Multiple testing

I What's a pvalue?

Until now, we have seen various ways to build a test Δ_α of level α.

$\Delta_\alpha \in \{0,1\}$, only depends on the observation

and $\forall \alpha \quad P_{H_0}(\Delta_\alpha = 1) \leq \alpha$

If we assume that Δ_α is monotonous that is

$$\forall \alpha < \alpha', \quad \Delta_\alpha \leq \Delta_{\alpha'}$$

then the pvalue of the test Δ_α is the limit value p which depends on the observations such that

$\forall \alpha < p \qquad \Delta_\alpha = 0 \qquad$ "It accepts H_0 at level α"

$\forall \alpha > p \qquad \Delta_\alpha = 1 \qquad$ "It rejects H_0 ________ "

Classical example

If Δ_α is defined as $\mathbb{1}_{T > c_\alpha}$ where T is the test statistic whose distribution is known under H_0 and $\boxed{\text{continuous}}$, and c_α is the $1-\alpha$ quantile of this dist°

Then $p = 1 - F_{H_0}(T)$ where F_{H_0} cdf of T under H_0

Correction

1/ It's of level α

$$\Delta_\alpha = \mathbb{1}_{T > \underbrace{F^{-1}(1-\alpha)}_{c_\alpha \text{ the } 1-\alpha \text{ quantile of } T \text{ under } H_0}}$$

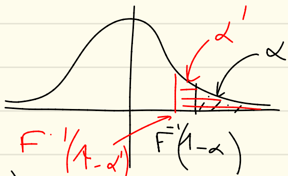
$$P_{H_0}(\Delta_\alpha = 1) = P_{H_0}(T > F^{-1}(1-\alpha)) = 1 - P_{H_0}(T \leq F^{-1}(1-\alpha))$$
$$= 1 - F(F^{-1}(1-\alpha))$$

F cdf is continuous then $F(F^{-1}(t)) = t$

$$P_{H_0}(\Delta_\alpha = 1) = 1 - (1-\alpha) = \alpha \quad \forall \alpha$$

2/ Monotonous

$\alpha < \alpha'$



then $F^{-1}(1-\alpha') < F^{-1}(1-\alpha)$

If $\Delta_\alpha = \mathbb{1}_{T > F^{-1}(1-\alpha)} = 1 \implies \Delta_{\alpha'} = \mathbb{1}_{T > F^{-1}(1-\alpha')} = 1$

$\Delta_\alpha \leq \Delta_{\alpha'}$ in this case

If $\Delta_{\alpha'} = \mathbb{1}_{T > F^{-1}(1-\alpha')} = 0 \implies \Delta_\alpha = \mathbb{1}_{T > F^{-1}(1-\alpha)} = 0$

$\implies \Delta_\alpha \leq \Delta_{\alpha'}$ in all cases

3/ Check $p(T) = 1 - F(T) \iff T = F^{-1}(1 - p(T))$

let $\alpha < p(T) \implies 1 - \alpha > 1 - p(T)$

$$\Delta_\alpha = \mathbb{1}_{T > F^{-1}(1-\alpha)} = \mathbb{1}_{\underbrace{F^{-1}(1-p) > F^{-1}(1-\alpha)}_{\text{this is wrong}}}$$

so $\Delta_\alpha = 0$ (the test accepts)

let $\alpha > p(T) \implies 1 - \alpha < 1 - p(T)$

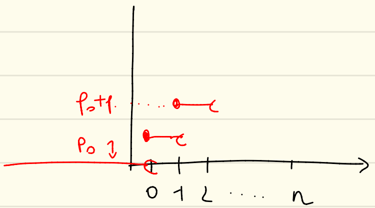
$\implies \Delta_\alpha = 1$ the test rejects

Small complements on cdf

X real variable, the cdf $F(x) = P(X \leq x)$

In general, $F \nearrow$ $\quad F(-\infty) = 0 \quad F(+\infty) = 1$

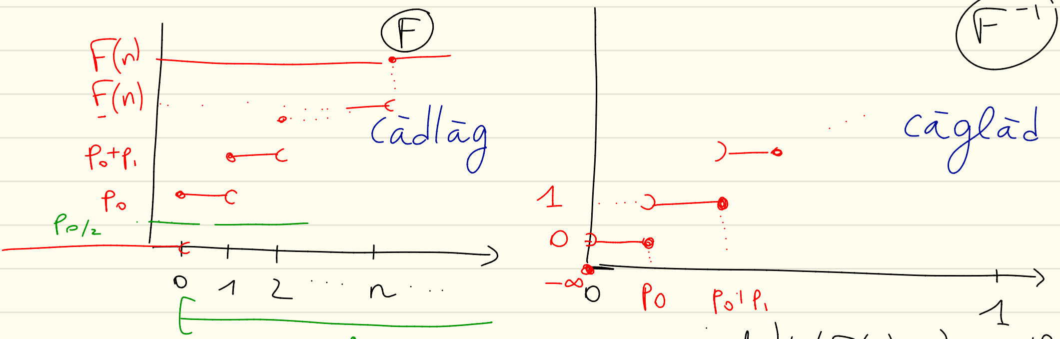
F càdlàg (continu à droite limité à gauche)
continuous on the right limited on the left



ex: Poisson

Official definition of a quantile

$$F^{-1}(u) = \inf \{ t \,/\, F(t) \geq u \}$$



$u = 0 \qquad \inf\{t \,/\, F(t) \geq 0\} = \inf \mathbb{R} = -\infty$

$u = \frac{p_0}{2} \qquad \inf\{t \,/\, F(t) \geq \frac{p_0}{2}\} = 0$

it also works for all $t \in (0, p_0]$

If T is discrete with càdlàg cdf under H_0,

and if $\Delta_\alpha = \mathbb{1}_{T > F^{-1}(1-\alpha)}$

then the pvalue is defined by $p(T) = 1 - \underbrace{F_-(T)}_{\text{limit when coming from the left}}$

Check:

1/ Should be of level α:

$$P_{H_0}(\Delta_\alpha = 1) = P_{H_0}(T > F^{-1}(1-\alpha)) = 1 - P(T \leq F^{-1}(1-\alpha))$$
$$= 1 - F(F^{-1}(1-\alpha))$$

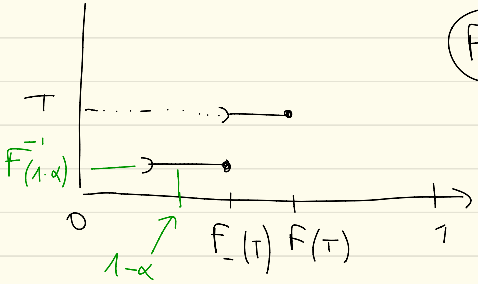
See the Poisson example $F(F^{-1}(0)) = 0$, but $F(F^{-1}(\frac{p_0}{2})) = p_0$!

In general, $\boxed{F(F^{-1}(t)) \geq t}$

So $P_{H_0}(\Delta_\alpha = 1) = 1 - \underbrace{F(F^{-1}(1-\alpha))}_{\geq 1-\alpha} \leq \alpha$

It's monotonous as before

It remains to check that $p(T) = 1 - F_-(T)$ is the limit value to pass from acceptation to rejection

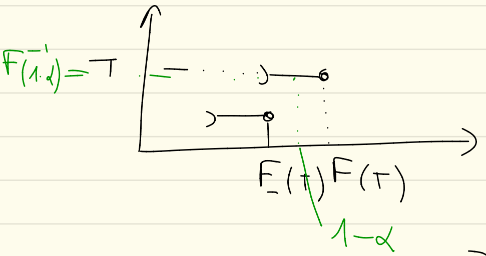


F^{-1}

take $\alpha = p(T) + \varepsilon$ small

$1-\alpha = 1 - p(T) - \varepsilon$
$= F_-(T) - \varepsilon$

So $T > F^{-1}(1-\alpha) \Rightarrow$ the test rejects



take $\alpha = p(T) - \varepsilon$ ($\varepsilon > 0$ small)

$1-\alpha = F_-(T) + \varepsilon$

So $T = F^{-1}(1-\alpha)$ and $T > F^{-1}(1-\alpha)$ is wrong

$\Rightarrow$ the test accepts

On simulations

We want to check the fundamental property of pvalue which is $\boxed{P_{H_0}(p \leq \alpha) \leq \alpha \quad \forall \alpha}$

– Gaussian case

observe $X \sim \mathcal{N}(\mu, 1)$

test H_0 : "$\mu = 0$" vs H_1 : "$\mu \neq 0$"

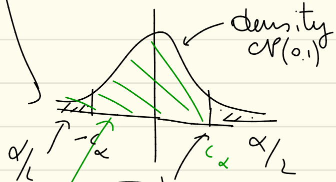
. Q on μ → estimated by X

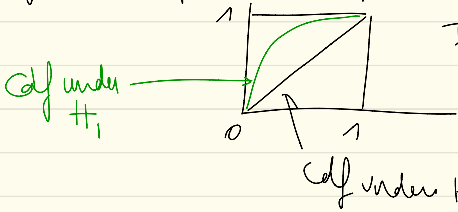
. reject when $|X| \geq \underset{c_\alpha}{\overset{?}{\shortparallel}}$ ← find it so that the test is exactly of level α

. compute the pvalue

(Use Φ the cdf of $\mathcal{N}(0,1)$)

On simulation → cdf of the pvalues (empirical on N simulations)





$P_{H_0}(p \leq \alpha) = \alpha$

(because it's C^0)

$\Phi^{-1}(1 - \frac{\alpha}{2}) = ?$

$1 - \alpha/2$

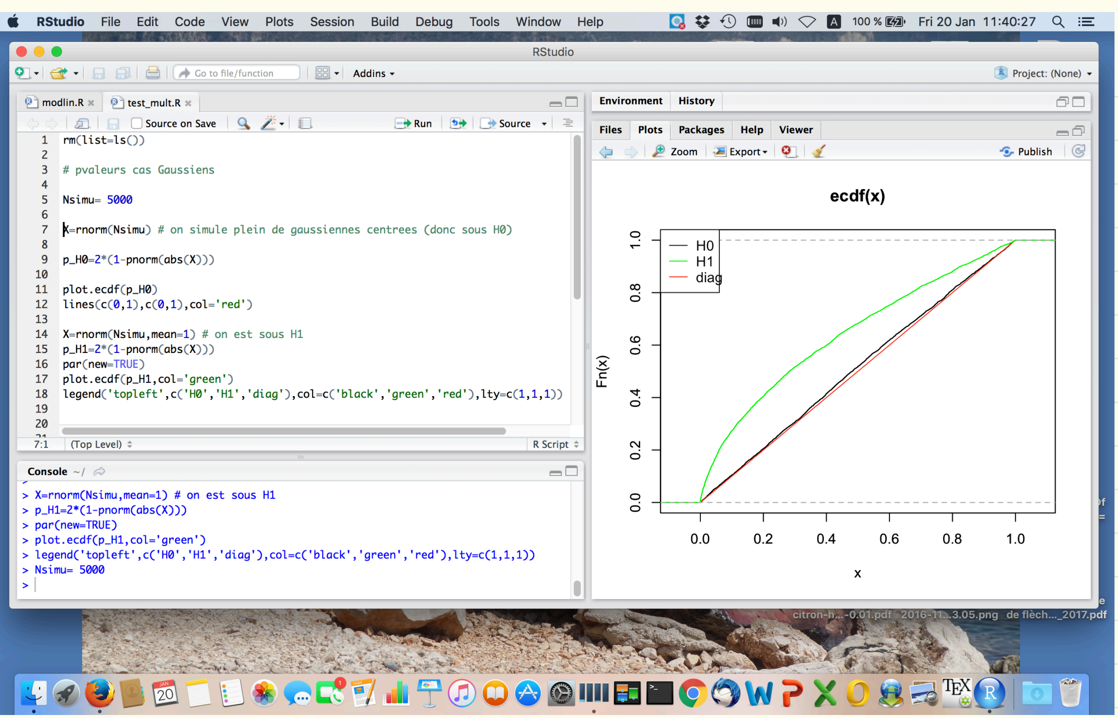
$$\underset{P(\mathcal{N}(0,1) \leq c_\alpha)}{\shortparallel}\Phi(c_\alpha) = 1 - \frac{\alpha}{2} \iff c_\alpha = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

pvalue p / $\alpha < p$ the test rejects
$\alpha > p$ the test accepts

$$\boxed{\begin{aligned} |X| &= \Phi^{-1}\left(1 - \frac{p}{2}\right) \\ \iff \Phi(|X|) &= 1 - \frac{p}{2} \\ \iff p &= 2(1 - \Phi(|X|)) \end{aligned}}$$

Test $\mathbb{1}_{|X| \geq \Phi^{-1}(1 - \frac{\alpha}{2})}$

Φ is C^0 $\Rightarrow$ Φ^{-1} takes all possible in $\mathbb{R}$ $\Rightarrow$ the pvalue p / p / $|X| = \Phi^{-1}(1 - \frac{p}{2})$



Discrete example

A company building tires, states in its ads that 99,5% of its tires can run 40000 km.

Observe 400 tires, 6 are flat after 40000 km.
What do you think? Has the company cheated?

Let θ = pty that a tire is flat after 40000 km

We want to test $H_0 : \theta = 0{,}005$ vs $H_1 : \theta > 0{,}005$

If T = number of observed flat tires n = total number of tires

then $\hat{\theta}$ is $\frac{T}{n}$

Reject when $\hat{\theta} > c_\alpha$ c_α to compute with the dist° under H_0

Under H_0, $T \sim \mathcal{B}(400, 0.005)$

Binomial / Poisson approx $\mathcal{P}(\underbrace{400 \times 0.005}_{=2})$

(See what we have done in the general case)

$\Delta_\alpha = \mathbb{1}_{T > F^{-1}(1-\alpha)}$ where F cdf of $\mathcal{B}(400, 0.005)$ or $\mathcal{P}(2)$

$= \mathbb{1}_{\hat{\theta} > \frac{F^{-1}(1-\alpha)}{n}}$

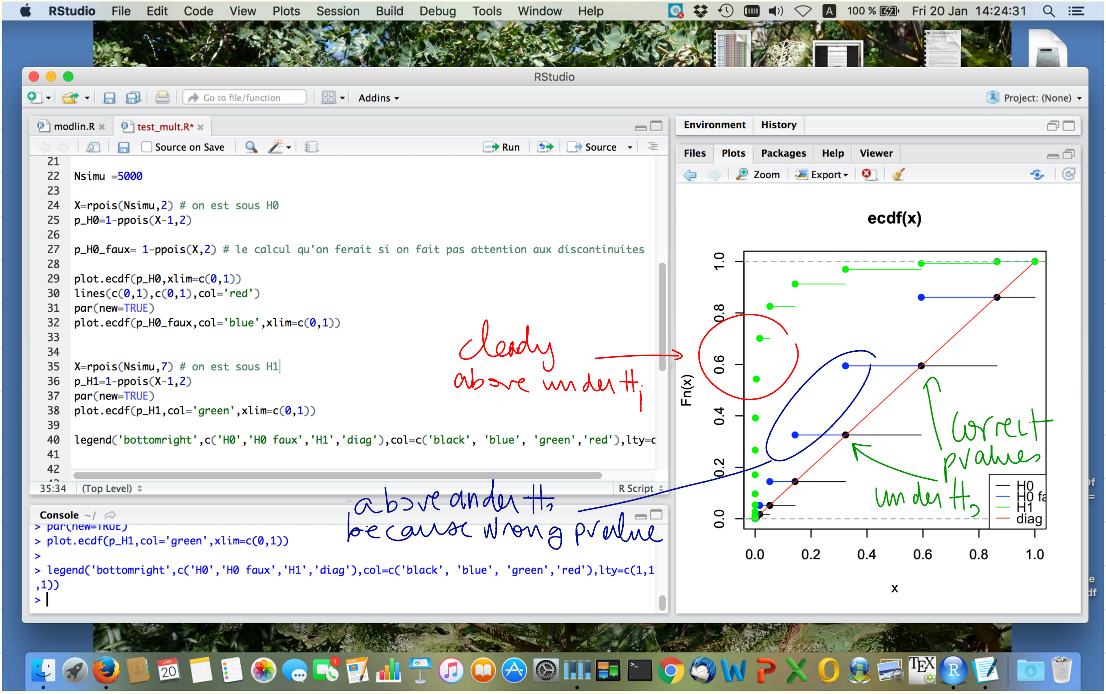
It's of level α $\forall \alpha$. And the associated pvalue is $1 - F_{-}(T)$

Here it is the same as $1 - F(T-1)$

↳ simulations of $\mathcal{P}(2)$ / $\mathcal{B}(400, 0.005)$ N_{sim}

Compute the true pvalue $1 - F(T-1)$
and the wrong one $1 - F(T)$

plot the corresponding ecdf and check that for the correct pvalue $\mathbb{P}_{H_0}(p \leq \alpha) \leq \alpha$



Observed pvalue is $1 - F(6-1) = 1,6\%$

quite small pvalue so the company was a bit too enthusiastic

Rmk: very useful in case of bootstrap/resampled dist^{o}
(maybe at the end of the course I'll explain how it works)

II Multiple testing formalism

We observe a variable X with distribution P

($\underline{ex}$ $X \sim \mathcal{N}_d(m, \Sigma)$)

We assume that $P \in \mathcal{P}$ (the set of all possible distributions of X)

($\underline{ex}$ $\mathcal{P} = \{ \mathcal{N}_d(m, \Sigma) \ / \ m \in \mathbb{R}^d, \ \Sigma \succcurlyeq 0 \text{ in } \mathbb{R}^{d \times d} \}$

or $\mathcal{P} = \{ \mathcal{N}_d(m, \sigma I_d), \ m \in \mathbb{R}^d, \ \sigma > 0 \}$

An hypothesis H is a $\underline{\underline{\text{subset}}}$ of $\mathcal{P}$.

$\underline{ex}$: if you want to test $m = 0 \to H = \{ \mathcal{N}(0, \Sigma) \ / \ \Sigma \succcurlyeq 0 \text{ in } \mathbb{R}^{d \times d} \}$

$$\complement \mathcal{P} = \{ \mathcal{N}(m, \Sigma) \ / \ m \in \mathbb{R}^d, \ \Sigma \succcurlyeq 0 \text{ in } \mathbb{R}^{d \times d} \}$$

if you want to test $m_1 = 0$

$$H = \left\{ \mathcal{N}\left(\begin{pmatrix} 0 \\ m_2 \\ \vdots \\ m_d \end{pmatrix}, \Sigma \right) \ / \ m_2, \dots m_d \in \mathbb{R}, \ \Sigma \succcurlyeq 0 \text{ in } \mathbb{R}^{d \times d} \right\}$$

If $\underline{H \text{ is true}}$ it means that P the distribution underlying the observation X is st $\boxed{P \in H}$

When we are dealing with multiple testing, we assume that we have a collection $\mathcal{H}$ of hypotheses H (usually finite) ex : $\mathcal{P} = \{ \mathcal{N}_d(m, I) \; / \; m \in \mathbb{R}^d \}$

$$H_1 = \left\{ \mathcal{N}_d\left(\begin{pmatrix} 0 \\ m_2 \\ \vdots \\ m_d \end{pmatrix}, I \right) \; / \; m_2, \dots, m_d \in \mathbb{R} \right\} \iff \text{" } m_1 = 0 \text{ "}$$

$\vdots$

$$H_j = \left\{ \mathcal{N}_d\left(\begin{pmatrix} \vdots \\ 0 \\ \vdots \end{pmatrix} \leftarrow j, I \right), \dots \right\} \iff \text{" } m_j = 0 \text{ "}$$

$\mathcal{H} = \{ H_1, \dots, H_d \}$ and therefore

H_j true $\iff P \in H_j \iff m_j = 0$

$P \in H_1 \cap H_2$ it means that H_1 and H_2 are true hence that $m_1 = m_2 = 0$

$P \in \bigcap_{H \in \mathcal{H}} H$, all H are true $\Rightarrow$ (for the example) $P = \mathcal{N}\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, I \right)$

The outcome of a multiple testing procedure is a random set $\mathcal{R}$ which only depends on the observation X such that $\boxed{\mathcal{R} \subset \mathcal{H}}$

It's the collection of rejected hypotheses

Ex By looking at the data, you have strong evidences that m_3 and $m_4 \neq 0$

$$\Rightarrow \mathcal{R} = \{ H_3, H_4 \}$$

R^c is the collection of accepted hypotheses

$\underline{ex}$ $\Rightarrow$ $R^c = \{1, 2, 5 \dots d\}$ $\Rightarrow$ you think that $m_1 = m_2 = m_5 \dots = m_d = 0$

$\underline{NB}$ if $R = \emptyset$ it means that you think that all hypotheses are true $\Rightarrow$ $P = \mathcal{N}(0, I)$

if $R = \{H_1 \dots H_d\} : \mathcal{H}$ $\Rightarrow$ none of the coeff are null (at least that's you think)

The aim of R is to estimate the set of false hypotheses of P

$$\mathcal{F}(P) = \{ H \in \mathcal{H} \,/\, P \notin H \}$$

$\underline{ex}$ $P = \mathcal{N}\left(\begin{pmatrix} 1 \\ 2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, I \right)$

$\underbrace{m_1 \text{ and } m_2 \neq 0}_{P \notin H_1,\; P \notin H_2}$, $\underbrace{m_3 = \dots = m_d = 0}_{P \in H_3 \cap \dots \cap H_d}$

$$\mathcal{F}(P) = \{H_1, H_2\}$$

The set of true hypotheses:

$$\mathcal{T}(P) = \{ H \in \mathcal{H} \,/\, P \in H \}$$

$\underline{ex}$ $\mathcal{T}(P) = \{H_3, \dots, H_d\}$

First kind error

With just one test, the first kind error is $P(\text{test rejects } H)$ when H true

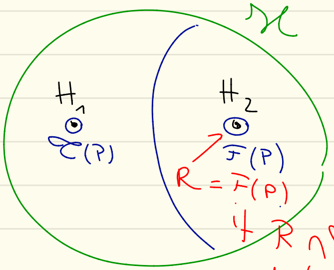
usually the test is of level α
$\Rightarrow$ the 1st kind error $\leq \alpha$

Now we want to do it for multiple testing
$\Longrightarrow$ you don't want to wrongly reject
$\Longrightarrow$ you want to control $R \cap \mathcal{C}(P)$

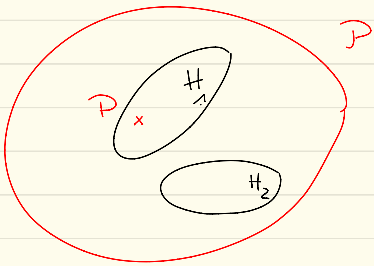
↪ various measures of that

<u>NB</u> the 2nd kind error $\longleftrightarrow$ $R^c \cap \mathcal{F}(P)$

<u>NB</u> If $\left.\begin{array}{l} R \cap \mathcal{C}(P) = \emptyset \\ R^c \cap \mathcal{F}(P) = \emptyset \end{array}\right\} \Longleftrightarrow R = \mathcal{F}(P)$ (you're perfect!)



if $R \cap \mathcal{C}(P) = \emptyset$
and $R^c \cap \mathcal{F}(P) = \emptyset$



formalism

$\mathcal{C}(P) = \{H_1\}$ $\quad \mathcal{H} = \{H_1, H_2\}$
$\mathcal{F}(P) = \{H_2\}$

Goeman and Solari (Ann of Stats 2010)
Roquain (review SFdS)

III Family Wise Error Rate

1/ Definition

$$\boxed{FWER(R) = \underbrace{\sup_{P \in \mathcal{P}}} P(R \cap \mathcal{T}(P) \neq \emptyset)}$$

NB depending on the author
$$FWER_P(R) = P(R \cap \mathcal{T}(P) \neq \emptyset)$$

NB: In my course, I'll always use the sup because it is coherent with what we do for classical test

When I write $P_{H_0}(A \text{ rejects}) \leq \alpha$

I meant $\forall P \in H_0, P(A \text{ rejects}) \leq \alpha$

i.e. $\sup_{P \in H_0} P(A \text{ rejects}) \leq \alpha$

We will look for procedures that guarantee $FWER(R) \leq \alpha$ (say 5%)

NB there is a weaker criterion: the weak Family Wise Error Rate

$$wFWER(R) = \sup_{P \in \bigcap_{H \in \mathcal{H}} H} P(R \cap \mathcal{T}(P) \neq \emptyset) = \sup_{P \in \bigcap_{H \in \mathcal{H}} H} P(R \neq \emptyset)$$

2/ Bonferroni

To each hypothesis $H \in \mathcal{H}$ we perform a test and define a pvalue p_H $\left(\forall \alpha \in [0,1) \ P(p_H \leq \alpha) \leq \alpha \text{ if } P \in H \right)$

$$\boxed{R_{Bonf} = \left\{ H \,/\, p_H \leq \frac{\alpha}{\#\mathcal{H}} \right\}}$$

Rmk: Before, Bonf $\Leftrightarrow$ performing tests at level $\alpha / \#$tests

Here $\#$tests $= \#\mathcal{H}$ and the test of hyp H is just $\mathbb{1}_{p_H \leq \text{level}}$

wFWER control

take $P \in \bigcap_{H \in \mathcal{H}} H$ so for all H, $P(p_H \leq u) \leq u \ \forall u \in (0,1)$

$$P(R_{Bonf} \neq \phi) = P\left(\exists H \in \mathcal{H},\ p_H \leq \frac{\alpha}{\#\mathcal{H}}\right)$$

$$\leq \sum_{H \in \mathcal{H}} P\left(p_H \leq \frac{\alpha}{\#\mathcal{H}}\right) \leq (\#\mathcal{H}) \frac{\alpha}{\#\mathcal{H}} = \alpha$$

Hence $\boxed{wFWER(R_{Bonf}) \leq \alpha}$

FWER control

$P \in \mathcal{P}$

if $P \in H$ then $P(p_H \leq u) \leq u \ \forall u \in (0,1)$

Hence $\forall H \in \mathcal{C}(P)$, $P(p_H \leq u) \leq u \ \forall u \in (0,1)$

$$P(R \cap \mathcal{C}(P) \neq \emptyset) = P(\exists H \in \mathcal{C}(P),\ H \in R)$$

$$= P\left(\exists H \in \mathcal{C}(P),\ p_H \leq \frac{\alpha}{\#\mathcal{H}}\right)$$

$$\leq \sum_{H \in \mathcal{C}(P)} P\left(p_H \leq \frac{\alpha}{\#\mathcal{H}}\right) \leq \frac{\#\mathcal{C}(P)}{\#\mathcal{H}}\alpha \leq \alpha$$

because $\mathcal{C}(P) \subset \mathcal{H}$

Hence $\boxed{\mathrm{FWER}(R_{Bonf}) \leq \alpha}$

Simulation

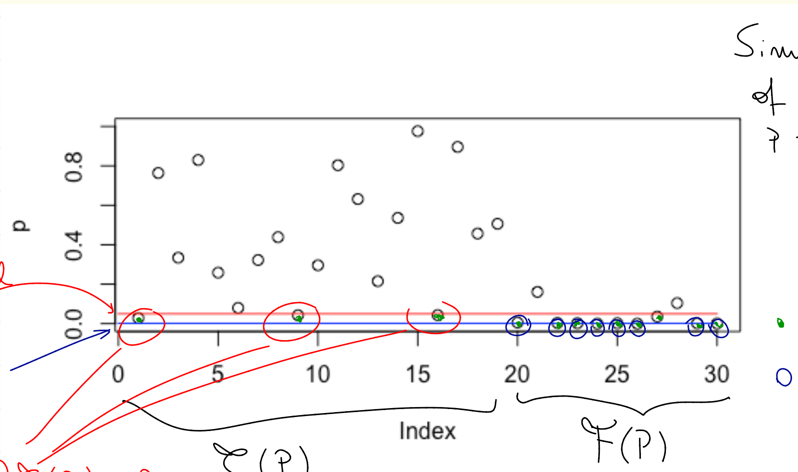
- Simulate $X \sim \mathcal{N}_d(0, I)$ $(d = 30)$ $\quad \mathcal{H} = \{H_1 \dots H_d\}$

  $H_i \Leftrightarrow m_i = 0$

  Compute on simulation

  wFWER for $R_{Bonf} = \left\{H \,/\, p_H \leq \frac{\alpha}{\#\mathcal{H}}\right\}$

  $R_{not\ corr} = \{H \,/\, p_H \leq \alpha\}$



Simulation of $\mathcal{N}\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 3 \\ \vdots \\ 3 \end{pmatrix} \begin{matrix} \}19 \\ \\ \}11 \end{matrix}, I\right)$

$p = 30$

at level 5%

$\frac{5\%}{30}$

$\mathcal{C}(P)$ $\quad \mathcal{F}(P)$

$R_{non\ corr} \cap \mathcal{C}(P)$ = false positive

• $R_{non\ corr}$

o R_{Bonf}

Code of the plot above

```
rm(list=ls())

### simulate a Gaussian vector
d=30 # 30 coordinates
level=0.05

X=rnorm(d)
X[20:30]=X[20:30]+3 ### the last ten have a mean which is non zero

### we perform a test of nullity for each of the coordinates and transform them into p-values

p=2*(1-pnorm(abs(X)))

plot(p,ylim=c(0,1))
### illustration of Bonferroni
lines(c(0,30),c(level,level),col='red')
lines(c(0,30),c(level/30,level/30),col='blue')
```

eleven

→ if you don't want to be in $\cap H$.

→ transform into pvalues (do not do it directly on X)

Code for checking wFER

```
### check that Bonferroni controls the wFWER

Nsimu=5000
wFWER_nocorr=0
wFWER_bonf=0
d=30
for(j in 1:Nsimu)
{
 X=rnorm(d)
 p=2*(1-pnorm(abs(X)))
 index_nocorr=which(p<level)
 if (length(index_nocorr)>0)
 {
  wFWER_nocorr=wFWER_nocorr+1
 }
 index_bonf=which(p<level/d)
 if (length(index_bonf)>0)
 {
  wFWER_bonf=wFWER_bonf+1
 }
}

wFWER_nocorr=wFWER_nocorr/Nsimu
wFWER_bonf=wFWER_bonf/Nsimu
```

← initialisation

← loop on the simulation

← sim of $\mathcal{N}(0,I)$ (only P in $\bigcap_{H\in\mathcal{H}} H$)

← pvalues

← $R_{noncorr}$

← if $R \neq \emptyset$

(Recall $\mathcal{H} = \{H_1 \dots H_d\}$, $H_j \Leftrightarrow m_j = 0$)

← R_{Bonf}

d: number of hypothesis

} final renormalisation

You should find $wFWER(R_{Bonf}) \approx 5\%$

and $wFWER(R_{noncorr}) \approx 78\%$ (Never compare a pvalue to 5% if you're doing it for several at a time!

```
Nsimu=5000
FWER=0


for(j in 1:Nsimu)
{
  X=rnorm(d)
  X[20:30]=X[20:30]+3
  p=2*(1-pnorm(abs(X)))
  index=which(p<level/d) # all the ones that are detected
  false_positive=which(index<20)
  if (length(false_positive)>0)
  {
    FWER=FWER+1
  }
}

FWER=FWER/Nsimu
```

Code for FWER (just for one P and not $\sup_P$)

$\mathcal{N}\left(\begin{pmatrix}0\\ \vdots\\ 0\\ 3\\ \vdots\\ 3\end{pmatrix}, I\right)$

R_{Bonf}

finding $R_{Bonf} \cap \mathcal{T}(P)$

$P(R \cap \mathcal{T}(P) \neq \phi) \leq \alpha$

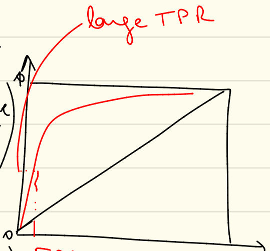
you should find $\approx 3\%$

$$\approx \frac{1}{N_{simu}} \sum_{i \leq N_{simu}} \mathbb{1}_{R_i \cap \mathcal{T}(P) \neq \phi}$$

NB $\underbrace{\frac{\overset{\#\mathcal{T}(P)}{19}}{30} \times 0.05}_{\text{Bound we computed}} \simeq 3\%$

The $H_i \in R \cap \mathcal{T}(P)$ are false positives

ROC curve



TPR $= E\left(\frac{\# \text{ true positive}}{\# \text{ possible positive}}\right)$

$= E\left(\frac{\#(R \cap F(P))}{\# F(P)}\right)$

$\text{FPR} = E\left(\frac{\# \text{ false positive}}{\# \text{ possible neg}}\right) = E\left(\frac{\#(R \cap \mathcal{T}(P))}{\# \mathcal{T}(P)}\right)$

Usually for a given class/method you have a parameter λ, $\to$ for each λ TPR_λ and $\text{FPR}_\lambda \to$ plot.

Plot the Roc curve for R_{Bonf} and $R_{non\ corr}$ for different α (replacing λ)

$\alpha = 0{,}001 \ldots 0{,}005$ step 0.001
$0{,}01 \ldots 1$ step 0.1

Code for Roc curves

```
Nsimu=5000
mean=3

level=c(seq(0.001,0.005,by=0.001),seq(0.01,1,by=0.01))
ll=length(level)
```

possible α

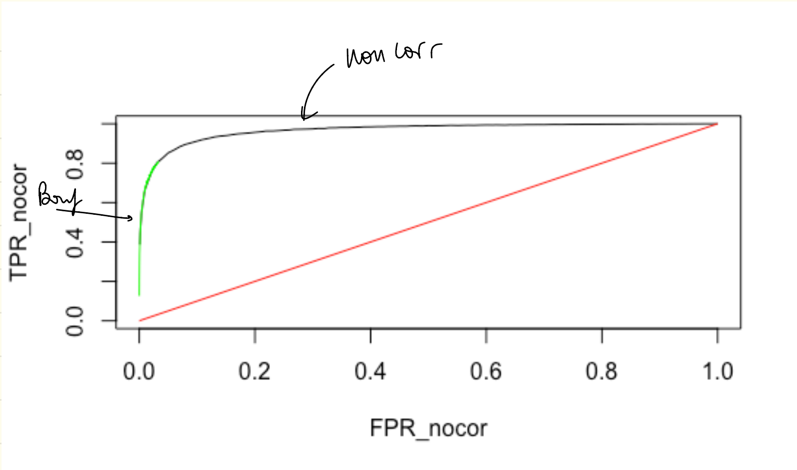
```
TPR_nocor=matrix(0,nrow=Nsimu,ncol=ll)
FPR_nocor=matrix(0,nrow=Nsimu,ncol=ll)
TPR=matrix(0,nrow=Nsimu,ncol=ll)
FPR=matrix(0,nrow=Nsimu,ncol=ll)
```

} initialisation

```
for(j in 1:Nsimu)
{
 for(l in 1:ll)
 {
  X=rnorm(d)
  X[20:30]=X[20:30]+mean
  p=2*(1-pnorm(abs(X)))
  index=which(p<level[l]) # all the ones that are detected = the positives without correction
  false_positive=which(index<20)
  true_positive=which(index>=20)
  TPR_nocor[j,l]=length(true_positive)/11 # 11 is the number of hypothesis you want to reject
  FPR_nocor[j,l]=length(false_positive)/19 # 19 is the number of hypothesis you want to accept
  index=which(p<level[l]/d) # all the ones that are detected = the positives with Bonferonni correction
  false_positive=which(index<20)
  true_positive=which(index>=20)
  TPR[j,l]=length(true_positive)/11 # 11 is the number of hypothesis you want to reject
  FPR[j,l]=length(false_positive)/19 # 19 is the number of hypothesis you want to accept


 }
}

FPR=colMeans(FPR)
TPR=colMeans(TPR)
FPR_nocor=colMeans(FPR_nocor)
TPR_nocor=colMeans(TPR_nocor)

plot(FPR_nocor,TPR_nocor,xlim=c(0,1),ylim=c(0,1),type='l')
lines(FPR,TPR, col='green')
lines(c(0,1),c(0,1),col='red')
```

$R_{non\ corr}$

R_{Bonf}



The ROC curves are the same except that they are not plotted on the same portion of space.

$\Longrightarrow$ from a classical point of view, R_{Bonf} and $R_{noncorr}$ are equivalent

(it's just a change of parametrization $\alpha \to \alpha/d$ for the same curve)

$\hookrightarrow$ multiple testing approaches ask more to the method

they want to guarantee that $FWER(R) \leq \alpha$

$\Longrightarrow$ which makes here the difference between R_{Bonf} and $R_{noncorr}$

2. The min-p procedure

Let F_0 be the cumulative dist$^\circ$ function

of $\min_{H \in \mathcal{H}} p_H$ under $P \in \bigcap_{H \in \mathcal{H}} H$ (and assume the p_H are continuous)

This assumes two things

1) the dist° of the minp is <u>known</u>

2) It doesn't depend on the P in $\bigcap_{H \in \mathcal{H}} H$

<u>ex</u> if test of $m_j = 0 \quad j = 1 \dots d$

- in $\mathcal{P} = \{ \mathcal{N}(m, I_n) \,/\, m \in \mathbb{R}^n \}$

then $P \subset \bigcap H \iff P = \mathcal{N}(0, I)$

↳ of course no problem

- in $\mathcal{P} = \{ \mathcal{N}(m, \sigma^2 I_n) \,/\, m \in \mathbb{R}^n, \sigma > 0 \}$

then I need to find p_H /

$\min p_H$ has the same dist° $\forall P = \mathcal{N}(0, \sigma^2 I_n)$, $\sigma > 0$

Then we define $\boxed{R_{\min p} = \{ H \,/\, p_H \leq F_0^{-1}(\alpha) \}}$

↑ the α quantile of F_0

<u>NB</u>: Useful only if the p_H are highly correlated

For instance if all the p_H's are equal to 1 pvalue p

(ie you're doing $\# \mathcal{H}$ times the same test) then

$\min p_H = p \qquad F_0^{-1}(\alpha) = \alpha$ (since $p \sim \mathcal{U}([0,1])$ under H_0)

(I'm considering only c° p_H) and you will reject more with $\min p$ $\mathcal{R} = \{H / p_H \leq \alpha\}$ (in this case) than with $R_{Bonf} = \{H / p_H \leq \frac{\alpha}{\#\mathcal{H}}\}$

Control of wFWER

Let $P \in \bigcap_{H \in \mathcal{H}} H$ $\quad (\mathcal{E}(P) = \mathcal{H})$

$$P(R_{\min p} \neq \phi) = P(\exists H \in \mathcal{H}, p_H \leq F_0^{-1}(\alpha))$$
$$= P(\min_{H \in \mathcal{H}} p_H \leq F_0^{-1}(\alpha))$$
$$= F_0(F_0^{-1}(\alpha)) = \alpha \quad (\text{we assume it's } c^\circ)$$

$\Rightarrow \sup_{P \in \bigcap_{H \in \mathcal{H}} H} P(R_{\min p} \neq \phi) \leq \alpha$ ie $\mathrm{wFER}(R_{\min p}) \leq \alpha$

Control of FWER

To do so we need another assumption

$(*)$ $\forall g \subset \mathcal{H}$, the dist° of $\min_{H \in g} p_H$ is the same $\forall P \in \bigcap_{H \in g} H$

Then $P \subset P$

$$P\left(R_{\min p} \cap \mathcal{C}(P) \neq \emptyset\right) = P\left(\exists H \in \mathcal{C}(P),\ p_H \leq F_0^{-1}(\alpha)\right)$$
$$= P\left(\min_{H \in \mathcal{C}(P)} p_H \leq F_0^{-1}(\alpha)\right)$$

Let's apply ⊛ with $\mathcal{G} = \mathcal{C}(P)$

$P \in \bigcap_{H \in \mathcal{G}} H$, where $\mathcal{G} = \mathcal{C}(P)$ | but there is also $Q \in \bigcap_{H \in \mathcal{H}} H$
that satisfies $Q \in \bigcap_{H \in \mathcal{G}} H$

By ⊛, $P\left(R_{\min p} \cap \mathcal{C}(P) = \emptyset\right) = Q\left(\min_{H \in \mathcal{G}} p_H \leq F_0^{-1}(\alpha)\right)$

(I can do as if all H in $\mathcal{H}$ are true!)

and $\leq Q\left(\min_{H \in \mathcal{H}} p_H \leq F_0^{-1}(\alpha)\right) = F_0\left(F_0^{-1}(\alpha)\right) = \alpha$

(because $\min_{\mathcal{H}} \leq \min_{\mathcal{G}}$)

So $\sup_{P \in \mathcal{P}} P\left(R_{\min p} \cap \mathcal{C}(P) \neq \emptyset\right) \leq \alpha$

ie $FWER(R_{\min p}) \leq \alpha$

3/ Step down improvement

Stepdown is an iterative procedure whose aim is to increase recursively the set R and still keeping a controlled FWER

To do so we need $\mathcal{N}$: $\mathcal{G} \subset \mathcal{H} \longmapsto \mathcal{N}(\mathcal{G}) \subset \mathcal{H}$
"Next"

$\mathcal{N}$ should satisfy:

<u>Monotony</u> $\forall \mathcal{G} \subset \mathcal{G}' \subset \mathcal{H}, \ \mathcal{N}(\mathcal{G}) \subset \mathcal{N}(\mathcal{G}') \cup \mathcal{G}'$

<u>False set control</u>

$$\forall P \in \mathcal{P}, \quad \mathbb{P}\big(\mathcal{N}(\mathcal{F}(P)) \subset \mathcal{F}(P)\big) \geq 1-\alpha$$

Then the stepdown is defined by

$$\boxed{\begin{cases} R_0 = \emptyset \\ R_{n+1} = R_n \cup \mathcal{N}(R_n) \end{cases} \quad \text{and} \quad R_{sd} = \text{"}\lim_{n \to +\infty}\text{"} R_n}$$

<u>NB</u> $\#\mathcal{H}$ is finite so $(R_n)_{n \geq 0}$ will be constant after a while

Control of the FWER

We want for a $P \in \mathcal{P}$ to control the proba under P of

$$R_{SD} \cap \mathcal{T}(P) \neq \emptyset$$

If this is the case $\exists n$ (random) such that $\begin{cases} \mathcal{N}(R_{n-1}) \cap \mathcal{T}(P) \neq \emptyset \\ R_n \cap \mathcal{T}(P) = \emptyset \end{cases}$

If $R_n \cap \mathcal{T}(P) = \emptyset$ it means that $R_n \subset \mathcal{F}(P)$

monotony $\Rightarrow$ $\mathcal{N}(R_n) \subset \mathcal{N}(\mathcal{F}(P)) \cup \mathcal{F}(P)$

So it means that $\mathcal{N}(\mathcal{F}(P)) \cap \mathcal{T}(P) \neq \emptyset$

By the false set control, this happens at most with probability α

$$\Rightarrow \quad P\left(R_{SD} \cap \mathcal{T}(P) \neq \emptyset\right) \leq \alpha$$

and $\mathrm{FWER}(R_{SD}) \leq \alpha$.

Examples of $\mathcal{N}$

Bonferroni

$$\mathcal{N}_B(g) = \left\{ H \,/\, p_H \leq \frac{\alpha}{\#\mathcal{H} - \#g} \right\}$$

<u>monotony</u> $\quad g \subset g' \Rightarrow \#g \leq \#g'$

hence $\dfrac{\alpha}{\#\mathcal{H} - \#g} \leq \dfrac{\alpha}{\#\mathcal{H} - \#g'}$

and $\mathcal{N}(g) \subset \mathcal{N}(g') \subset \mathcal{N}(g') \cup g'$

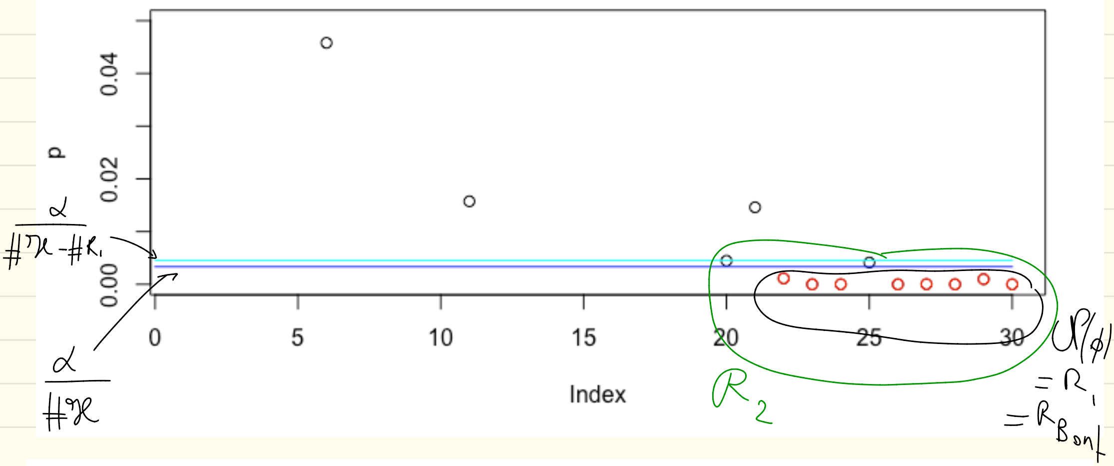
<u>False set control</u>

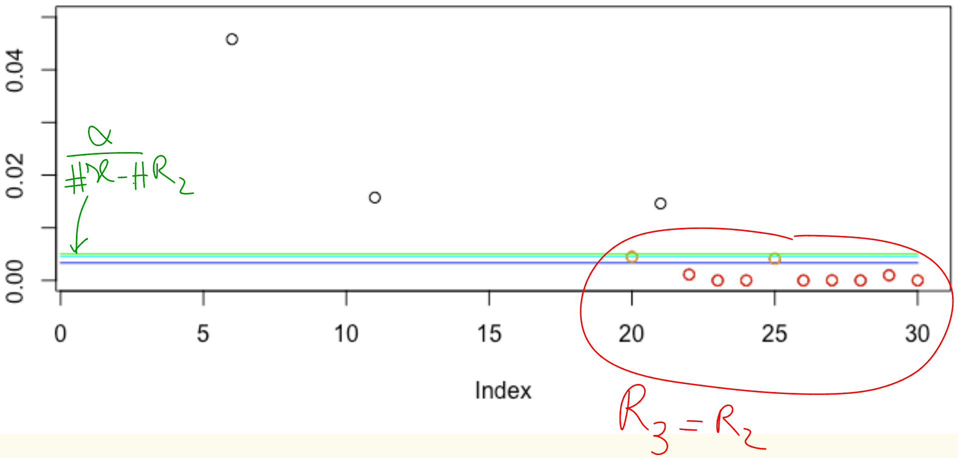
$P \in \mathcal{P}$

$$\mathbb{P}\left(\mathcal{N}(\mathcal{F}(P)) \cap \mathcal{T}(P) \neq \emptyset \right)$$

$$\leq \mathbb{P}\left(\exists H \in \mathcal{T}(P),\ p_H \leq \frac{\alpha}{\underbrace{\#\mathcal{H} - \#\mathcal{F}(P)}_{\#\mathcal{T}(P)}} \right)$$

$$\leq \sum_{H \in \mathcal{T}(P)} \mathbb{P}\left(p_H \leq \frac{\alpha}{\#\mathcal{T}(P)} \right) \leq \frac{\#\mathcal{T}(P)}{\#\mathcal{T}(P)} \alpha = \alpha$$





$\Rightarrow$ stop $R_{sd} = R_2$

Code for the plots above

```
d=30 # 30 coordinates
level=0.1

X=rnorm(d)
X[20:30]=X[20:30]+4
p=2*(1-pnorm(abs(X)))

## step 0
plot(p,ylim=c(0,0.05))
### illustration of Bonferroni
lines(c(0,30),c(level/30,level/30),col='blue')

index=which(p<level/d)

points(index, p[index],col='red')

### step 1

pstep1=p # one removes the points already detected by assigning them a new pval =1 (therefore removng them form the study)
pstep1[index]=rep(1,length(index))

index1=which(pstep1<level/(d-length(index)))
lines(c(0,30),c(level/(d-length(index)),level/(d-length(index))),col='cyan')

index1

points(index1,p[index1],col='orange')

pstep2=pstep1 # one removes the points already detected by assigning them a new pval =1 (therefore removng them form the study)
pstep2[index1]=rep(1,length(index1))

index2=which(pstep2<level/(d-length(index)-length(index1)))
lines(c(0,30),c(level/(d-length(index)-length(index1)),level/(d-length(index)-length(index1))),col='green')

index2
```

NB one can also define $\mathcal{R}_m$ with min pvalue

$$\mathcal{R}_m(\mathcal{G}) = \{ H \,/\, p_H \leq F_{\mathcal{G}^c}^{-1}(\alpha) \}$$

Where $F_{\mathcal{G}^c}$ is the cdf of $\min_{H \in \mathcal{G}^c} p_H$ under $P \in \bigcap_{H \in \mathcal{G}^c} H$

As an exercise, check monotony and false set control for $\mathcal{N}_m$

Simulation of R_{sd} built with $\mathcal{N}_{Bonf}$

function to compute R_{sd} with $\mathcal{N}_{Bonf}$

p the vector of pvalues
level : α

$\#\mathcal{H} \to$ d=length(p)
$R_{Bonf} \to$ index=which(p<level/d)
R_1

new "dim" $\#\mathcal{H} - \#R_n \to$ d=d-length(index)

index=which(p<level/d) $\leftarrow \mathcal{N}(R_n)$
reject=c(reject,index) $\leftarrow R_{n+1} = R_n \cup \mathcal{N}($

```
SDbonf=function(p,level)
{
 d=length(p)
 index=which(p<level/d)
 reject=index
 while(length(index)>0)
 {
   p[index]=1
   d=d-length(index)
   index=which(p<level/d)
   reject=c(reject,index)
 }
 return(reject)
}
```

$wFER(R_{sd/Bonf})$

```
#### control wFWER StepDown

Nsimu=5000
wFWER=0
level=0.05

for(j in 1:Nsimu)
{
 X=rnorm(d)
 p=2*(1-pnorm(abs(X)))
 index=SDbonf(p,level)
 if (length(index)>0)
 {
   wFWER=wFWER+1
 }
}

wFWER=wFWER/Nsimu
```

```
## control FWER Step Down
```

FWER($R_{Sd/Bonf}$)
(just for 1 P)

```
Nsimu=5000
FWER=0


for(j in 1:Nsimu)
{
  X=rnorm(d)
  X[20:30]=X[20:30]+3
  p=2*(1-pnorm(abs(X)))
  index=SDbonf(p,level)# all the ones that are detected
  false_positive=which(index<20)
  if (length(false_positive)>0)
  {
    FWER=FWER+1
  }
}

FWER=FWER/Nsimu


### comparison ROC curves / TPR /FPR
```

ROC curves

```
Nsimu=5000
mean=5

level=c(seq(0.001,0.005,by=0.001),seq(0.01,1,by=0.01))
ll=length(level)


TPR_SDBonf=matrix(0,nrow=Nsimu,ncol=ll)
FPR_SDBonf=matrix(0,nrow=Nsimu,ncol=ll)
TPR_Bonf=matrix(0,nrow=Nsimu,ncol=ll)
FPR_Bonf=matrix(0,nrow=Nsimu,ncol=ll)



for(j in 1:Nsimu)
{
  for(l in 1:ll)
  {
    X=rnorm(d)
    X[20:30]=X[20:30]+mean
    p=2*(1-pnorm(abs(X)))
    index=which(p<level[l]/d) # all the ones that are detected = the positives without correction
    false_positive=which(index<20)
    true_positive=which(index>=20)
    TPR_Bonf[j,l]=length(true_positive)/11 # 11 is the number of hypothesis you want to reject
    FPR_Bonf[j,l]=length(false_positive)/19 # 19 is the number of hypothesis you want to accept
    index=SDbonf(p,level[l]) # all the ones that are detected = the positives with Bonferonni correction
    false_positive=which(index<20)
    true_positive=which(index>=20)
```

```
TPR_SDBonf[j,l]=length(true_positive)/11 # 11 is the number of hypothesis you want to reject
    FPR_SDBonf[j,l]=length(false_positive)/19 # 19 is the number of hypothesis you want to accept


  }
}

FPR_Bonf=colMeans(FPR_Bonf)
TPR_Bonf=colMeans(TPR_Bonf)
FPR_SDBonf=colMeans(FPR_SDBonf)
TPR_SDBonf=colMeans(TPR_SDBonf)

plot(FPR_Bonf,TPR_Bonf,xlim=c(0,1),ylim=c(0,1),type='l') ### almost no difference
lines(FPR_SDBonf,TPR_SDBonf, col='green')
lines(c(0,1),c(0,1),col='red')


plot(level,TPR_Bonf,type='l') ### but for fixed level, improvement of the TPR for SDBonf
lines(level, TPR_SDBonf,type='l', col='green')

plot(level,FPR_Bonf,type='l') ### justification : more FPR with SDBonf but still controlled FWER
so SDbonf is less "conservative" but does not improve the ROC curve
lines(level, FPR_SDBonf,type='l', col='green')
```

↳ You should see that the green curve (R_{sd}) is above the black (Bonf) for TPR / level and FPR / level

→ It means that one makes more discoveries with SD! (and still control FWER ⟹ It's better)

IV False Discovery Rate

1/ Definition

$$\boxed{FDR(R) = \sup_{P \in \mathcal{P}} E\left(\frac{\#R \cap \mathcal{C}(P)}{\#R}\right)}$$

This is the false Discovery Rate with the convention $\frac{0}{0} = 0$

If no discoveries no mistake

The idea is that in some cases, you want to allow more discoveries by saying that as long as the percentage of false discoveries among the set of discoveries is low, you're happy!

Be Careful: as a statistician you have responsabilities!

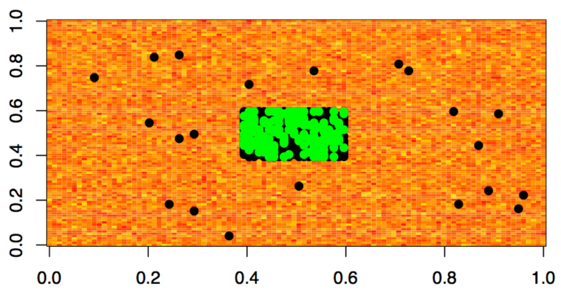
Ex:

If you need to take a decision and that one false positive has a lot of nasty implications
$\Longrightarrow$ control the FWER (medicine on a market, military app....)

If you want to screen and you know that your decisions will be checked again by different means
$\Longrightarrow$ Control the FDR

FDR is less "asking" guarantee than FWER (see the image below)

In black the discoveries with BH, in green with Bonferroni
BH controls the FDR, Bonferroni controls the FWER
True positive = the inner square



If $P \in \bigcap_{H \in \mathcal{H}} H$ $\left(\mathcal{T}(P) = \mathcal{H},\ F(P) = \emptyset\right)$

then $E\left(\dfrac{\# R \cap \mathcal{T}(P)}{\# R}\right) = P(R \neq \emptyset)$

So if there is nothing to discover both criteria gives the same control

when there is something to discover, there is a $\neq$ ce.

In general, $\text{FWER}(R) \geq \text{FDR}(R)$

2/ Benjamini and Hochberg method (95) (step up with more generally)

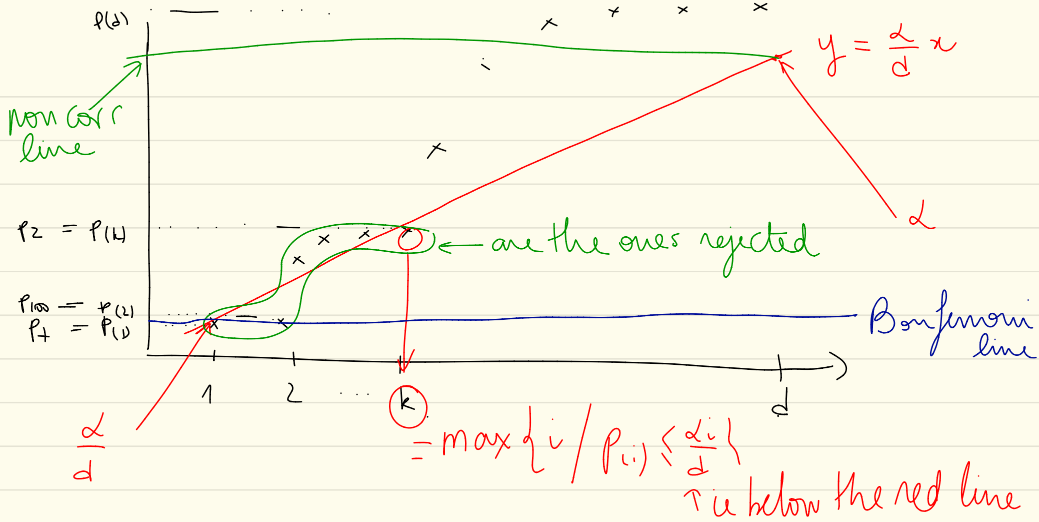
One of the most cited paper in statistics
In 20 years, it became the classical method imposed in many scientific areas (esp. genomics)

Principle:

1/ Sort the pvalues $p_{(1)} \leq p_{(2)} \cdots \leq p_{(d)}$

2/ Define $\boxed{k = \max\left\{ i \,/\, p_{(i)} \leq \frac{\alpha i}{d} \right\}}$ $(d = \# \mathcal{H})$

3/ Reject all the indices $\leftrightarrow p_{(1)} \cdots p_{(k)}$



We can prove that

1/ If the p_H are independent, $FDR(R) \leq \alpha$

2/ If the p_H are coming from one sided test on Gaussian vectors
If the gaussian vector is positively corr
then $FDR(R) \leq \alpha$

3/ Most of the time even if these assumptions are not fulfilled
$FDR(R) \leq \alpha$ (thorough simulation)

4/ However in any case

$$FDR(R) \leq \alpha \underbrace{\left(1 + \frac{1}{2} + \dots + \frac{1}{d}\right)}_{\approx \log(d)}$$

See interactive web page of E. Roquain FDR

Function for computing R_{BH}

```
BH=function(p,level)
{
  d=length(p)
  psort=sort(p,index.return=TRUE)
  vec=level*(1:d)/d
  ind=which(psort$x<=vec)
  if(length(ind)>0)
  {
    reject=psort$ix[1:max(ind)]
  }else{reject=c()}
  return(reject)
}
```

ask for original indices

sort pvalues

red line

indices below the line

the value sorted

go until the max of indices

the original indices of the first 1...k sorted pvalues

Code for computing FDR/FWER of BH

```
Nsimu=5000
d=30
mean=10
FDR=0
FWER=0

for(j in 1: Nsimu)
{
  X=rnorm(d)
  X[20:30]=X[20:30]+mean
  p=2*(1-pnorm(abs(X)))
  index=BH(p,level) # all the ones that are detected = the positives without correction
  false_positive=which(index<20)
  if(length(index)>0)
  {
    FDR=FDR+length(false_positive)/length(index)
  }
  if(length(false_positive)>0)
  {
    FWER=FWER+1
  }
}

FDR=FDR/Nsimu
FWER=FWER/Nsimu
```

Code for the visualization with the square

```
### visual effect of FDR control versus FWER control

dim=100
X=matrix(rnorm(dim*dim),nrow=dim,ncol=dim)
mean=4

image(X)

X[(2*dim/5):(3*dim/5),(2*dim/5):(3*dim/5)]=X[(2*dim/5):(3*dim/5),(2*dim/5):(3*dim/5)] +mean

image(X)

p=2*(1-pnorm(abs(X)))

level=0.05

index=BH(p,level)
col_index=ceiling(index/dim)
row_index=index-((col_index-1)*dim)

ref=seq(0,1, length.out=dim)

points(ref[row_index],ref[col_index],pch=19) ### FDR

index=which(p<level/(dim*dim)) ### with Bonferroni and FWER control

col_index=ceiling(index/dim)
row_index=index-((col_index-1)*dim)

points(ref[row_index],ref[col_index],pch=19,col='green')

index=which(p<level) ### with no control

col_index=ceiling(index/dim)
row_index=index-((col_index-1)*dim)

points(ref[row_index],ref[col_index],pch=19,col='blue')
```