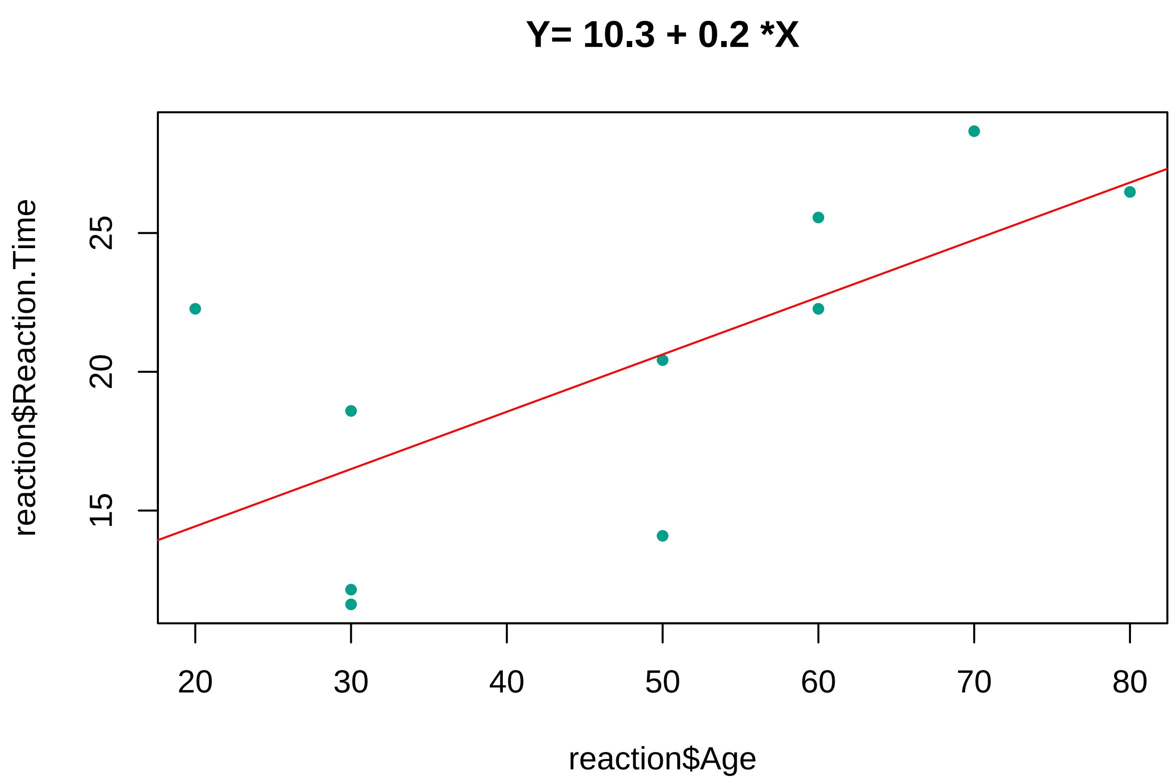
and therefore the least squares are the sum of the squared residuals: $\sum_{i=1}^{n}(y_i - \hat{\beta}_0 + \hat{\beta}_1 x_i)^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

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?

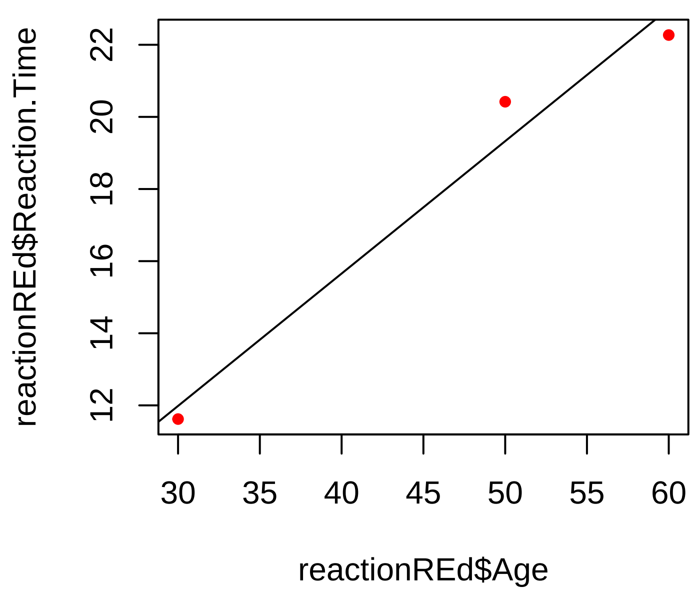
- **Null Hypothesis** $H_0: \ \beta_1 = 0$ (the **true** β_1, not its estimate $\hat{\beta}_1$!). There is no relationship between X and Y.
- **Alternative Hypothesis** $H_1: \ \beta_1 > 0$ The relationship is positive.

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
```

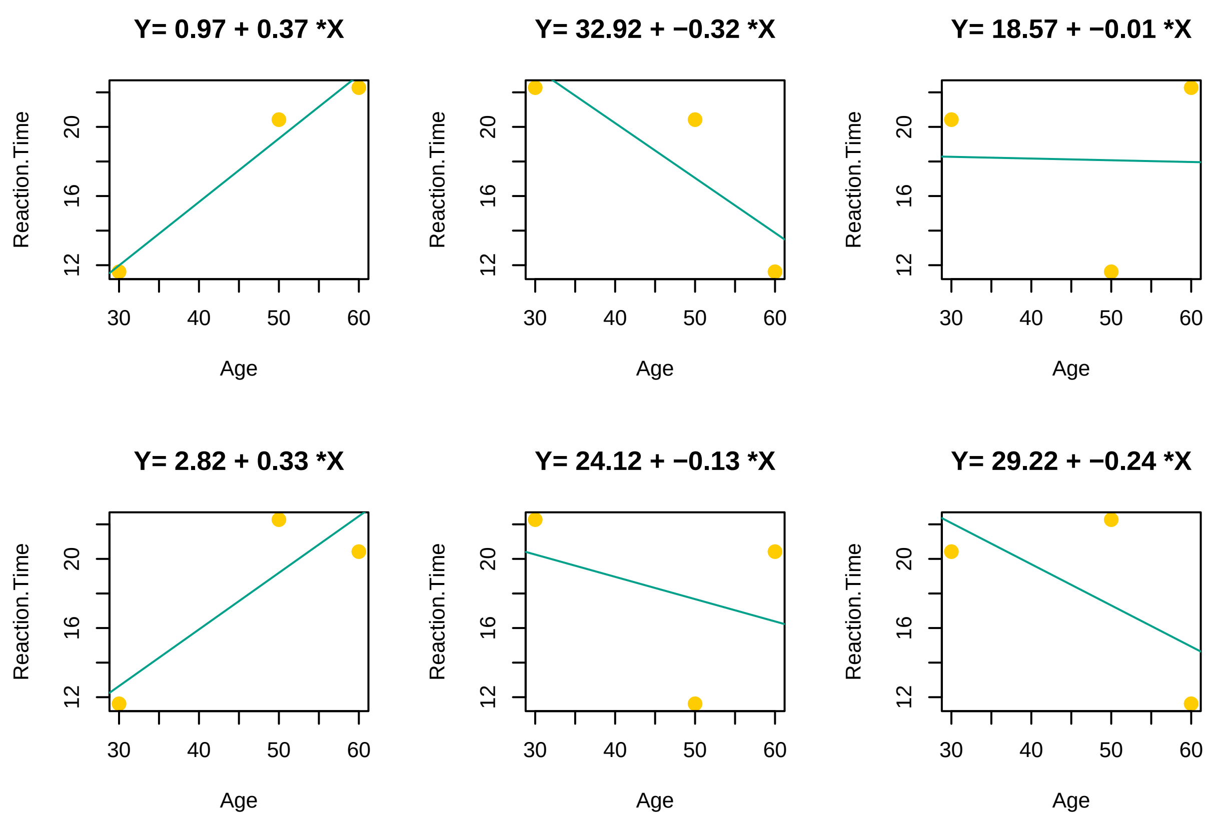


- *If* H_0 is true: there is no linear relationship between X and Y
- Therefore, the trend observed on the data is due to chance.

- 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 Yobserved? $3*2*1 = 3! = 6$ 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 $f(y_1, \ldots, y_n)$ does not change when the ordering of $y_1, \ldots, y_n$ 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…

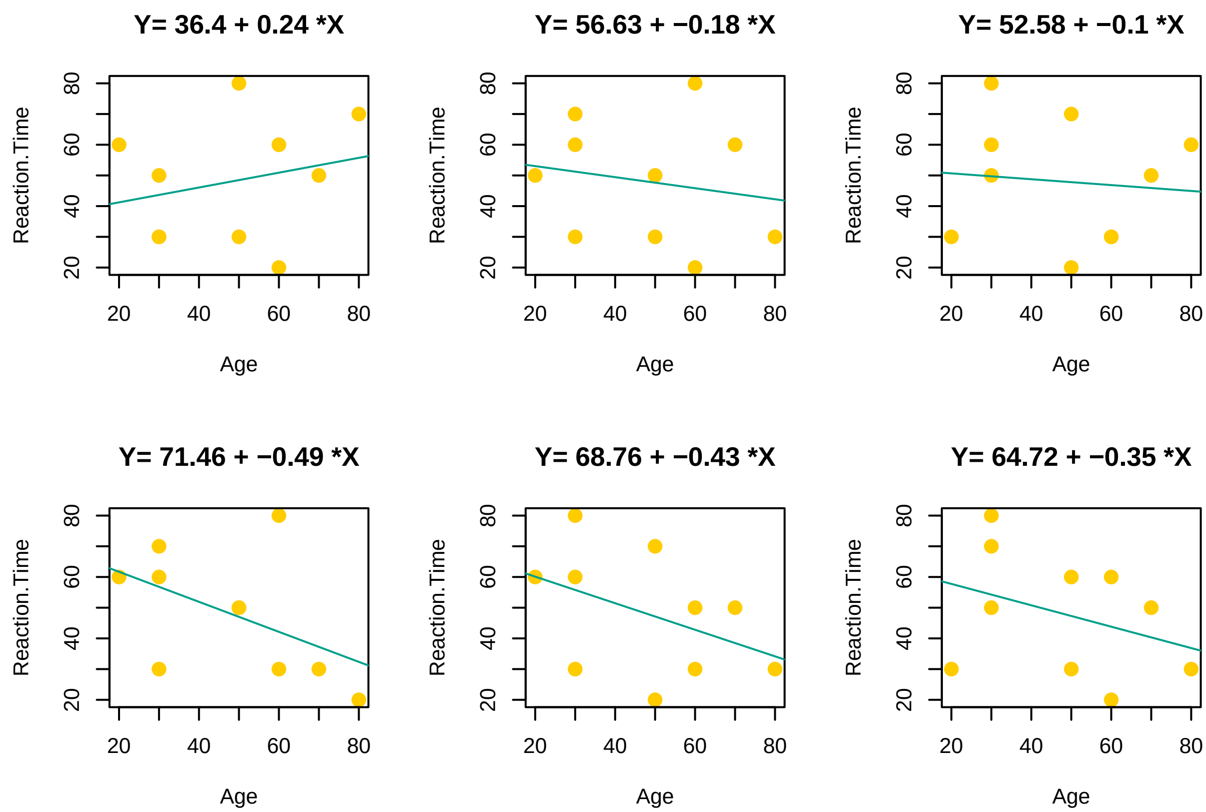
How many permutations of the vector $y_1, \ldots, y_n$ are possible? $n! = 10! = 3628800$.

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.

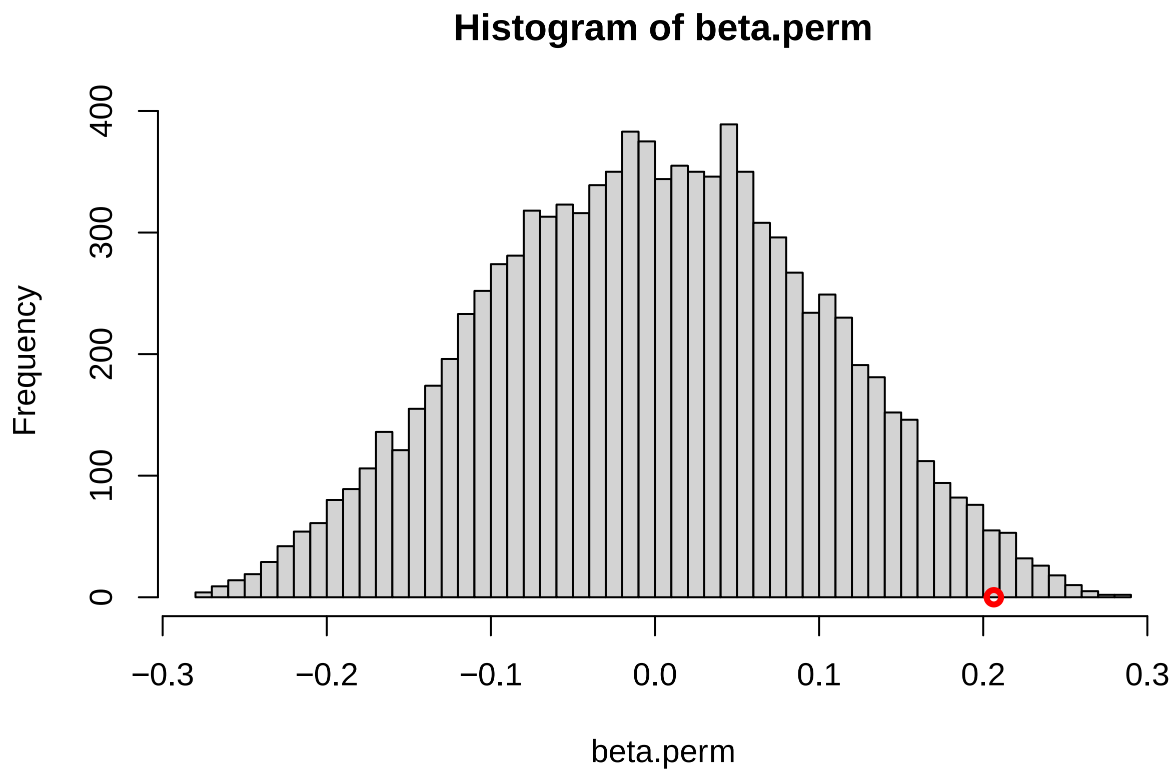
here some example

`Age` vs a permutations of `Reaction.Time`



We repeat 10^4 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=10^4 permutations we got 9830 times a $\hat{\beta}_1^{*b} \leq \hat{\beta}_1^{obs}$.

The p-value (significance) is $p = \frac{\#(\hat{\beta}_1^{*b} \geq \hat{\beta}_1^{obs})}{B} = 0.0172$

($\hat{\beta}_1^{obs}$ 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 $p = 0.0172$, 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** ($T: \mathbb{R}^n \to \mathbb{R}$) 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: $T = \hat{\beta}_1 = \hat{\sigma}_{xy}/\hat{\sigma}_{yy}$ 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:

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

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

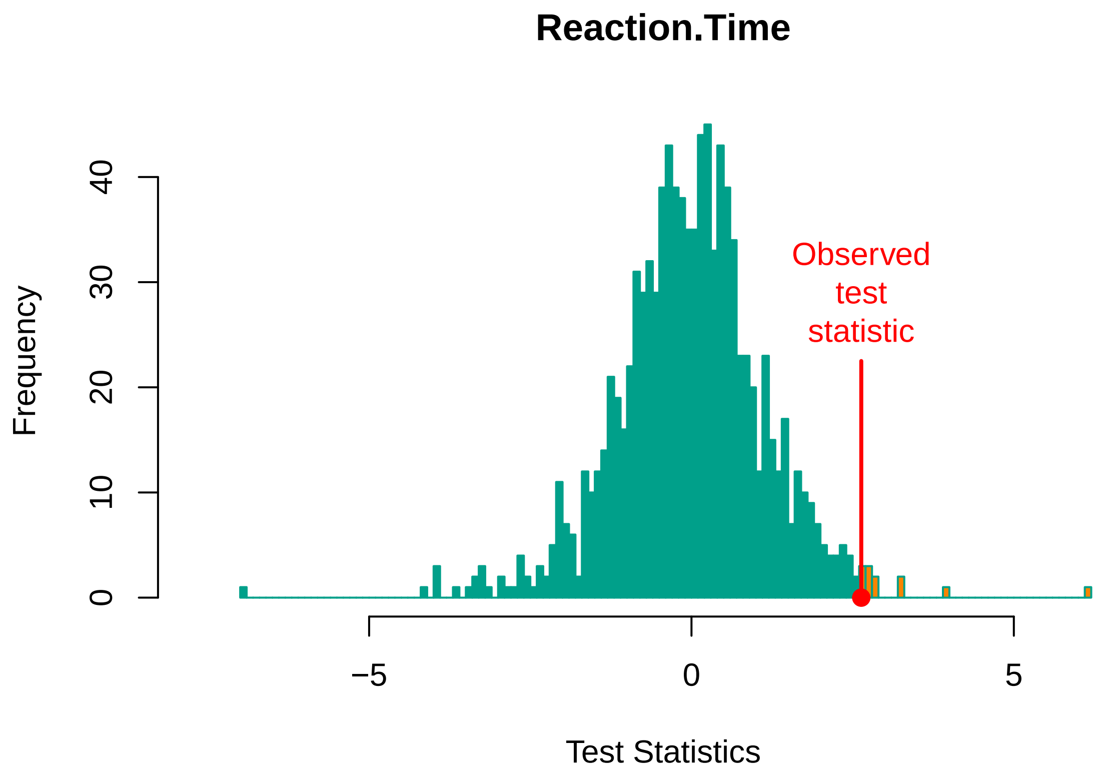
2.2.1 We make it in flip

```
library(flip)

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

```
##
##                Test  Stat tail p-value
## Reaction.Time    t 2.633    >  0.0120
```

```
## 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

$p = \frac{\#(|\hat{\beta}_1^{*b}| \geq |\hat{\beta}_1^{obs}|)}{B} = 0.0374$

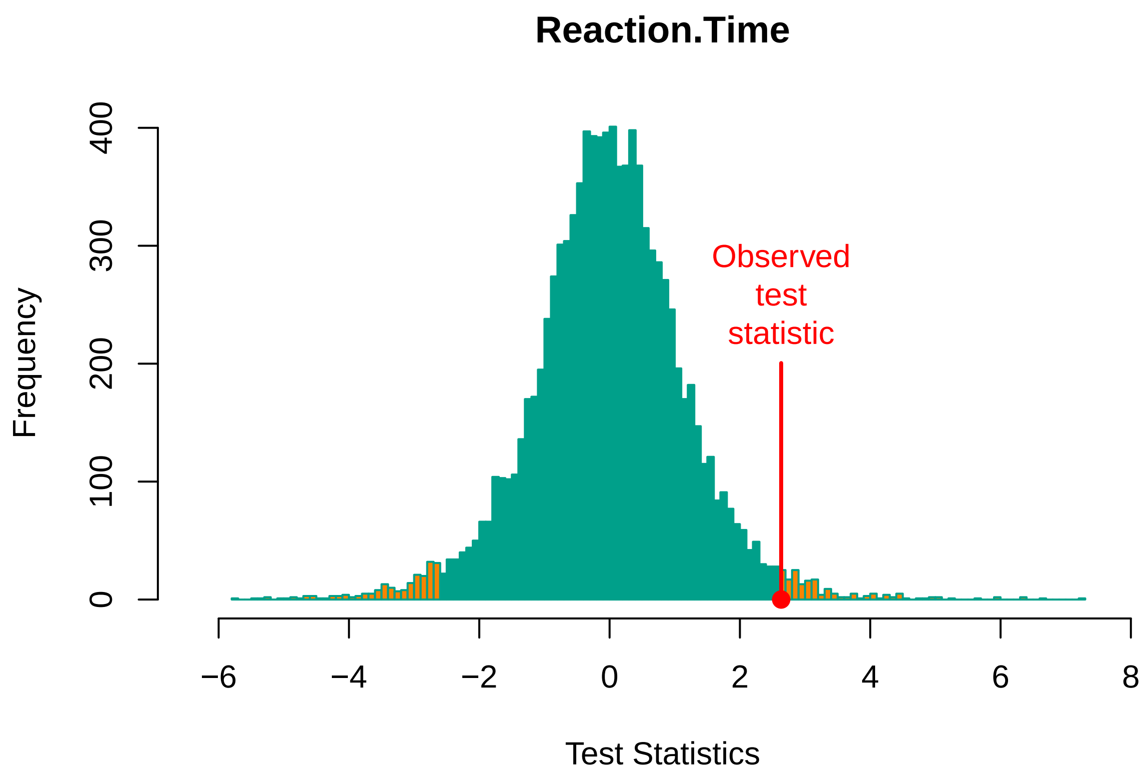
(remark: the observed test stat is included among the permuted one)

In `flip`:

```
library(flip)
(res=flip(Reaction.Time~Age,data=reaction,tail=0,perms=10000))
```

```
##
##                 Test  Stat tail p-value
## Reaction.Time      t 2.633   ><  0.0369
```

```
plot(res)
```



2.3 A more formal approach

(see also Pesarin, 2001; Hemerik & Goeman, 2017)

Let Y be data taking values in a sample space $\mathcal{Y}$. Let Π be a finite set of transformations $\pi : \mathcal{Y} \to \mathcal{Y}$, 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
$H_0 : Y \in \Omega_0$

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 $T(Y): \mathbb{R}^n \to \mathbb{R}$

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

Define the test:

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

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

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.)

More about permutation testing

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 tye I error, while permutation approach provides exactness.

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

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

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

(due to group structure)

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

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

2.3.1 Properties (see Pesarin, 2001)

The theorem above proves that the permutation tests have **exact control of the type I error**, i.e. $P(p-value \leq \alpha|H_0) = \alpha$ assuming $\alpha \in \{1/|\mathcal{O}|, 2/|\mathcal{O}|, \ldots, 1\}$ - 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).

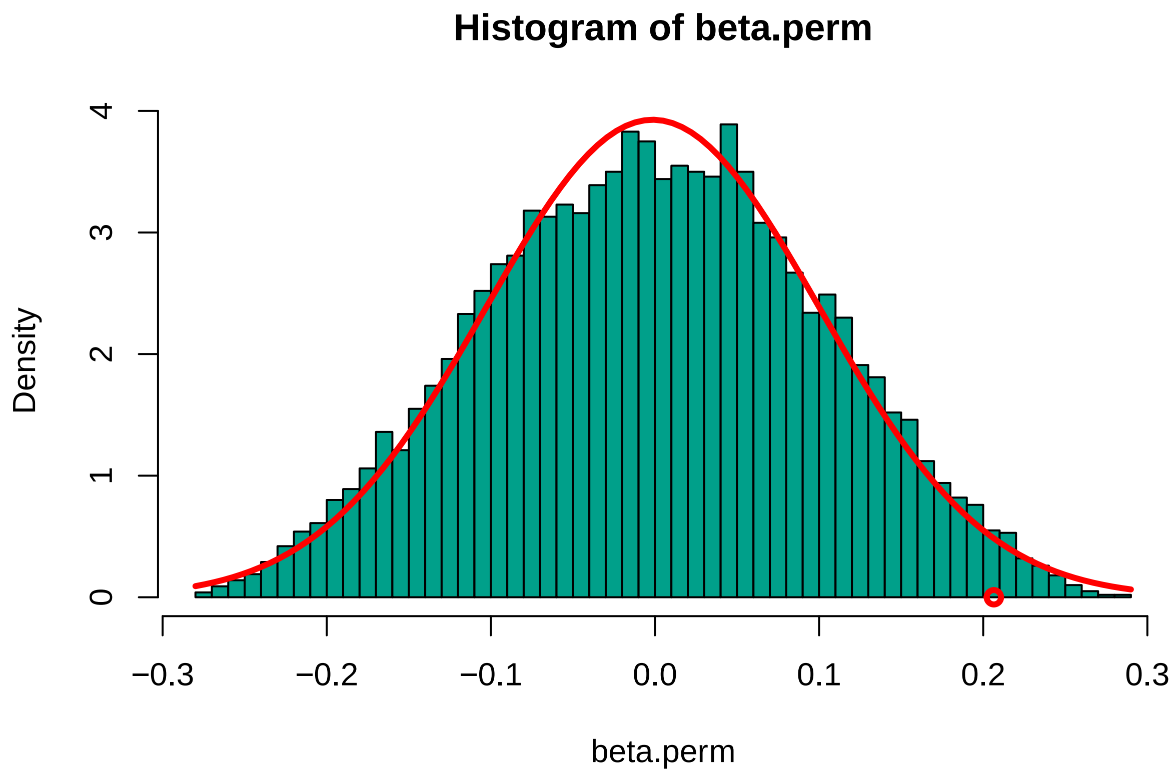
Further properties:

- The permutations tests are **Unbiased**: $P(p-value \leq \alpha|H_1) > \alpha$
- The test is **Consistent**: $P(p-value \leq \alpha|H_1) \to 1$ when $n \to \infty$
- 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:

$Y =$ linear part $+$ normal error

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

Assumptions of the linear model

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

$t = \frac{\hat{\beta_1}}{std.dev\ \hat{\beta_1}} = \frac{\hat{\beta_1}}{\sqrt{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 / \sum (x_i - \bar{x})^2 / (n-2)}}$

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

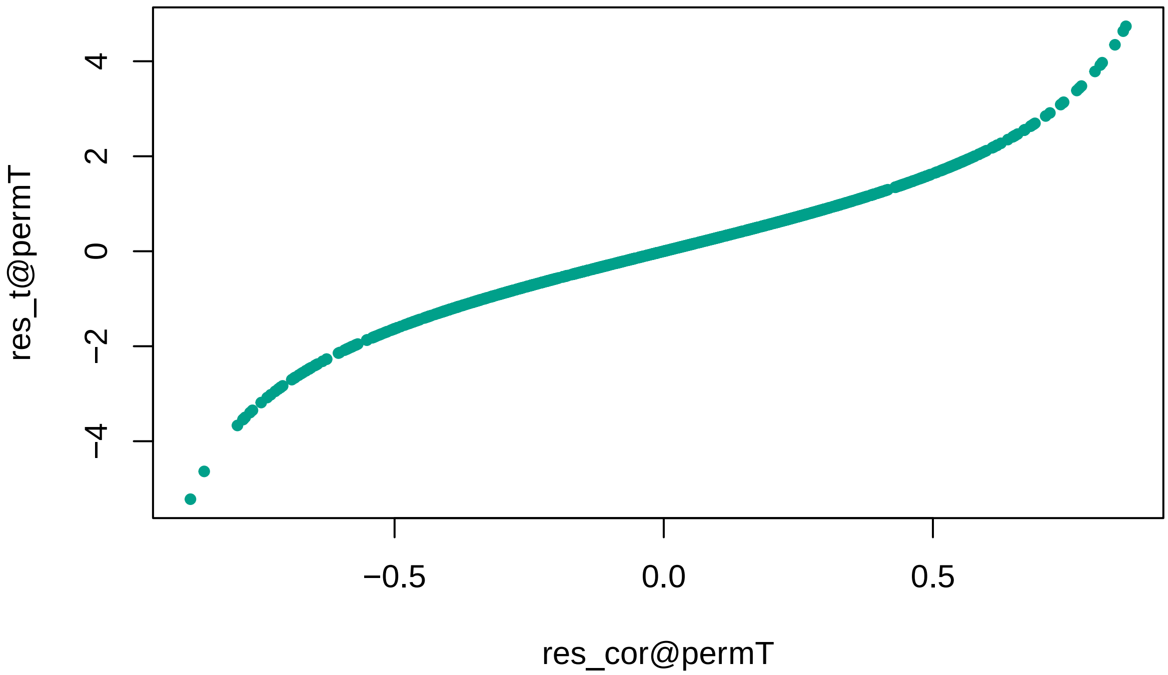
```
set.seed(1)
(res_cor=flip(Reaction.Time~Age,data=reaction,statTest = "cor"))
```

```
##
##               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 do 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 = 10000,statTest  = "rank"))
```

```
##
##                     Test  Stat tail p-value
## Reaction.Time Wilcoxon 2.179   ><  0.0210
```

```
# to see the rank correlation use the workaround:
(res=flip(rank(reaction$Reaction.Time)~rank(reaction$Age),perms = 10000,statTest  = "cor"))
```

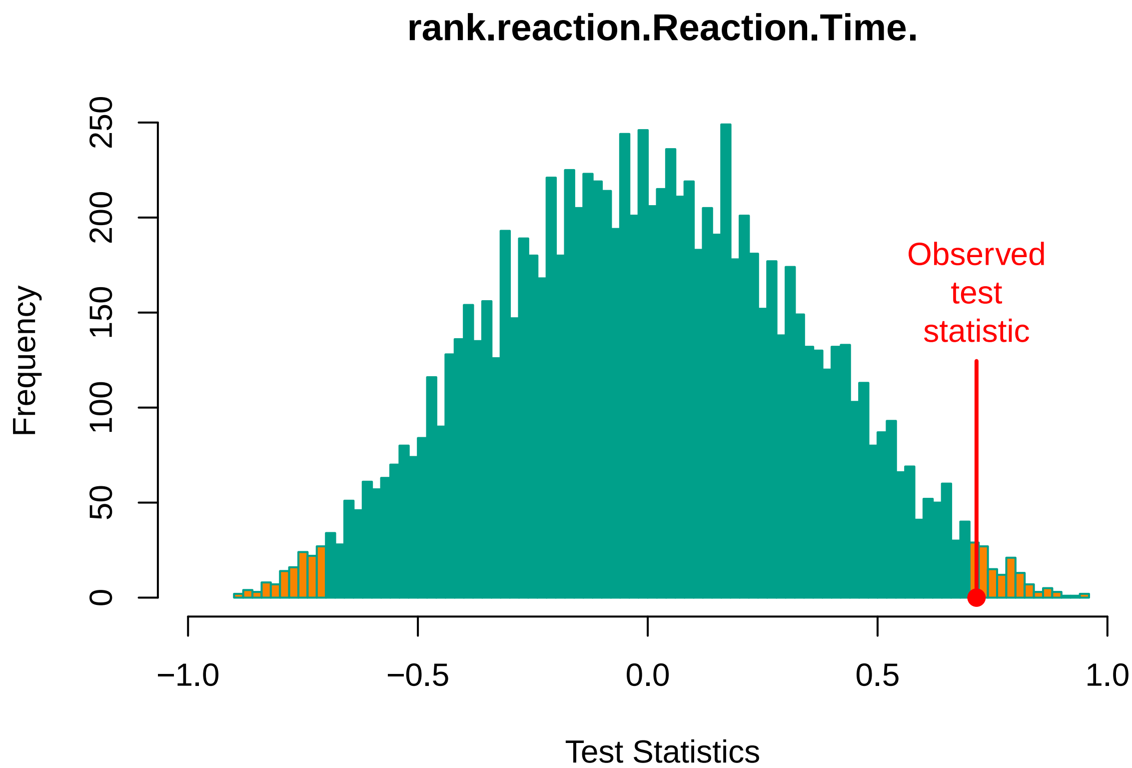
```
##
##                                 Test   Stat tail p-value
## rank.reaction.Reaction.Time.  cor 0.7153   ><  0.0221
```

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

```
## Warning in cor.test.default(reaction$Reaction.Time, reaction$Age, method =
## "spe"): Cannot compute exact p-value with 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 do 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

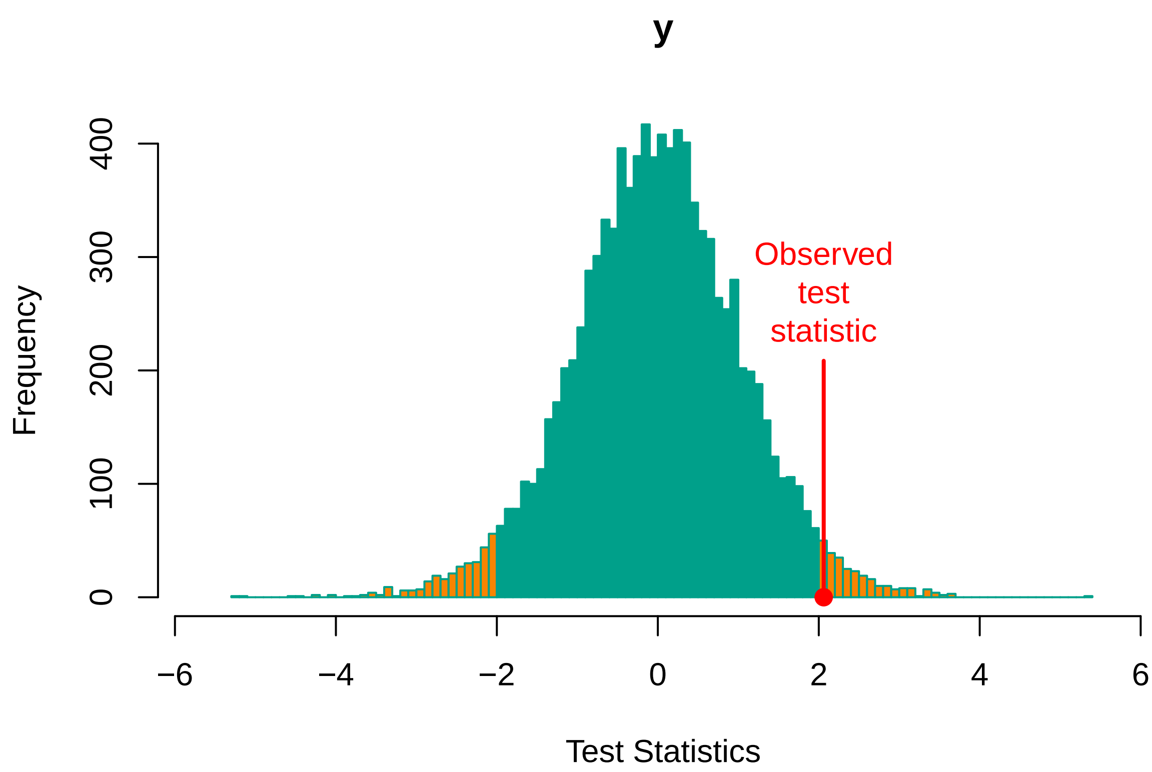
```r
data("seeds")
seeds=na.omit(seeds)

(res=flip(y~grp,data=seeds,perms = 10000))
```

```
##
##   Test  Stat tail p-value
## y    t 2.061   ><  0.0511
```

```r
(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
```

```r
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=10000))
```

```
##
##       Test Stat tail p-value
## y Wilcoxon 2.13   ><  0.0317
```

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

```
## Warning in wilcox.test.default(x = c(12.54, 14.81, 16.71, 7.53, 7.02, 8.09, :
## cannot compute exact p-value with 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 cathegorical 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
```

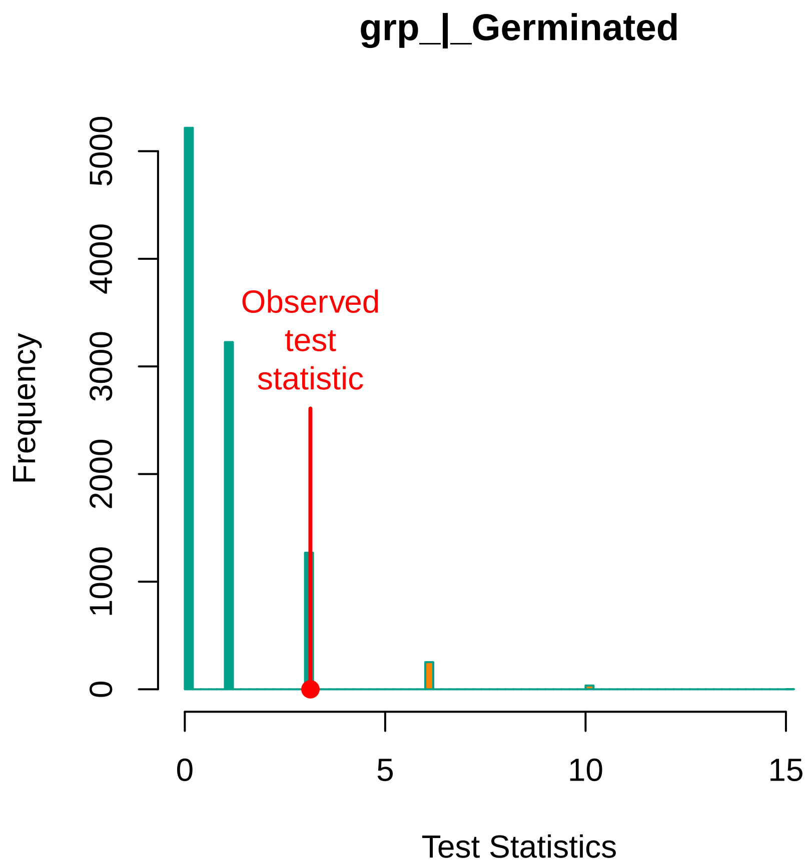
```
chisq.test(seeds$grp,seeds$Germinated)
```

```
##
##  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=10000))
```

```
##
##                           Test  Stat tail p-value
## grp_|_Germinated Chi Squared 3.135    >  0.1557
```

```
plot(res)
```



… and the Fisher test:

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

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

```
(flip(Germinated~grp,data=seeds,perms=10000))
```

```
##
##                    Test   Stat tail p-value
## GerminatedFALSE       t -1.798   ><  0.1542
## GerminatedTRUE        t  1.798   ><  0.1542
```

3.4 ANOVA (C-sample)

e.g. 3 groups of `Age`: young [18-35), middle age [35-60), old [60-100)] lines(d$Dose$, (dAUC),col=s,lwd=2)})] lines(d$Dose$, (dCmax),col=s,lwd=2)})] lines(d$Dose$, (dTmax),col=s,lwd=2)})

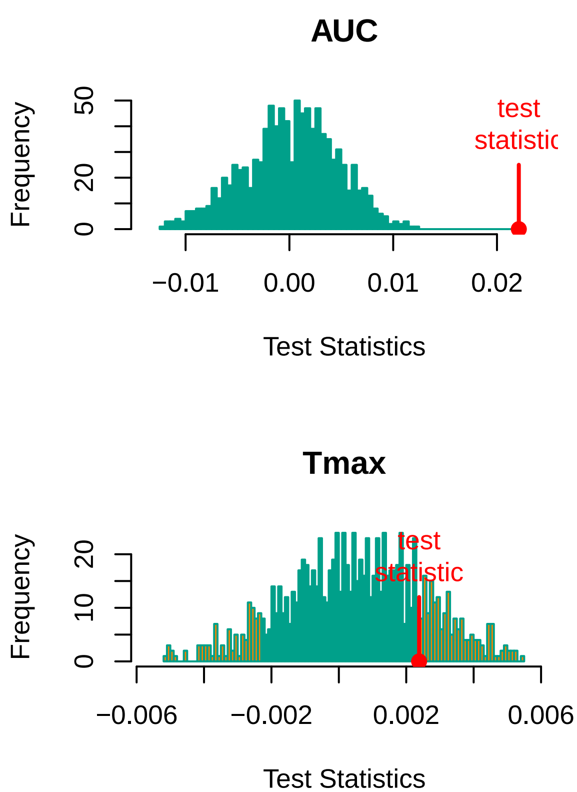
\begin{center}\includegraphics{perm_files/figure-latex/unnamed-chunk-41-1} \end{center}

Now the analysis: A simple solution could be:

```
```r
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   ><  0.2680
```

```
#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 15.13  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        0.0030   **
## Cmax coeff 0.0208   ><  0.0010      0.0010      0.0030        0.0020   **
## Tmax coeff 0.0024   ><  0.2680      0.2680      0.2680        0.2680
```

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

4 (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