What's a p-value?

Until now, we have seen various ways to build a test $\Delta_x$ of level $\alpha$.

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

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

If we assume that $\Delta_x$ is monotonous, that is

$\forall x < x' \quad \Delta_x \leq \Delta_{x'}$

Then the p-value of the test $\Delta_x$ is the limit value $p$ which depends on the observations such that

$\forall x < p \quad \Delta_x = 0$ "I'll accept $H_0$ at level $\alpha'$

$\forall x > p \quad \Delta_x = 1$ "I reject $H_0$"

Classical example

If $\Delta_x$ is defined as $\Pi T > c_x$, where $T$ is the test statistic whose distribution is known under $H_0$ and continuous, and $c_x$ is the 1- level quantile of this distribution.

Then $p = 1 - F_{H_0}(T)$, where $F_{H_0}$ is the distribution of $T$ under $H_0$. 
1. It's of level \( \alpha \)

\[
\Delta_\alpha = \{ t \mid T > F^{-1}(1-\alpha) \}
\]

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

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

2. Monotonous

\( \alpha < \alpha' \)

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

If \( \Delta_\alpha = \{ t \mid T > F^{-1}(1-\alpha) \} = \emptyset \Rightarrow \Delta_{\alpha'} = \{ t \mid T > F^{-1}(1-\alpha') \} = \emptyset \)

\( \Delta_\alpha \subseteq \Delta_{\alpha'} \) in this case

If \( \Delta_{\alpha'} = \{ t \mid T > F^{-1}(1-\alpha') \} = \emptyset \Rightarrow \Delta_\alpha = \{ t \mid T > F^{-1}(1-\alpha) \} = \emptyset \)

\( \Rightarrow \Delta_\alpha \subseteq \Delta_{\alpha'} \) in all cases
3/ Check \( p(T) = 1 - F(T) \) \( \Rightarrow \) \( T = F^{-1}(1 - p(T)) \)

let \( \alpha < p(T) \) \( \Rightarrow \) \( 1 - \alpha > 1 - p(T) \)

\[ \Delta_\alpha = \frac{\| T > F^{-1}(1-\alpha) \|}{\| F^{-1}(1-p) > F^{-1}(1-\alpha) \|} \]

this is wrong

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

let \( \alpha > p(T) \) \( \Rightarrow \) \( 1 - \alpha < 1 - p(T) \)

\( \Rightarrow \Delta_\alpha = 1 \) (the test rejects)

**Small complements on cdf**

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

In general, \( F \) \( \overset{\text{a.s.}}{\rightarrow} \) \( F(-\infty) = 0 \) \( F(+\infty) = 1 \)

\( F \) \( \tilde{\text{cadlàg}} \) (continu à droite limité à gauche)

continous on the right, limited on the left

\( \text{ex: Poisson} \)

**Official definition of a quantile**
If $T$ is discrete with càdlàg cdf under $H_0$, and if $\Delta_{\alpha} = \sup \{ t : F^{-1}(1 - \alpha) \leq t \}
abla$, then the p-value is defined by $p(t) = 1 - \frac{F(t)}{F^{-1}(1 - \alpha)}$. 

Check: $\Delta_{\alpha}$ should be of level $\alpha$:

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

$$= 1 - F(F^{-1}(1 - \alpha))$$

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

$$F^{-1}(u) = \inf \{ t : F(t) \geq u \}$$
In general, \( F(F^{-1}(t)) > t \)

So \( P_{H_0}(A_\alpha = 1) = 1 - F(F^{-1}(1 - \alpha)) \leq \alpha \)

\( \geq 1 - \alpha \)

It's monotonous as before

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

\[ F^{-1} \]

Take \( \alpha = p(T) + \varepsilon \) small

\[ 1 - \alpha = 1 - p(T) - \varepsilon = F_-(T) - \varepsilon \]

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

\[ F^{-1}(1 - \alpha) \]

Take \( \alpha = p(T) - \varepsilon \) small

\[ 1 - \alpha = F_-(T) + \varepsilon \]

So \( T = F^{-1}(1 - \alpha) \) and \( T \geq F^{-1}(1 - \alpha) \)

In wrong

On simulations

We want to check the fundamental property of p-value

which is \( P_{H_0}(p \leq \alpha) \leq \alpha \) \& \( 1 - \alpha \)
- Gaussian case

obsrv $X \sim N(\mu, 1)$

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

$\mu$ on $\bar{X}$ estimated by $\bar{X}$

reject when $|\bar{X}| \geq ?$ = find $\alpha/2$ so that the test is exactly of level $\alpha$

compute the p-value $c_{\alpha}$ (Use $\Phi$ the cdf of $N(0, 1)$)

On simulation → cdf of the p-values (Empirical on N simulation)

$P_{H_0}(|X| \leq c_{\alpha}) = \alpha$

(proof)

$P_{H_0}(|X| \leq \Phi^{-1}(1 - \alpha/2))$

$p$ value $p/\alpha > p$ the test rejects
$\alpha > p$ the test accepts

Test $H_1: |X| > \Phi^{-1}(1 - \alpha/2)$

$\Phi$ is c.d.f. $\Rightarrow \Phi^{-1}$ takes all positive in $\mathbb{R}$ $\Rightarrow$ the p-value $p = P/|X| = \Phi^{-1}(1 - \alpha/2)$
Increte 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 \( \Theta \) = pty that a tire will fail after 40000 km.

We want to test \( H_0: \Theta = 0.005 \) vs \( H_1: \Theta > 0.005 \)
If \( T \) = number of observed flat tines, \( n \) = total number of tines, then \( \theta \) is \( \frac{T}{n} \)

Reject when \( \theta > c_1 \) to compute with the distribution under \( H_0 \).

Under \( H_0 \), \( T \sim B(400, 0.005) \)

Binomial/Poisson approx \( \frac{B(400, 0.005)}{\sqrt{n}} \)

(See what we have done in the general case)

\[ A_\alpha = \{ T > F^{-1}(1-\alpha) \} \text{ when } F \text{ cdf of } B(400, 0.005) \text{ or } P(2) \]

\[ = \{ \theta > F^{-1}(1-\alpha) \} \]

It's of level \( \alpha \). And the associated p-value is \( 1 - F(T) \)

Here, \( T \) is the same as \( 1 - F(T-1) \)

Simulations of \( P(2) / B(400, 0.005) \)

Compute the true p-value \( 1 - F(T-1) \)

and the wrong one \( 1 - F(T) \)

Plot the corresponding ecdf and check that for the correct p-value \( P_{H_0}(P < \alpha) \leq \alpha \)
Observed p-value is $1 - F(6-1) = 1.6\%$

quite small p-value so the company was a bit too enthusiastic

Rank: very useful in case of bootstrap/resampled data

(Many at the end of the course I'll explain how it works)
We observe a variable $X$ with distribution $P$
\[
\text{ex: } X \sim \mathcal{N}_d(m, \Sigma)
\]
We assume that $P \in \mathcal{P}$ (the set of all possible distributions of $X$)
\[
\text{ex: } \mathcal{P} = \left\{ \mathcal{N}_d(m, \Sigma) / m \in \mathbb{R}^d, \Sigma > 0 \text{ in } \mathbb{R}^{d \times d} \right\}
\]
\[
\text{or } \mathcal{P} = \left\{ \mathcal{N}_d(m, \sigma I_d) / m \in \mathbb{R}^d, \sigma > 0 \right\}
\]

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

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

$C\mathcal{P} = \left\{ \mathcal{N}(m, \Sigma) / m \in \mathbb{R}^d, \Sigma > 0 \text{ in } \mathbb{R}^{d \times d} \right\}$

If you want to test $m_1 = 0$

\[
H = \left\{ \mathcal{N}((0, \ldots, 0), \Sigma) / \Sigma > 0 \text{ in } \mathbb{R}^{d \times d} \right\}
\]

If $H$ is true, it means that the distribution underlying the observation $X$ is at $P \in H$

When we are dealing with multiple testing, we assume that we have a collection $\mathcal{H}$ of hypotheses $H$ (usually finite)
\[ \mathcal{P} = \left\{ \mathcal{N}_d(m, \mathcal{I}) \ \middle| \ m \in \mathbb{R}^d \right\} \]

\[ H_1 = \left\{ \mathcal{N}_d\left(\begin{bmatrix} \cdots \end{bmatrix}, \mathcal{I}\right) \ \middle| \ m_1 = 0 \right\} \]

\[ H_j = \left\{ \mathcal{N}_d\left(\begin{bmatrix} \cdots \end{bmatrix}, \mathcal{I}\right) \ \middle| \ m_j = 0 \right\} \]

\[ \mathcal{H} = \left\{ H_1, \ldots, H_d \right\} \text{ and therefore} \]

\[ H_j \text{ true } \Rightarrow P \in H_j \Leftrightarrow m_j = 0 \]

\[ P \in H_1 \cap H_2 \text{ it means that } H_1 \text{ and } H_2 \text{ are true} \]

\[ \text{hence that } m_1 = m_2 = 0 \]

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

The outcome of a multiple testing procedure is a random set \( \mathcal{R} \) which only depends on the observation \( X \) which is \( \mathcal{R} \subset \mathcal{H} \).

It's the collection of rejected hypotheses.

Ex. By looking at the data, you have strong evidence that \( m_3 \) and \( m_4 \neq 0 \)

\[ \Rightarrow \mathcal{R} = \left\{ H_3, H_4 \right\} \]
$R^c$ is the collection of accepted hypotheses

$$R^c = \{1, 2, 5, \ldots \} \Rightarrow \text{you think that } m_1 = m_2 = m_5 = \ldots = m_d = 0$$

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

If $R = \{H_1, \ldots, H_d\}$, then $\Rightarrow$ none of the null hypotheses are null (at least that’s what you think)

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

$$\mathcal{F}(P) = \{+ \in \mathcal{R} | P \neq +\}$$

$$\begin{align*}
\text{ex} & \quad P = \mathcal{N}\left(\left(\begin{array}{c} 1 \\ 2 \\ 0 \\
end{array}\right), I\right) \\
& \quad m_1 \text{ and } m_2 \neq 0 \quad m_3 = \ldots = m_d = 0 \\
& \quad P \neq +_1 \quad P \in +_3 \cap \ldots \cap +_d \\
& \quad P \neq +_2
\end{align*}$$

$$\mathcal{F}(P) = \{+_1, +_2\}$$

The set of true hypotheses:

$$\mathcal{C}(P) = \{H \in \mathcal{R} | P \in H\}$$

$$\begin{align*}
\text{ex} & \quad \mathcal{C}(P) = \{+_3, \ldots, +_d\}
\end{align*}$$
First kind error

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

usually the test n of level \( \alpha \)

\( \Rightarrow \) the 1st kind error \( \leq \alpha \)

Now we want to do it for multiple testing

\( \Rightarrow \) you don’t want to wrongly reject

\( \Rightarrow \) you want to control \( R \cap \mathcal{C}(P) \)

\( \Rightarrow \) various measures of that

NB the 2nd kind error \( \rightarrow R^c \cap \mathcal{F}(P) \)

NB 1 if \( R \cap \mathcal{C}(P) = \emptyset \) \( \Rightarrow \) \( R = \mathcal{F}(P) \) \( \Rightarrow \) you repudiate

Formalism

Geman and Solari (Ann of Stats 2010)

Roquain (review SfDS)

III Family Wise Error Rate
Definition

\[ \text{FWER} (R) = \sup_{P \in \mathcal{P}} P (R \cap \mathcal{C}(P) \neq \emptyset) \]

NB: depending on the author

\[ \text{FWER}_P (R) = P (R \cap \mathcal{C}(P) \neq \emptyset) \]

NB: In my course, I'll always use the sup

because it is similar to what we do for classical tests.

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

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

\[ \leq \sup_{P \in H_0} P (\text{rejects}) \leq \alpha \]

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

NB: there is a weaker criterion, the weak family wise error rate

\[ \text{WFWER} (R) = \sup_{P \in \mathcal{P}} P (R \cap \mathcal{C}(P) \neq \emptyset) = \sup_{P \in \mathcal{P}} P (R \neq \emptyset) \]

2/ Bonferroni
To each hypothesis $H \in \mathcal{H}$ we perform a test and define a p-value $p_H$ \( \forall \alpha \in [0, 1] \ P(p_H \leq \alpha) \leq \alpha \) if $p \in H$.

\[
R_{\text{Bonf}} = \{ H / p_H \leq \frac{\alpha}{\# \mathcal{H}} \}
\]

Remark: Before Bonferroni testing at level $\alpha / \# \mathcal{H}$, the number of tests $\# \mathcal{H}$ and the test of hypothesis $H$ is post. $p_H \leq \alpha$.

**WFWER control**

Take $p \in \mathcal{H}$ so in all $H$, $P(p_H \leq u) \leq u \ \forall \ u \in [0, 1]$

\[
P(R_{\text{Bonf}} \neq \emptyset) = P(\exists H \in \mathcal{H}, \ p_H \leq \frac{\alpha}{\# \mathcal{H}}) \\
\leq \sum_{H \in \mathcal{H}} P(p_H \leq \frac{\alpha}{\# \mathcal{H}}) \leq (\# \mathcal{H}) \frac{\alpha}{\# \mathcal{H}} \leq \alpha
\]

Hence \( \text{WFWER}(R_{\text{Bonf}}) \leq \alpha \)

**FWER control**

$\leq \alpha$

$\text{P} < \alpha$ if $p \in H$ then $P(p_H \leq u) \leq u \ \forall u \in [0, 1]$

Hence \( \forall H \in \mathcal{H}, \ P(p_H \leq u) \leq u \ \forall u \in [0, 1] \)
\[ P(\mathcal{R} \cap \mathcal{C}(p) \neq \emptyset) = P(\exists \mathcal{H} \in \mathcal{C}(p), \mathcal{H} \in \mathcal{R}) \]
\[ = P(\exists \mathcal{H} \in \mathcal{C}(p), \mathcal{H} \in \mathcal{R}) \]
\[ \leq \frac{\alpha}{\# \mathcal{C}(p)} \leq \frac{\# \mathcal{C}(p)}{\# \mathcal{R}} \alpha \leq \frac{\alpha}{\# \mathcal{R}} \]

Hence \( \text{FWER}(R_{\text{Bonf}}) \leq \alpha \) because \( \mathcal{C}(p) \subset \mathcal{R} \)

Simulation

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

Compute on simulation

\[ w_{\text{FWER}} \quad \text{for} \quad R_{\text{Bonf}} = \{ H / \mathcal{H} \leq \frac{\alpha}{\# \mathcal{H}} \} \]
\[ R_{\text{not corr}} = \{ H / \mathcal{H} \leq \alpha \} \]

Simulation

\[ \hat{p} = \frac{r}{d(2, 1, 2, 3, 1, 1)} \]

\[ R \text{ not corr} \]

\[ R \text{ Bonf} \]
### simulate a Gaussian vector

d=30  # 30 coordinates
level=0.05

\[ X = \text{rnorm}(d) \]
\[ X[20:30] = X[20:30] + 3 \]

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

\[ p = 2(1 - \text{pnorm}(\text{abs}(X))) \]

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')

### check that Bonferroni controls the wFWER

Nsimu=5000
wFWER_nocorr=0
wFWER_bonf=0

d = 30

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

wFWER_nocorr = wFWER_nocorr / Nsimu
wFWER_bonf = wFWER_bonf / Nsimu

\[ \text{You should find } w\text{FWER}(R_{\text{bonf}}) \approx 5\% \]

\[ \text{and } w\text{FWER}(R_{\text{nocorr}}) \approx 78\% \]
Code for FWER (just for one p and not sup)

\[
\text{for}(j \text{ in } 1: \text{Nsimu}) \\
\{ \\
    X = \text{rnorm}(d) \\
    p = 2(1 - \text{pnorm}(|X|)) \\
    \text{index} = \text{which}(p < \text{level}/d) \quad \# \text{all the ones that are detected} \\
    \text{false\_positive} = \text{which}(\text{index} < 20) \\
    \text{if} (\text{length(false\_positive)} > 0) \\
    \{ \\
        \text{FWER} = \text{FWER} + 1 \\
    \} \\
\}
\]

\[
\text{FWER} = \frac{\text{FWER}}{\text{Nsimu}}
\]

You should find \( \approx 3\% \)

\[
\frac{\#(\text{index})}{\text{Nsimu}} \times 0.05 \approx 3\%
\]

Bound we computed

The \( R \cap \mathcal{C}(p) \) are false positives

\[
P(R \cap \mathcal{C}(p) \neq \emptyset) \leq \frac{1}{Nsimu} \sum_{j=1}^{Nsimu} 1_{R \cap \mathcal{C}(p) \neq \emptyset}
\]

ROC curve

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

\[ \text{FPR} = \mathbb{E}\left( \frac{\#(\text{false positive})}{\#(\text{possible negative})} \right) \]

\[ 1 \quad \text{FPR} = \mathbb{E}\left( \frac{\#(\text{false positive})}{\#(\text{possible negative})} \right) \]

for each, \( \text{TPR}, \text{FPR} \rightarrow \text{plot} \)

usually for a given clam method

\[ \mathbb{E}\left( \frac{\#(R \cap \mathcal{C}(p))}{\#(\mathcal{C}(p))} \right) \]
Plot the ROC curve for \( R_{Bonf} \) and \( R_{non \ cor} \) for different \( \alpha \) (up to 0.1)

\[
\alpha = 0.001, 0.005, \ldots, 0.1
\]

Code for ROC curves

\[
\text{Nsimu}=5000
\]
\[
\text{mean}=3
\]

\[
\text{level=}\text{c(seq(0.001,0.005,by=0.001),seq(0.01,1,by=0.01))}
\]
\[
\text{ll=}\text{length(level)}
\]

\[
\text{TPR}_{\text{nocor}}=\text{matrix(0,nrow=Nsimu,ncol=ll)}
\]
\[
\text{FPR}_{\text{nocor}}=\text{matrix(0,nrow=Nsimu,ncol=ll)}
\]
\[
\text{TPR}=\text{matrix(0,nrow=Nsimu,ncol=ll)}
\]
\[
\text{FPR}=\text{matrix(0,nrow=Nsimu,ncol=ll)}
\]

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

\[
\text{FPR=colMeans(FPR)}
\]
\[
\text{TPR=colMeans(TPR)}
\]
\[
\text{FPR}_{\text{nocor}}=\text{colMeans(FPR}_{\text{nocor}})
\]
\[
\text{TPR}_{\text{nocor}}=\text{colMeans(TPR}_{\text{nocor}})
\]

\[
\text{plot(FPR}_{\text{nocor}},\text{TPR}_{\text{nocor}},\text{xlim=c(0,1)},\text{ylim=c(0,1)},\text{type='l'})
\]
\[
\text{lines(FPR,TPR, col='green')}
\]
\[
\text{lines(c(0,1),c(0,1),col='red')}
\]
The ROC curves are the same except that they are not plotted on the same portion of paper.

⇒ from a classical point of view, $R_{Bonf}$ and $R_{nocor}$ are equivalent.

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

⇒ multiple testing approaches ask more to the method they want to guarantee that $FWER(R) \leq \alpha$

⇒ which makes here the difference between $R_{Bonf}$ and $R_{nocor}$

9. The min-$p$ procedure

Let $F_0$ be the cumulative distribution function

of $\min p^+ \under P \in \mathcal{H}^+$ under $F_0 \in \mathcal{H}^+$ (and among the $p^+$ are continuous)
This assumes two things:

1) the dist\(^{10}\) of the min\(\bar{p}\) is known

2) \(H_0\) doesn't depend on the \(P\) in \(\cap H\)

\[ \text{ex. if test of } m_j = 0 \quad j = 1, \ldots, k \]

\[ \text{in } P = \{ \mathcal{N}(m, I) / m \in \mathbb{R}^n \} \]

then \(P \cap H \equiv P = \mathcal{N}(0, I) \)

\[ \text{of course no problem} \]

\[ \text{in } P = \{ \mathcal{N}(m, \sigma^2 I) / m \in \mathbb{R}^n, \sigma > 0 \} \]

then I need to find \(p_{11} \)

\[ \min p_{11} \text{ has the same dist}^{10} \quad P = \mathcal{N}(0, \sigma^2 I) \]

\[ \sigma > 0 \]

Then we define

\[ R_{\min p} = \{ \# / p_{11} \leq F_0^{-1}(\alpha) \} \]

\(F_0\) quantile of \(F_0\)

\[ \text{NB. Useful only if the } p_{11} \text{ are highly correlated} \]

For instance if all the \(p_{11}\)'s are equal to 1 prove \(p\)

(i.e., you're doing \#H times the same test) then

\[ \min p_{11} = p \quad F_0^{-1}(\alpha) = \alpha \quad (\text{since } p \sim \mathcal{U}(0,1) \text{ under } H_0) \]
(I'm considering only $\phi_{+}$) and you will reject more
with $\min p \leq \alpha$ (in this case)
than with $R_{\text{conf}} = \{ h \in \mathcal{X} | \frac{1}{\# \mathcal{X}} \sum_{x \in \mathcal{X}} p_{+}(x) \leq \alpha \}$

Control of $\text{FWER}$

Let $\mathcal{P} \in \mathcal{N}^{+}$ \quad ($\mathcal{X}(\mathcal{P}) = \mathcal{Y}$)

$P(R_{\text{min}} \neq \phi) = P(\exists h \in \mathcal{Y}, p_{+} \leq F^{-1}_{0}(\alpha))$

$= P(\min_{h \in \mathcal{X}} p_{+} \leq F^{-1}_{0}(\alpha))$

$= F_{0}(F_{0}^{-1}(\alpha)) = \alpha$ \quad (we assume it is $\leq \alpha$

$\Rightarrow \sup_{\mathcal{P} \in \mathcal{N}^{+}} P(R_{\text{min}} \neq \phi) \leq \alpha \quad \text{ie FWER}(R_{\text{min}}) \leq \alpha$

Control of $\text{FWER}$

To do so we need another assumption

$\forall \mathcal{G} \subset \mathcal{X}$, the dist of $\min p_{+}$ is the same \forall $h \in \mathcal{H}_{\mathcal{G}}$

Then $P < P$
\[ \mathbb{P}(R_{\min} \cap \mathcal{C}(P) \neq \emptyset) = \mathbb{P}(\exists \mathcal{H} \in \mathcal{C}(P), \mathbb{P}_{\mathcal{H}} \leq F_0^{-1}(\alpha)) \]

\[ = \mathbb{P}(\min_{\mathcal{H} \in \mathcal{C}(P)} \mathbb{P}_{\mathcal{H}} \leq F_0^{-1}(\alpha)) \]

Let's apply \( \otimes \) with \( \mathcal{Y} = \mathcal{C}(P) \)

\[ \forall \mathcal{H} \in \mathcal{H} \cup \mathcal{H} \text{ there is also } \mathcal{Q} \in \mathcal{H} \cup \mathcal{H} \text{ that satisfies } \mathcal{Q} \in \mathcal{H} \cup \mathcal{H} \]

By \( \otimes \), \( \mathbb{P}(R_{\min} \cap \mathcal{C}(P) = \emptyset) = \mathbb{Q}(\min_{\mathcal{H} \in \mathcal{Y}} \mathbb{P}_{\mathcal{H}} \leq F_0^{-1}(\alpha)) \)

(I can do as \( \forall \mathcal{H} \) in \( \mathcal{H} \) one time!)

and \( \leq \mathbb{Q}(\min_{\mathcal{H} \in \mathcal{X}} \mathbb{P}_{\mathcal{H}} \leq F_0^{-1}(\alpha)) = F_0(F_0^{-1}(\alpha)) \)

(because \( \min_{\mathcal{X} \in \mathcal{Y}} \leq \min_{\mathcal{Q} \in \mathcal{Y}} \) = \( \alpha \)

So \( \sup_{P \in \mathcal{P}} \mathbb{P}(R_{\min} \cap \mathcal{C}(P) \neq \emptyset) \leq \alpha \)

\( \alpha \) \( \Rightarrow \) \( \text{FWER}(R_{\min}) \leq \alpha \)
Stepdown is an iterative procedure whose aim is to increase recursively the set $R$ and still keeping a controlled FWER $P$

To do so we need $\mathcal{A} : g \in \mathcal{X} \rightarrow \mathcal{A}(g) \subset \mathcal{X}$

"Next"

$\mathcal{A}$ should satisfy:

Monotony

$\forall g, g' \in \mathcal{X}, \mathcal{A}(g) \subset \mathcal{A}(g')$  \( \forall g, g' \in \mathcal{X}, \mathcal{A}(g) \subset \mathcal{A}(g') \)

False reject control

$\forall P \in \mathcal{P}, P(\mathcal{A}(F(P)) \subset F(P)) > 1-\alpha$

Then the stepdown is defined by

$$
\begin{align*}
R_0 &= \emptyset \\
R_{n+1} &= R_n \cup \mathcal{A}(R_n)
\end{align*}
$$

and $R_{\text{sd}} = \lim_{n \to \infty} R_n$

\[ \text{NB: \# } R_n \text{ is \# } \text{mile so } (R_n) \text{ will be constant after a while.} \]
Control of the FWER

We want for a PEP to control the proba under P of

\[ R_{sd} \cap C(P) \neq \emptyset \]

If this is the case \( \exists n \) (random) such that

\[
\begin{cases} 
    \mathcal{N}(R_n) \cap C(P) \neq \emptyset \\
    R_n \cap C(P) = \emptyset 
\end{cases}
\]

If \( R_n \cap C(P) = \emptyset \) it means that \( R_n \subset F(P) \)

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

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

By the false set control, this happens almost

with probability \( \alpha \)

\[
\Rightarrow \quad P \left( R_{sd} \cap C(P) \neq \emptyset \right) \leq \alpha
\]

and \( \text{FWER}(R_{sd}) \leq \alpha \)

Examples of \( \mathcal{N} \)

Bonferroni
\[ \mathcal{N}_{B}^\mathcal{P}(g) = \{ h \mid \#h \leq \frac{\alpha}{\#x - \#g} \} \]

Monotony: \[ g \leq g' \Rightarrow \#g \leq \#g' \]

Hence: \[ \frac{\alpha}{\#x - \#g} \leq \frac{\alpha}{\#x - \#g'} \]

And: \[ \mathcal{N}(g) \subseteq \mathcal{N}(g') \subseteq \mathcal{N}(g') \cup g' \]

False set control: \[ \forall \mathcal{P} \in \mathcal{P} \]

\[ \mathcal{P}(\mathcal{N}(\mathcal{F}(\mathcal{P})) \cap \mathcal{C}(\mathcal{P}) \neq \emptyset) \]

\[ \leq \mathcal{P}(\exists h \in \mathcal{C}(\mathcal{P}), \#h \leq \frac{\alpha}{\#x - \#\mathcal{F}(\mathcal{P})} \leq \#\mathcal{C}(\mathcal{P}) \]

\[ \leq \exists \mathcal{P}(\forall h \leq \frac{\alpha}{\#\mathcal{C}(\mathcal{P})} \leq \#\mathcal{C}(\mathcal{P}) \]
\[
\frac{\frac{\alpha}{\#R - \#R_2}}{\#R - \#R_2}
\]

\[
\Rightarrow \text{stop } R_{sd} = R_2
\]
d = 30 # 30 coordinates
level = 0.1

X = rnorm(d)
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 removing them from 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 removing them from 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 $\mathcal{N}_m^c$ with min p-value

$$\mathcal{N}_m^c(G) = \{ H \mid p_H \leq F^{-1}_G(x) \}$$

where $F_G$ is the cdf of min $P(H \in C)$ under $P(E \in \mathcal{H})$. 

$$\mathcal{N}_m^c$$
As an exercise, check monotony and falsa et control for

$$R_{sd}$$

Simulation of $$R_{sd}$$ built with $$N_{Bonf}$$

Function to compute $$R_{sd}$$ with $$N_{Bonf}$$

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)
}

### 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

\[ FWER = \frac{R_{sd/\text{Bonf}}}{n} \]

\[ (\text{just for } 1 \text{ P}) \]

\[ \text{FWER} = 0 \]

\[ \text{FWER} = \frac{\text{FWER}}{\text{Nsimu}} \]

\[ \text{Nsimu} = 5000 \]

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

\[ \text{FWER} = \frac{\text{FWER}}{\text{Nsimu}} \]

### comparison ROC curves / TPR /FPR

\[ \text{Nsimu} = 5000 \]

\[ \text{mean} = 5 \]

\[ \text{level} = \text{c} (\text{seq} (0.001, 0.005, \text{by} = 0.001), \text{seq} (0.01, 1, \text{by} = 0.01)) \]
\[ \text{ll} = \text{length} (\text{level}) \]

\[ \text{TPR}_{\text{SDBonf}} = \text{matrix} (0, \text{nrow} = \text{Nsimu}, \text{ncol} = \text{ll}) \]
\[ \text{FPR}_{\text{SDBonf}} = \text{matrix} (0, \text{nrow} = \text{Nsimu}, \text{ncol} = \text{ll}) \]
\[ \text{TPR}_{\text{Bonf}} = \text{matrix} (0, \text{nrow} = \text{Nsimu}, \text{ncol} = \text{ll}) \]
\[ \text{FPR}_{\text{Bonf}} = \text{matrix} (0, \text{nrow} = \text{Nsimu}, \text{ncol} = \text{ll}) \]

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

As you should see that the green curve (SDB) is above the black (Bonf) for TPR/level and FPR/level
It means that one makes more discoveries with SD! (and still control FWER
⇒ +1 is better)

IV False Discovery Rate

1/ Definitions
\[
\text{FDR}(R) = \sup_{P \in \mathcal{P}} \mathbb{E}\left( \frac{\#R_{\text{false}}(P)}{\#R} \right)
\]

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 responsibilities!

**Ex:**

If you need to take a decision and that one false positive has a lot of nasty implications

\[\Rightarrow\text{control the FWR} \quad (\text{medicine, market, military application})\]

If you want to screen and you know that your decisions will be checked again by different means

\[\Rightarrow\text{control the FDR}\]
FDR is less "stringent" guarantee than FWER
(see the image below)

If \( P \in \cap \{ H : \mathbb{E}(P) = x, F(P) = \emptyset \} \)
then \( \mathbb{E}(\frac{\# R_{\text{not}(P)}}{\# R}) = P(R \neq \emptyset) \)
So if there is nothing to discover, both criteria give the same control.

When there is something to discover, there is a tie.

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

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

One of the most cited papers in statistics.

In 20 years, it became the classical method employed in many scientific areas (e.g., genomics).

**Principle:**

1/ Sort the p-values \( P_{(1)} \leq P_{(2)} \leq \cdots \leq P_{(d)} \)

2/ Define \( k = \max \{ i \mid P_{(i)} \leq \frac{\alpha i}{d} \} \) \( (d=\#R) \)

3/ Reject all the indices \( \Rightarrow P_{(1)} \cdots P_{(k)} \)
We can prove that

1/ If the \( p_i \) are independent, \( \text{FDR} (R) \leq \alpha \)

2/ If the \( p_i \) are coming from one-sided test on Gaussian vectors, then \( \text{FDR} (R) \leq \alpha \)

3/ Most of the time even if these assumptions are not fulfilled \( \text{FDR} (R) \leq \alpha \) (Monte Carlo simulation)

4/ However in any case

\[
\text{FDR} (R) \leq \alpha \left( 1 + \frac{1}{2} + \ldots + \frac{1}{d} \right) \approx \log (d)
\]

See interactive Web page of E. Reaven FDR
Function for computing BH

```r
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)
}
```

Code for computing FDR/FWER of BH

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

for(j in 1: Nsimu)
{
  X=rnorm(d)
  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')