

Graphical approaches to parametric and resampling based multiple testing procedures

Florian Klinglmueller

<http://float.lefant.net>

FWF P23167 & J3721

Acknowledgment: Livio Finos

University of Padua, Italy

Department of Statistics

Overview

1. Tests for intersection hypotheses - beyond Bonferroni
2. The graph as a weighting algorithm
3. Putting it all together:
Graphical approaches to parametric and resampling based multiple testing procedures
4. Case Study I: Two treatments vs. control
5. Case Study II: Neuroimaging data

Graphical procedures so far ...

- ▶ ... are based on the **Bonferroni** test:
 - ▶ Controls the Type I error rate in any scenario -
 - ▶ Correlations between test statistics are not considered
 - ▶ Conservative for positively correlated test statistics
- ▶ Often test statistics are positively correlated (e.g. clinical endpoints, absolute changes in gene expression)
- ▶ Sometimes there is even more information:
 - ▶ *Example:* Two doses of a new treatment versus common control → Test statistics correlated - (assuming i.i.d. normal outcomes) correlations are known

Tests for intersection hypotheses beyond Bonferroni

Consider,

- ▶ m elementary null hypotheses H_1, \dots, H_m , and index set $I = \{1, \dots, m\}$;
- ▶ unadjusted p -values p_i from tests of H_i (e.g. t -tests).

Then rejecting the intersection hypotheses $H_I = \bigcap_{i \in I} H_i$ that all H_i $i \in I$ are simultaneously true if any,

$$p_i \leq \frac{\alpha}{m}$$

controls the Type I error rate at level α .

Note: rejecting H_i for all i with $p_i \leq \alpha/m$ also controls the family wise error rate in the strong sense.

Consider,

- ▶ m elementary null hypotheses H_1, \dots, H_m , and index set $I = \{1, \dots, m\}$;
- ▶ unadjusted p -values p_i from tests of H_i (e.g. t -tests).

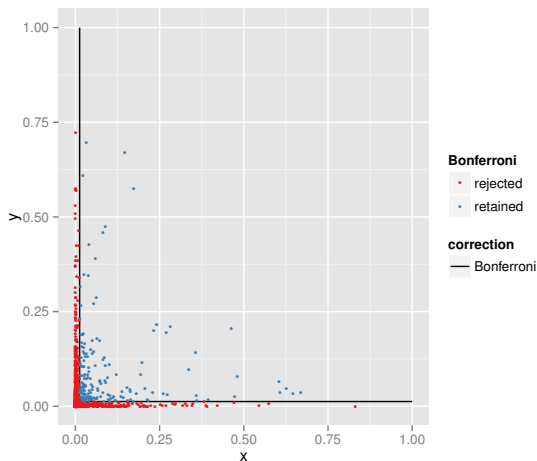
Then rejecting the intersection hypotheses $H_I = \bigcap_{i \in I} H_i$ that all H_i $i \in I$ are simultaneously true if any,

$$p_i \leq \frac{\alpha}{m}$$

controls the Type I error rate at level α .

Note: rejecting H_i for all i with $p_i \leq \alpha/m$ also controls the family wise error rate in the strong sense.

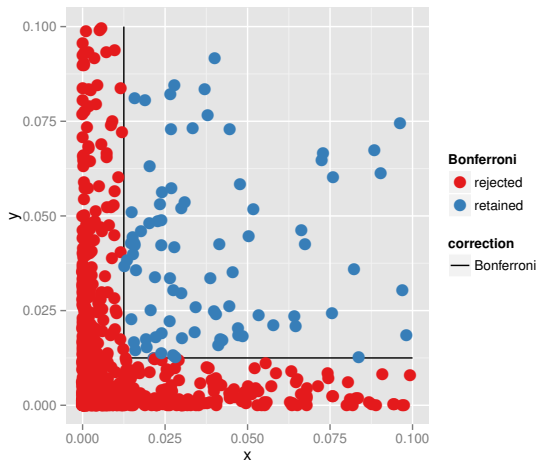
Bonferroni test: One sided, one sample z-Test, $\alpha = .025$



Std. Normal $\mu = (2.4, 2.4)$, Correlation $\rho = 0$, Power .80

Bonferroni test: One sided, one sample

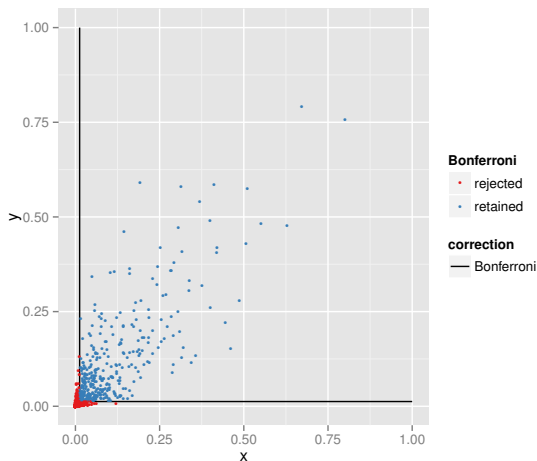
z -Test, $\alpha = .025$



Std. Normal $\mu = (2.4, 2.4)$, Correlation $\rho = 0$, Power .80

Bonferroni test: One sided, one sample

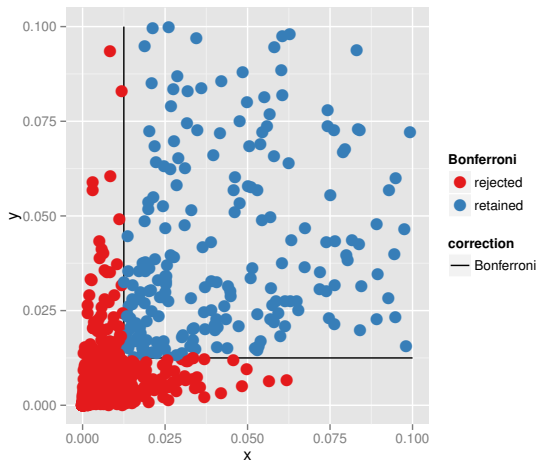
z -Test, $\alpha = .025$



Std. Normal $\mu = (2.4, 2.4)$, Correlation $\rho = .9$, Power .64

Bonferroni test: One sided, one sample

$z\text{-Test}, \alpha = .025$



Std. Normal $\mu = (2.4, 2.4)$, Correlation $\rho = .9$, Power .64

Weighted Bonferroni test

Consider,

- ▶ weights $w_i(I)$, such that $\sum_{i \in I} w_i(I) = 1$;

Then rejecting the intersection hypotheses $H_I = \bigcap_{i \in I} H_i$ that all H_i $i \in I$ are simultaneously true if any,

$$p_i \leq w_i(I)\alpha$$

controls the Type I error rate at level α .

Note: the weights need to be pre-fixed!

Note: rejecting H_i for all i with $p_i \leq w_i(I)\alpha$ also controls the family wise error rate in the strong sense.

Weighted Bonferroni test

Consider,

- ▶ weights $w_i(I)$, such that $\sum_{i \in I} w_i(I) = 1$;

Then rejecting the intersection hypotheses $H_I = \bigcap_{i \in I} H_i$ that all H_i $i \in I$ are simultaneously true if any,

$$p_i \leq w_i(I)\alpha$$

controls the Type I error rate at level α .

Note: the weights need to be pre-fixed!

Note: rejecting H_i for all i with $p_i \leq w_i(I)\alpha$ also controls the family wise error rate in the strong sense.

Weighted Bonferroni test

Consider,

- ▶ weights $w_i(I)$, such that $\sum_{i \in I} w_i(I) = 1$;

Then rejecting the intersection hypotheses $H_I = \bigcap_{i \in I} H_i$ that all H_i $i \in I$ are simultaneously true if any,

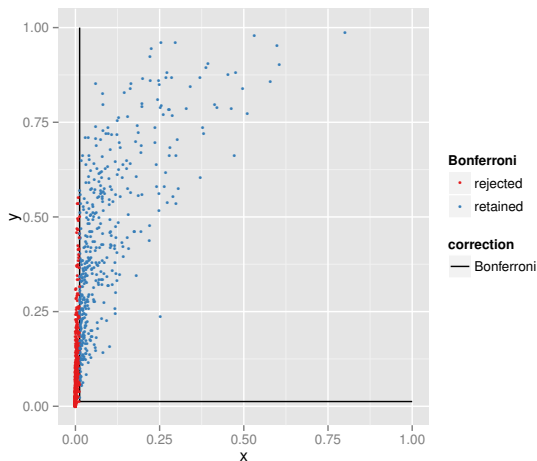
$$p_i \leq w_i(I)\alpha$$

controls the Type I error rate at level α .

Note: the weights need to be pre-fixed!

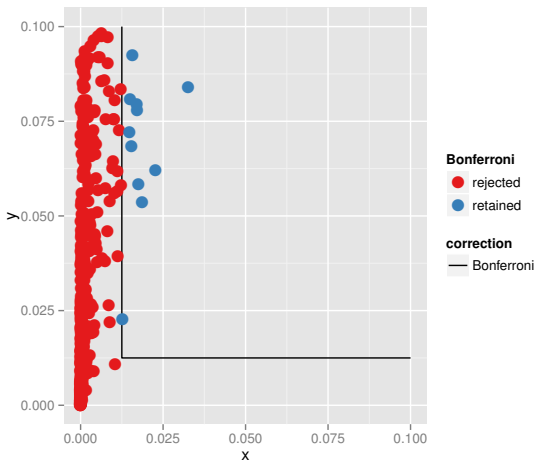
Note: rejecting H_i for all i with $p_i \leq w_i(I)\alpha$ also controls the family wise error rate in the strong sense.

Weighted Bonferroni test: One sided, one sample z-Test, $\alpha = .025$



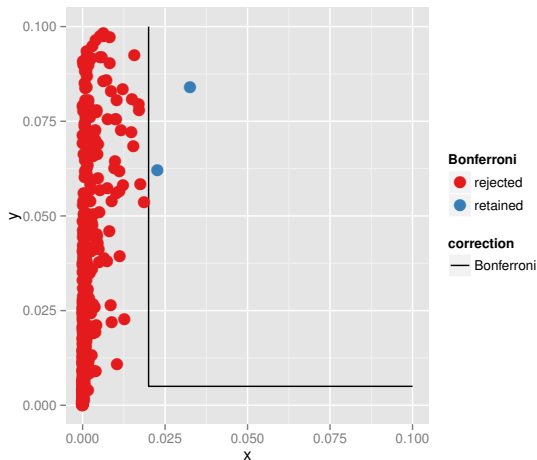
Std. Normal $\mu = (2.4, 1.0)$, Correlation $\rho = .5$, Weights $w = (.5, .5)$, Power .59

Weighted Bonferroni test: One sided, one sample z-Test, $\alpha = .025$



Std. Normal $\mu = (2.4, 1.0)$, Correlation $\rho = .5$, Weights $w = (.5, .5)$, Power .59

Weighted Bonferroni test: One sided, one sample z-Test, $\alpha = .025$



Std. Normal $\mu = (2.4, 1.0)$, Correlation $\rho = .5$, Weights $w = (.8, .2)$, Power .64

Weighted min- p test WESTFALL & YOUNG '93

If the joint distribution of p_i under H_I is known

- ▶ Find the largest c_I such that

$$P_{H_I} \left(\min_{i \in I} \{p_i / w_i(I)\} \leq c_I \alpha \right) = P_{H_I} \left(\bigcup_{i \in I} \{p_i \leq c_I w_i(I) \alpha\} \right) \leq \alpha.$$

- ▶ Reject H_I if there is an $i \in I$ such that $p_i \leq c_I w_i(I) \alpha$.
- ▶ $c_I \geq 1$: factor describing relative gain over Bonferroni obtained from exploiting the parametric model (typically multivariate Normal- or t-distribution with a certain correlation structure).
- ▶ There is a modification of the c_I -calculation for the case that only some correlations are known BRETZ ET AL., '11.

Note: rejecting H_i for all i with $p_i \leq c_I w_i(I) \alpha$ does **not** control the family wise error rate in the strong sense.

Weighted min- p test WESTFALL & YOUNG '93

If the joint distribution of p_i under H_I is known

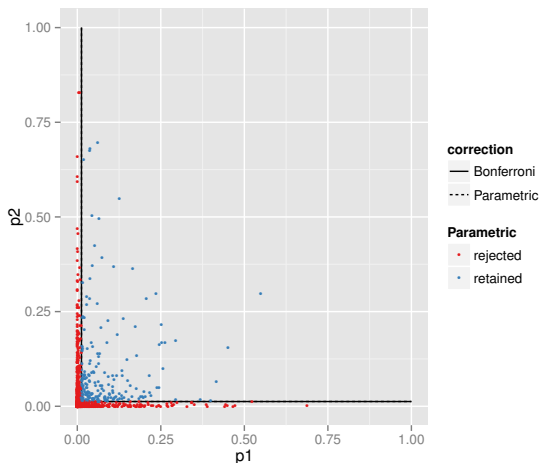
- Find the largest c_I such that

$$P_{H_I} \left(\min_{i \in I} \{p_i / w_i(I)\} \leq c_I \alpha \right) = P_{H_I} \left(\bigcup_{i \in I} \{p_i \leq c_I w_i(I) \alpha\} \right) \leq \alpha.$$

- Reject H_I if there is an $i \in I$ such that $p_i \leq c_I w_i(I) \alpha$.
- $c_I \geq 1$: factor describing relative gain over Bonferroni obtained from exploiting the parametric model (typically multivariate Normal- or t-distribution with a certain correlation structure).
- There is a modification of the c_I -calculation for the case that only some correlations are known BRETZ ET AL., '11.

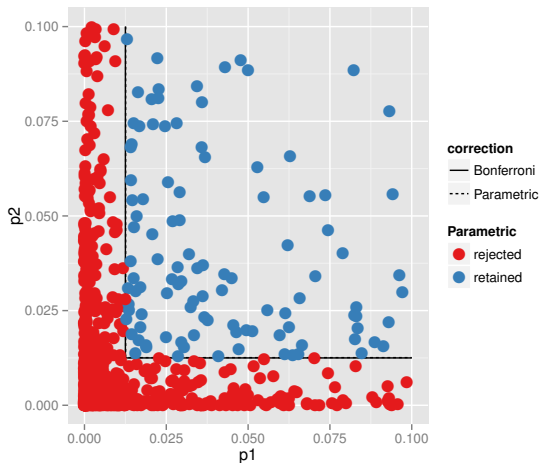
Note: rejecting H_i for all i with $p_i \leq c_I w_i(I) \alpha$ does **not** control the family wise error rate in the strong sense.

Weighted min- p test: One sided, one sample z-Test, $\alpha = .025$



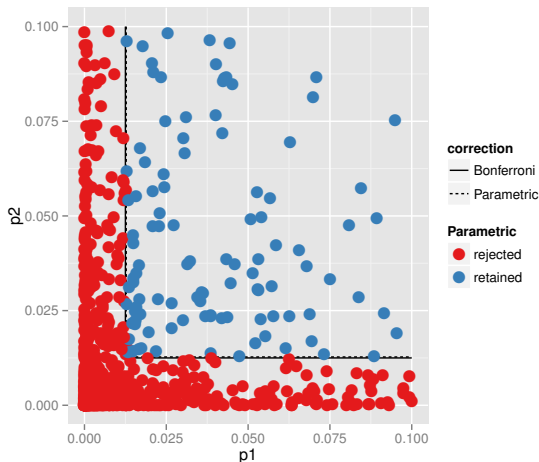
Std. Normal $\mu = (2.4, 2.4)$, Correlation $\rho = 0$, Power .80

Weighted min- p test: One sided, one sample z-Test, $\alpha = .025$



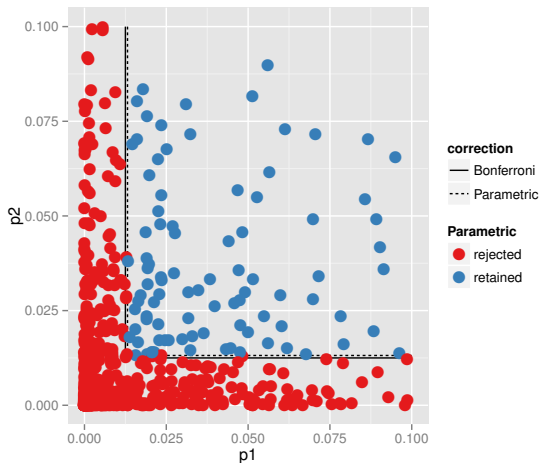
Std. Normal $\mu = (2.4, 2.4)$, Correlation $\rho = 0$, Power .80

Weighted min- ρ test: One sided, one sample z-Test, $\alpha = .025$



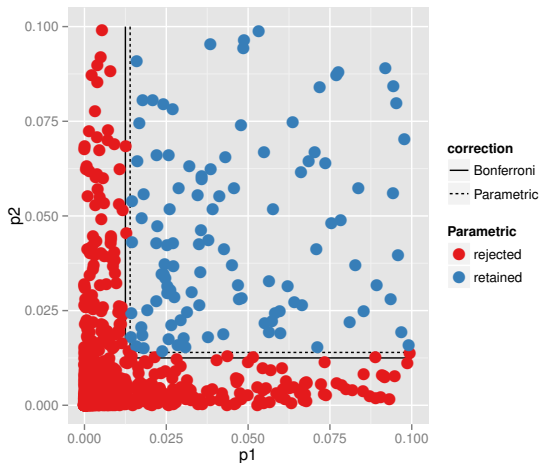
Std. Normal $\mu = (2.4, 2.4)$, Correlation $\rho = .2$, Power .77 vs.
.78

Weighted min- ρ test: One sided, one sample z-Test, $\alpha = .025$



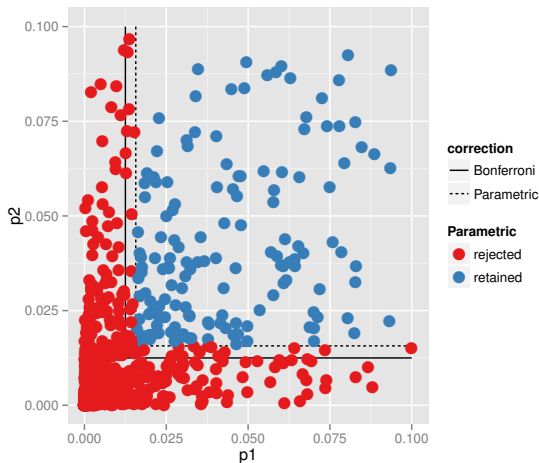
Std. Normal $\mu = (2.4, 2.4)$, Correlation $\rho = .4$, Power .74 vs.
.75

Weighted min- ρ test: One sided, one sample z-Test, $\alpha = .025$



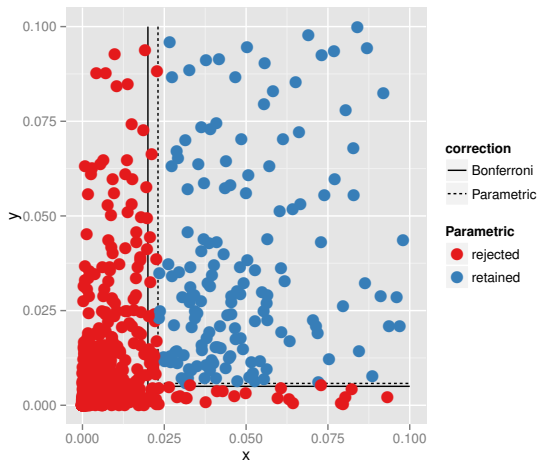
Std. Normal $\mu = (2.4, 2.4)$, Correlation $\rho = .6$, Power .73 vs.
.75

Weighted min- ρ test: One sided, one sample z-Test, $\alpha = .025$



Std. Normal $\mu = (2.4, 2.4)$, Correlation $\rho = .8$, Power .67 vs.
.70

Weighted min- ρ test: One sided, one sample z-Test, $\alpha = .025$



Std. Normal $\mu = (2.4, 2.4)$, Correlation $\rho = .8$, Weights
 $\mathbf{w} = (.8, .2)$

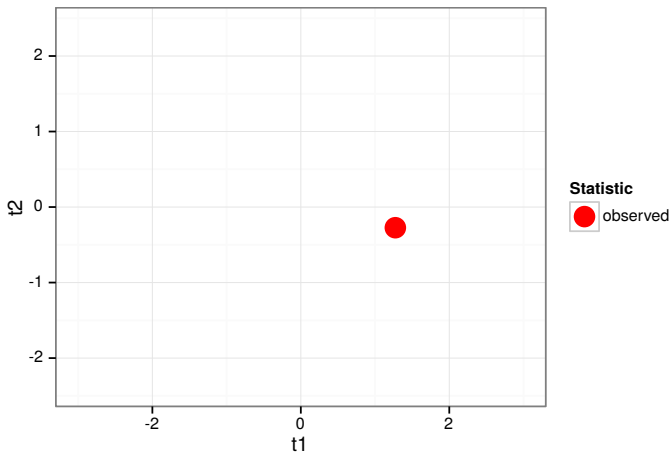
Nonparametric min- p test WESTFALL & YOUNG '93

- ▶ If the joint distribution of p -values is unknown we can use a permutation approach to compute the joint null distribution of the test staticistics/ p -values.
- ▶ Here the null hypotheses is exchangeability of outcome vectors.
- ▶ Based on the joint distribution of p -values the permutation distribution of the minimum (weighted) p -value (or other functionals) can be easily computed.
- ▶ Requires few assumptions about the data generating process.
- ▶ More powerfull than Bonferroni as it uses data-based distributional characteristics (discreteness, correlation structure)
- ▶ Asymptotically as good (or better) as parametric alternatives (e.g., MEINSHAUSEN ET AL. '11)

Nonparametric min- p test - treatment vs. control

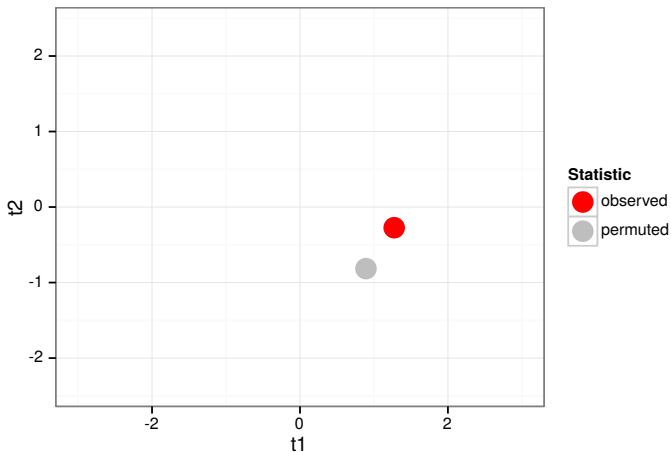
- ▶ m outcome variables $\mathbf{x}_k = (x_{k,1}, \dots, x_{k,m})$ of subjects k in $1, \dots, N$;
- ▶ Treatment assignments $\mathbf{g} = (g_1, \dots, g_N)$, where $g_k \in \{C, T\}$
- ▶ Test statistic $T(\mathbf{X}, \mathbf{g}) = (t_1, \dots, t_m)$ (e.g. t -statistics)
- ▶ Under the null hypotheses that the \mathbf{x}_k exchangeable (*i.e.*, independent of treatment assignment) the joint null distribution is given by the values of $T(\mathbf{X}, \mathbf{g}')$ for all permutations \mathbf{g}' of the treatment assignments
- ▶ The joint distribution of p -values is computed by replacing the values of T for the permuted (observed) data by component-wise quantiles.
- ▶ By taking the minimum over p -values p_i , $i \in I$ for each permutation we get the null distribution of the min- p statistic under H_I .

Permutation test: Illustration (t-statistic)



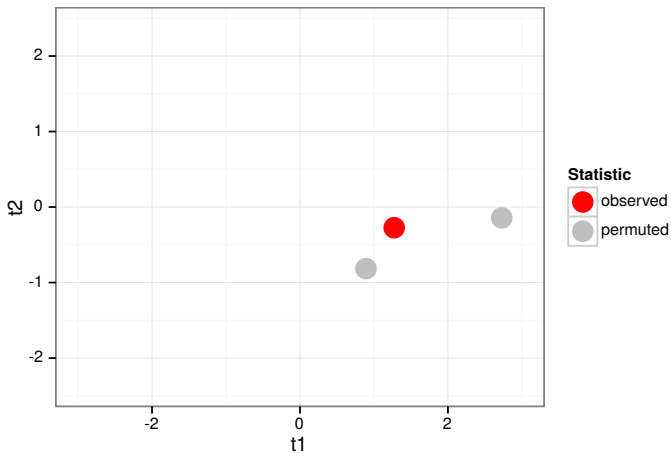
g	T	T	T	C	C	C
x1	0.59	0.71	-0.11	-0.45	0.61	-1.82
x2	0.63	-0.28	-0.28	-0.92	-0.12	1.82

Permutation test: Illustration (t-statistic)



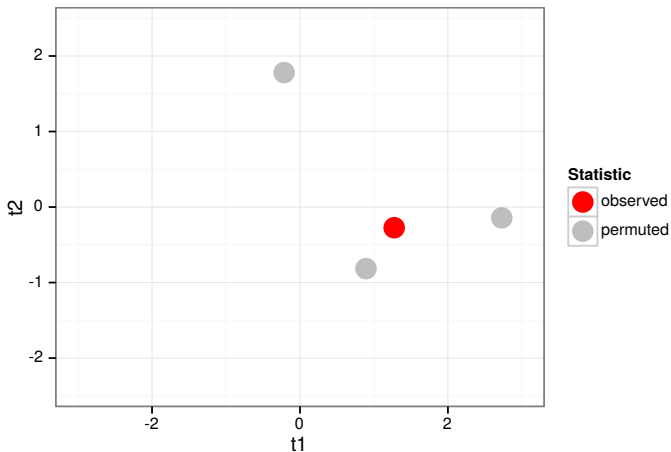
	g	T	T	C	T	C	C
x1	0.59	0.71	-0.11	-0.45	0.61	-1.82	
x2	0.63	-0.28	-0.28	-0.92	-0.12	1.82	

Permutation test: Illustration (t-statistic)



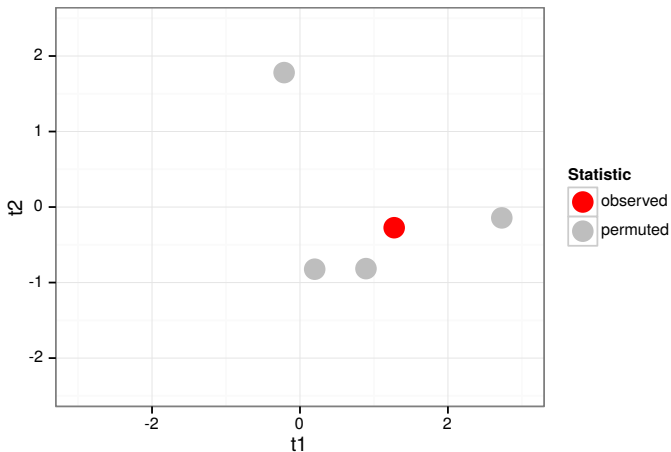
g	T	T	C	C	T	C
x1	0.59	0.71	-0.11	-0.45	0.61	-1.82
x2	0.63	-0.28	-0.28	-0.92	-0.12	1.82

Permutation test: Illustration (t-statistic)



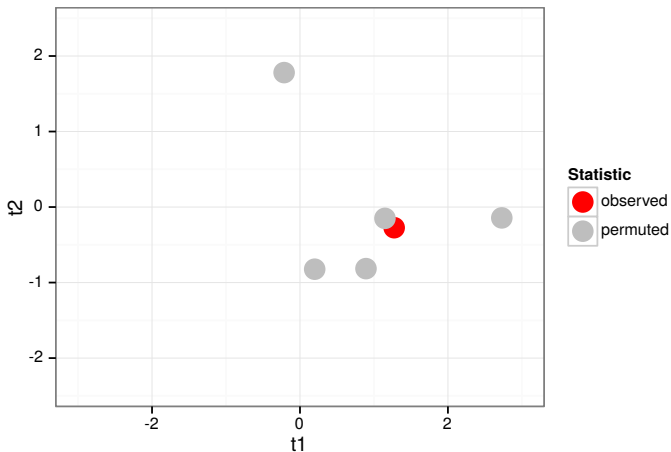
g	T	T	C	C	C	T
x1	0.59	0.71	-0.11	-0.45	0.61	-1.82
x2	0.63	-0.28	-0.28	-0.92	-0.12	1.82

Permutation test: Illustration (t-statistic)



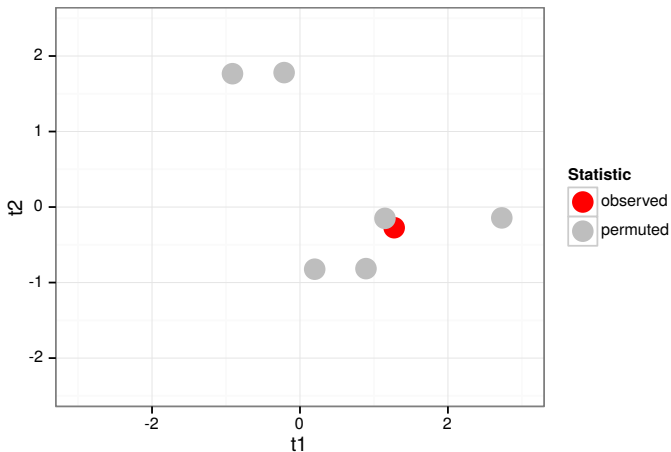
g	T	C	T	T	C	C
x1	0.59	0.71	-0.11	-0.45	0.61	-1.82
x2	0.63	-0.28	-0.28	-0.92	-0.12	1.82

Permutation test: Illustration (t-statistic)



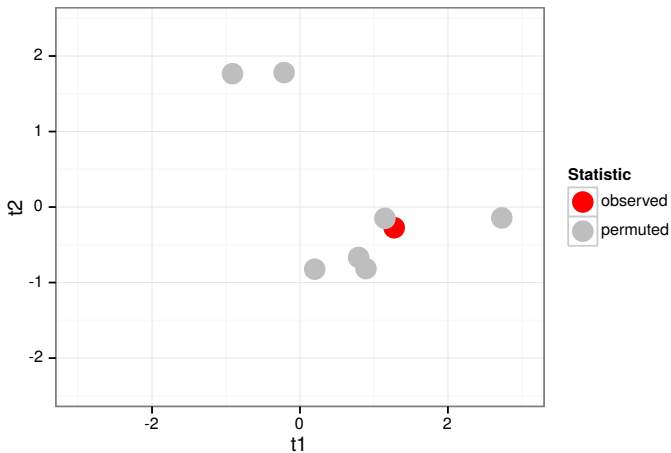
g	T	C	T	C	T	C
x1	0.59	0.71	-0.11	-0.45	0.61	-1.82
x2	0.63	-0.28	-0.28	-0.92	-0.12	1.82

Permutation test: Illustration (t-statistic)



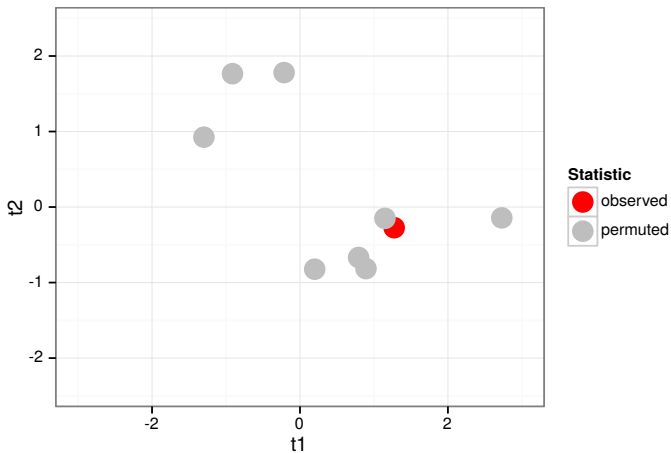
	g	T	C	T	C	C	T
x1	0.59	0.71	-0.11	-0.45	0.61	-1.82	
x2	0.63	-0.28	-0.28	-0.92	-0.12	1.82	

Permutation test: Illustration (t-statistic)



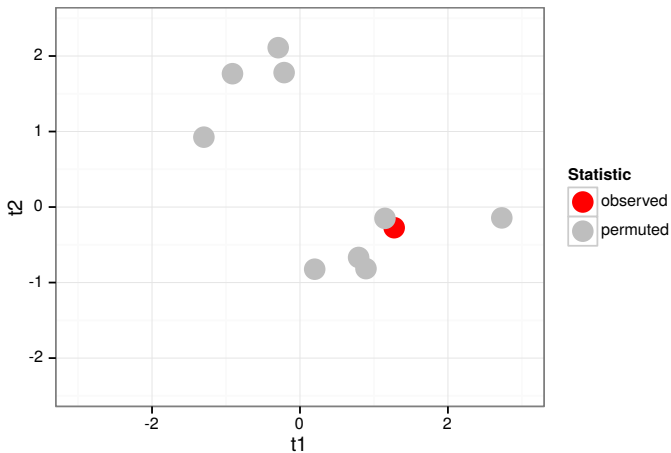
g	T	C	C	T	T	C
x1	0.59	0.71	-0.11	-0.45	0.61	-1.82
x2	0.63	-0.28	-0.28	-0.92	-0.12	1.82

Permutation test: Illustration (t-statistic)



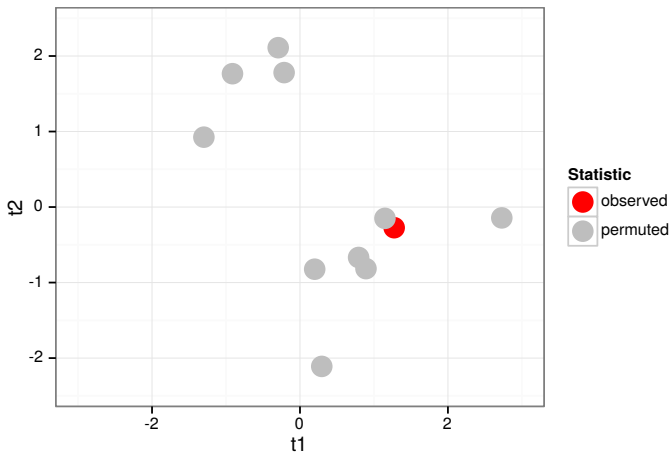
	g	T	C	C	T	C	T
x1		0.59	0.71	-0.11	-0.45	0.61	-1.82
x2		0.63	-0.28	-0.28	-0.92	-0.12	1.82

Permutation test: Illustration (t-statistic)



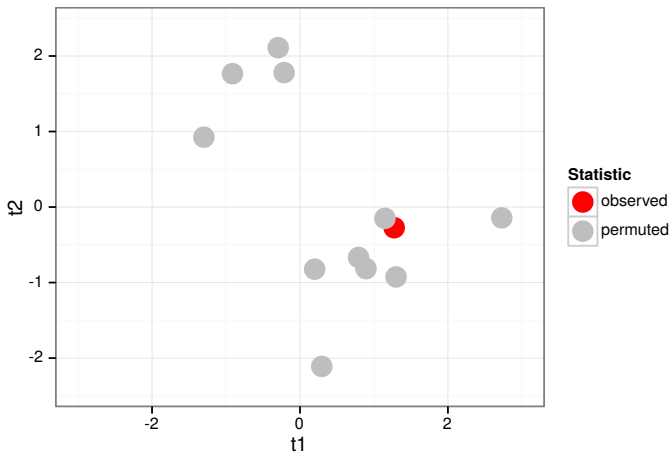
g	T	C	C	C	T	T
x1	0.59	0.71	-0.11	-0.45	0.61	-1.82
x2	0.63	-0.28	-0.28	-0.92	-0.12	1.82

Permutation test: Illustration (t-statistic)



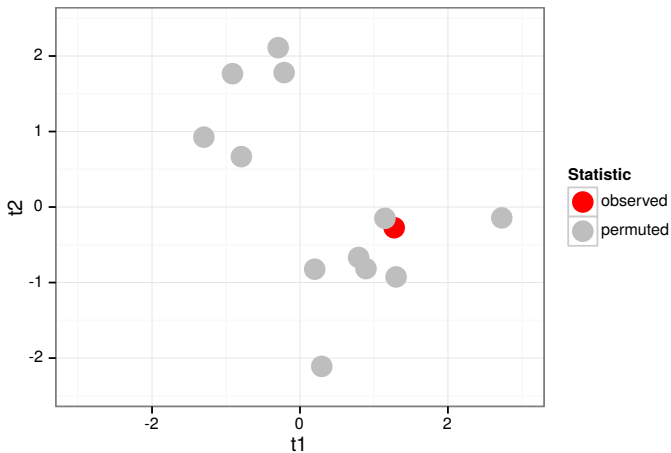
	g	C	T	T	T	C	C
x1		0.59	0.71	-0.11	-0.45	0.61	-1.82
x2		0.63	-0.28	-0.28	-0.92	-0.12	1.82

Permutation test: Illustration (t-statistic)



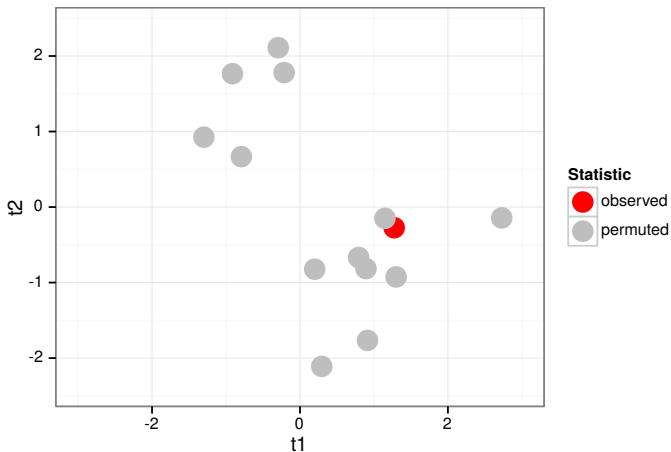
	g	C	T	T	C	T	C
x1		0.59	0.71	-0.11	-0.45	0.61	-1.82
x2		0.63	-0.28	-0.28	-0.92	-0.12	1.82

Permutation test: Illustration (t-statistic)



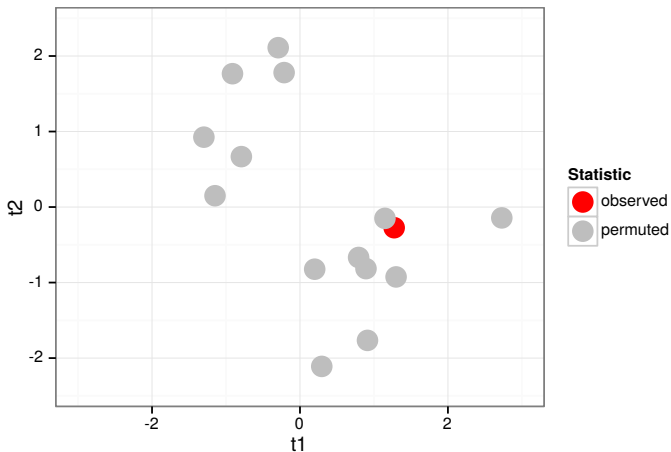
g	C	T	T	C	C	T
x1	0.59	0.71	-0.11	-0.45	0.61	-1.82
x2	0.63	-0.28	-0.28	-0.92	-0.12	1.82

Permutation test: Illustration (t-statistic)



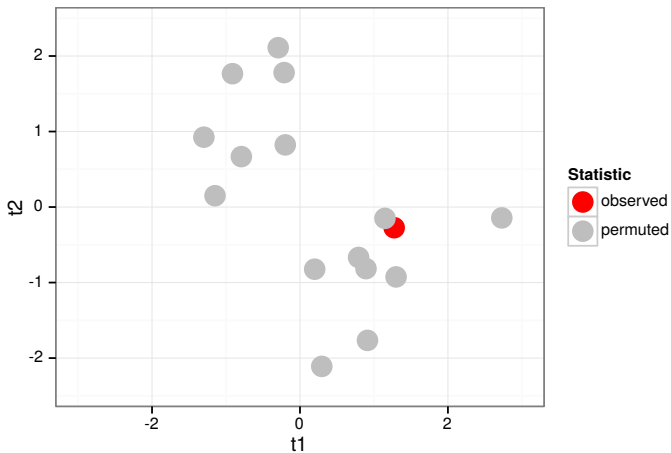
g	C	T	C	T	T	C
x1	0.59	0.71	-0.11	-0.45	0.61	-1.82
x2	0.63	-0.28	-0.28	-0.92	-0.12	1.82

Permutation test: Illustration (t-statistic)



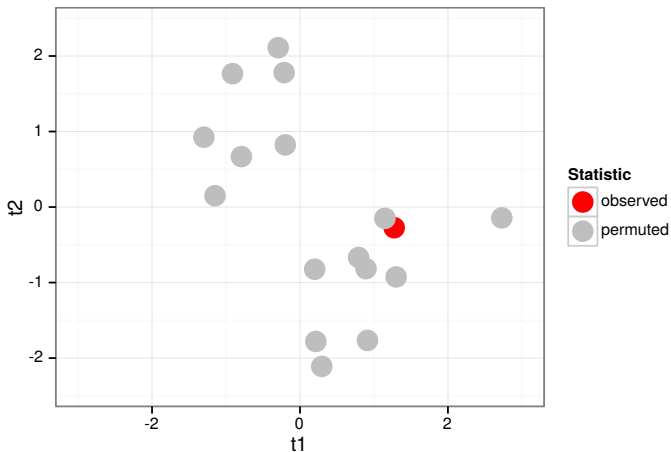
g	C	T	C	T	C	T
x1	0.59	0.71	-0.11	-0.45	0.61	-1.82
x2	0.63	-0.28	-0.28	-0.92	-0.12	1.82

Permutation test: Illustration (t-statistic)



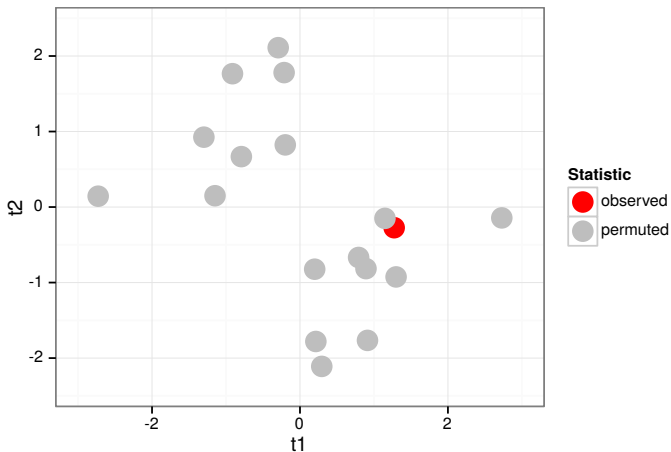
g	C	T	C	C	T	T
x1	0.59	0.71	-0.11	-0.45	0.61	-1.82
x2	0.63	-0.28	-0.28	-0.92	-0.12	1.82

Permutation test: Illustration (t-statistic)



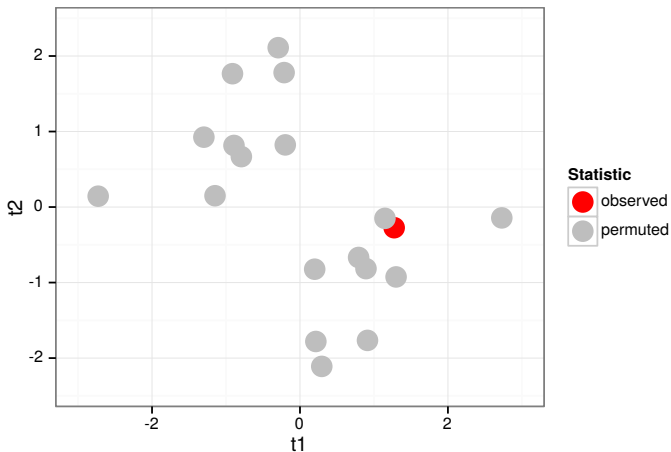
g	C	C	T	T	T	C
x1	0.59	0.71	-0.11	-0.45	0.61	-1.82
x2	0.63	-0.28	-0.28	-0.92	-0.12	1.82

Permutation test: Illustration (t-statistic)



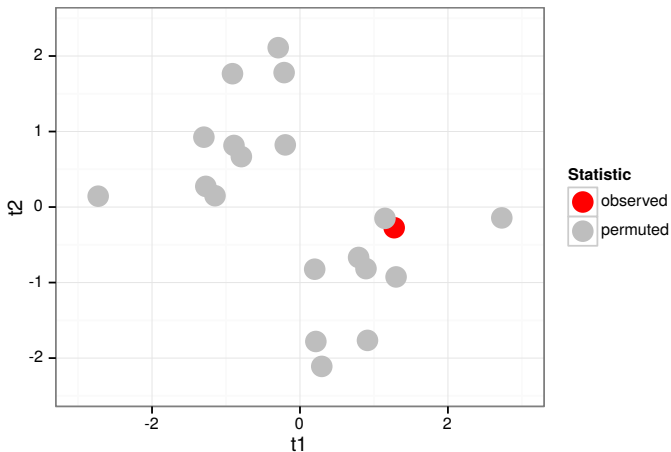
g	C	C	T	T	C	T
x1	0.59	0.71	-0.11	-0.45	0.61	-1.82
x2	0.63	-0.28	-0.28	-0.92	-0.12	1.82

Permutation test: Illustration (t-statistic)



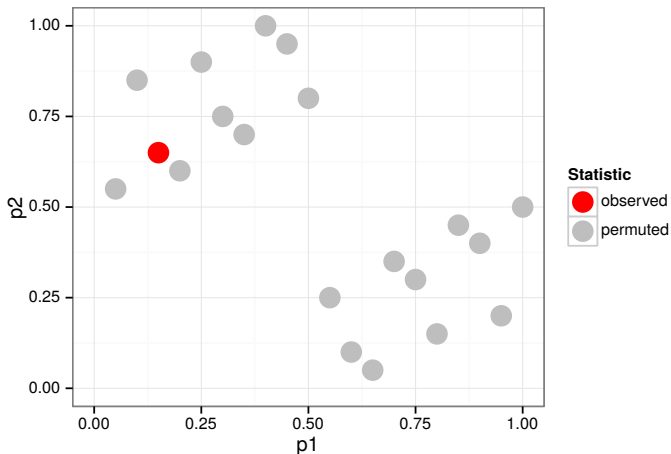
	g	C	C	T	C	T	T
x1	0.59	0.71	-0.11	-0.45	0.61	-1.82	
x2	0.63	-0.28	-0.28	-0.92	-0.12	1.82	

Permutation test: Illustration (t-statistic)



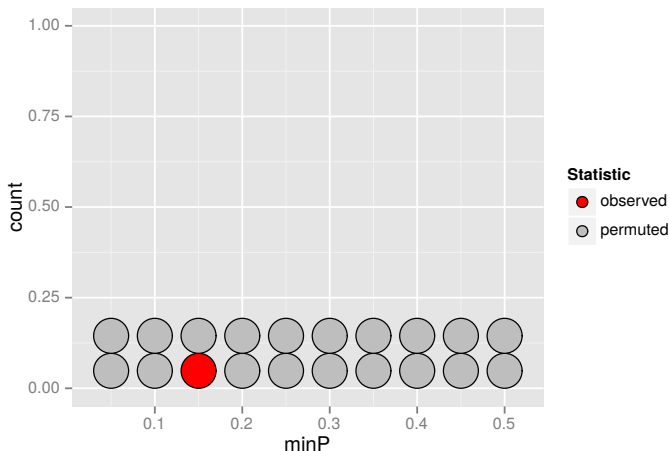
g	C	C	C	T	T	T
x1	0.59	0.71	-0.11	-0.45	0.61	-1.82
x2	0.63	-0.28	-0.28	-0.92	-0.12	1.82

Permutation test: Illustration (p-values)



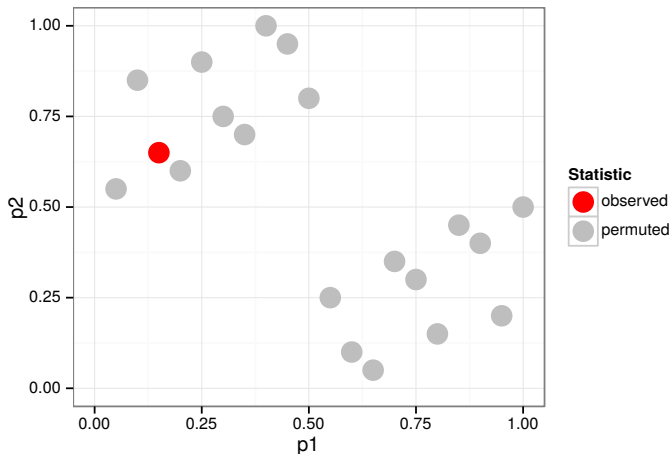
	g	C	C	C	T	T	T
x1		0.59	0.71	-0.11	-0.45	0.61	-1.82
x2		0.63	-0.28	-0.28	-0.92	-0.12	1.82

Permutation test: min- p



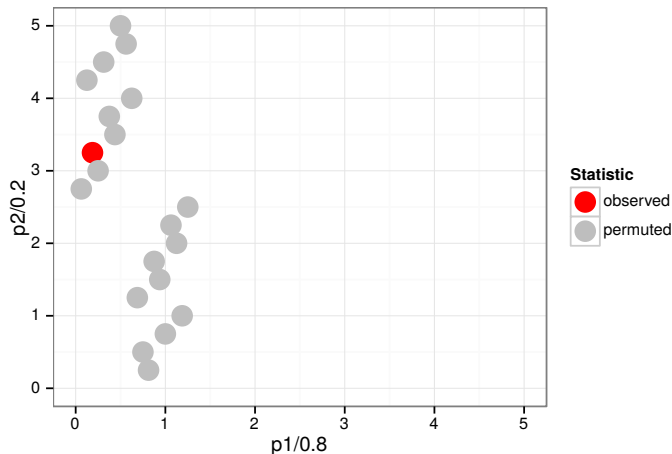
6 out of 20 permutation min- p values are equal or smaller to the observed min- p -value (.15). p -value of test of $H_1 \cap H_2$ is .3.

Permutation test: weighted min- p



Divide p_1 by weight of .8 and p_2 by weight of .2.
 p -value of test of $H_1 \cap H_2$ is .15.

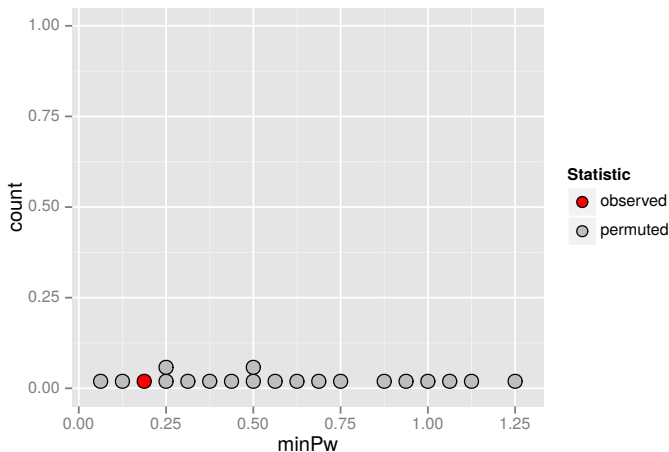
Permutation test: weighted min- p



Divide p_1 by weight of .8 and p_2 by weight of .2.

p -value of test of $H_1 \cap H_2$ is .15.

Permutation test: weighted min- p



Divide p_1 by weight of .8 and p_2 by weight of .2.
 p -value of test of $H_1 \cap H_2$ is .15.

Nonparametric combination PESARIN & SALMASO '10

Based on the joint permutation null distribution of test-statistics/p-values we may compute the distribution of many other combinations of component-wise test-statistics/p-values:

maxT: $\max_{i \in I} w_i(I) t_i$

Fisher: $-2 * \sum_{i \in I} w_i(I)^{(-1)} \ln p_i(X, g)$

... Direct sum, Inverse Normal, Mahalanobis, ...

Note: rejection of H_I using a nonparametric combination test does **not** provide us with a test of elementary hypotheses H_i .

Note: the combination function needs to be prespecified, and can in general not be chosen data dependent

Nonparametric combination PESARIN & SALMASO '10

Based on the joint permutation null distribution of test-statistics/p-values we may compute the distribution of many other combinations of component-wise test-statistics/p-values:

maxT: $\max_{i \in I} w_i(I) t_i$

Fisher: $-2 * \sum_{i \in I} w_i(I)^{(-1)} \ln p_i(X, g)$

... Direct sum, Inverse Normal, Mahalanobis, ...

Note: rejection of H_I using a nonparametric combination test does **not** provide us with a test of elementary hypotheses H_i .

Note: the combination function needs to be prespecified, and can in general not be chosen data dependent

Nonparametric combination PESARIN & SALMASO '10

Based on the joint permutation null distribution of test-statistics/p-values we may compute the distribution of many other combinations of component-wise test-statistics/p-values:

maxT: $\max_{i \in I} w_i(I) t_i$

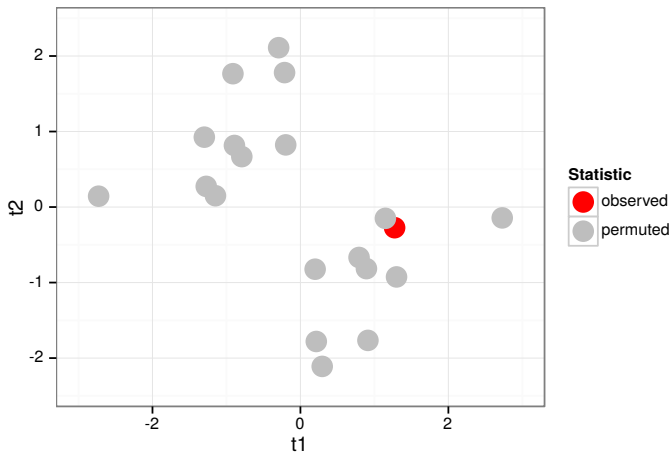
Fisher: $-2 * \sum_{i \in I} w_i(I)^{(-1)} \ln p_i(X, g)$

... Direct sum, Inverse Normal, Mahalanobis, ...

Note: rejection of H_I using a nonparametric combination test does **not** provide us with a test of elementary hypotheses H_i .

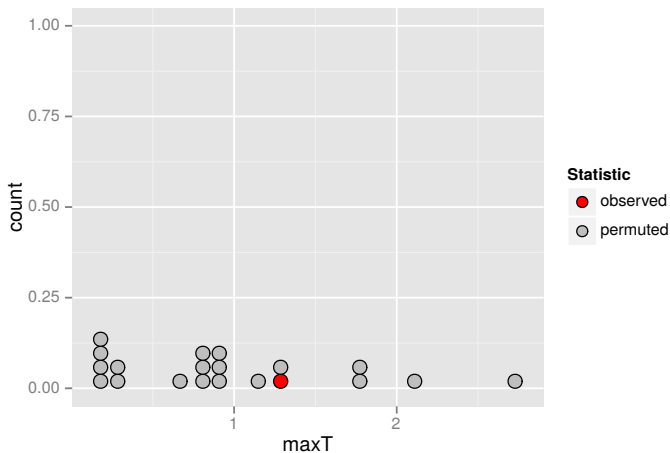
Note: the combination function needs to be prespecified, and can in general not be chosen data dependent

Nonparametric combination: Illustration



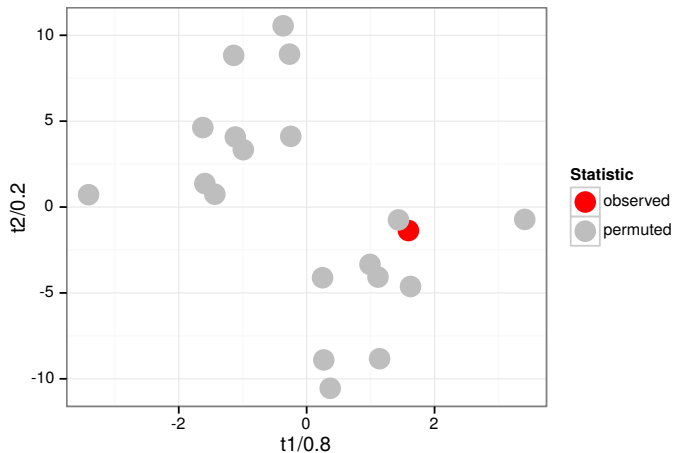
maxT: p -value of test of $H_1 \cap H_2$ is .3.

Nonparametric combination: Illustration



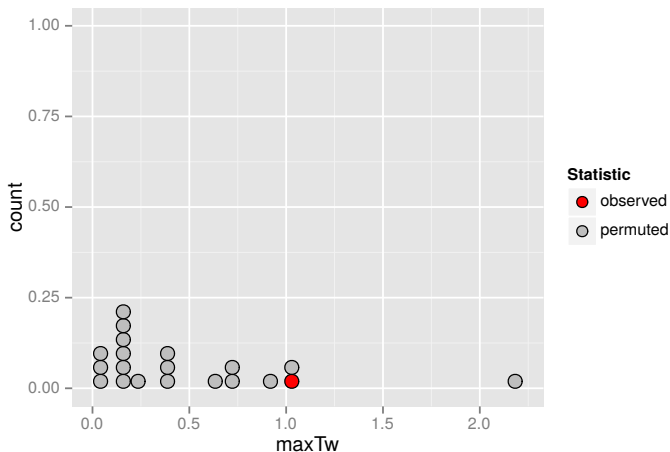
maxT: p -value of test of $H_1 \cap H_2$ is .3.

Nonparametric combination: Illustration



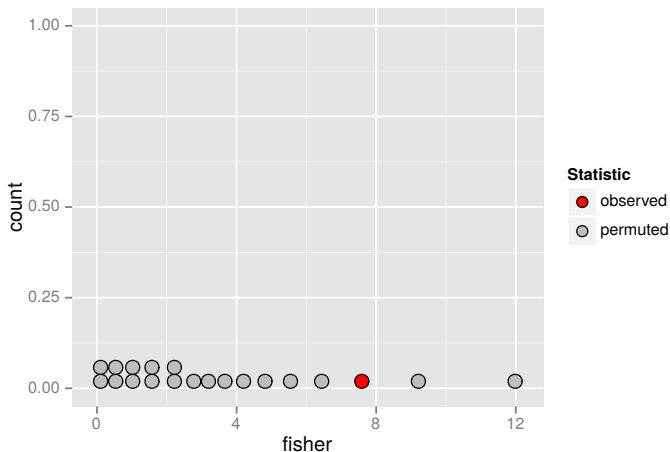
Weighted maxT: p -value of test of $H_1 \cap H_2$ is .15.

Nonparametric combination: Illustration



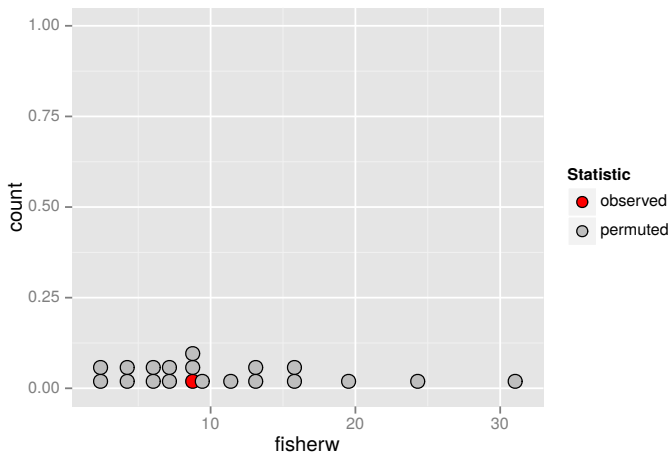
Weighted maxT: p -value of test of $H_1 \cap H_2$ is .15.

Nonparametric combination: Illustration



Fisher combination: p -value of test of $H_1 \cap H_2$ is .15.

Nonparametric combination: Illustration



Weighted Fisher combination: p -value of test of $H_1 \cap H_2$ is .15.

**The graph as a way to generate
weights**

General definition of the graph

General definition of the weights

- ▶ $\mathbf{w} = (w_1, \dots, w_m)$, $\sum_{i=1}^m w_i = 1$, initial weights
- ▶ $\mathbf{G} = (g_{ij}) : m \times m$ transition matrix
 g_{ij} with $0 \leq g_{ij} \leq 1$, $g_{ii} = 0$ and $\sum_{j=1}^m g_{ij} = 1$ for all $i = 1, \dots, m$.
- ▶ g_{ij} , fraction of the level of H_i that is allocated to H_j .
- ▶ \mathbf{G} and α fully determine the graph.

General weighting algorithm

To obtain the weights for some subset J of $\{1, \dots, m\}$ set $I = \{1, \dots, m\}$.

1. If $I = J$ stop.
2. Let j be any $j \in I \setminus J$
3. Update the graph:

$$I \rightarrow I / \{j\}$$
$$w_\ell \rightarrow \begin{cases} w_\ell + w_j g_{j\ell}, & \ell \in I \\ 0, & \text{otherwise} \end{cases}$$
$$g_{\ell k} \rightarrow \begin{cases} \frac{g_{\ell k} + g_{ej} g_{jk}}{1 - g_{ej} g_{je}}, & \ell, k \in I, \ell \neq k, g_{lj} g_{jl} < 1 \\ 0, & \text{otherwise} \end{cases}$$

4. Go to step 1.

Note: The order in which we select elements from $I \setminus J$ does not matter, the resulting weights are the same.

General weighting algorithm

To obtain the weights for some subset J of $\{1, \dots, m\}$ set $I = \{1, \dots, m\}$.

1. If $I = J$ stop.
2. Let j be any $j \in I \setminus J$
3. Update the graph:

$$I \rightarrow I / \{j\}$$
$$w_\ell \rightarrow \begin{cases} w_\ell + w_j g_{j\ell}, & \ell \in I \\ 0, & \text{otherwise} \end{cases}$$
$$g_{\ell k} \rightarrow \begin{cases} \frac{g_{\ell k} + g_{ej} g_{jk}}{1 - g_{ej} g_{je}}, & \ell, k \in I, \ell \neq k, g_{lj} g_{jl} < 1 \\ 0, & \text{otherwise} \end{cases}$$

4. Go to step 1.

Note: The order in which we select elements from $I \setminus J$ does not matter, the resulting weights are the same.

General weighting algorithm

To obtain the weights for some subset J of $\{1, \dots, m\}$ set $I = \{1, \dots, m\}$.

1. If $I = J$ stop.
2. Let j be any $j \in I \setminus J$
3. Update the graph:

$$I \rightarrow I / \{j\}$$
$$w_\ell \rightarrow \begin{cases} w_\ell + w_j g_{j\ell}, & \ell \in I \\ 0, & \text{otherwise} \end{cases}$$
$$g_{\ell k} \rightarrow \begin{cases} \frac{g_{\ell k} + g_{\ell j} g_{jk}}{1 - g_{\ell j} g_{j\ell}}, & \ell, k \in I, \ell \neq k, g_{\ell j} g_{j\ell} < 1 \\ 0, & \text{otherwise} \end{cases}$$

4. Go to step 1.

Note: The order in which we select elements from $I \setminus J$ does not matter, the resulting weights are the same.

General weighting algorithm

To obtain the weights for some subset J of $\{1, \dots, m\}$ set $I = \{1, \dots, m\}$.

1. If $I = J$ stop.
2. Let j be any $j \in I \setminus J$
3. Update the graph:

$$I \rightarrow I / \{j\}$$
$$w_\ell \rightarrow \begin{cases} w_\ell + w_j g_{j\ell}, & \ell \in I \\ 0, & \text{otherwise} \end{cases}$$
$$g_{\ell k} \rightarrow \begin{cases} \frac{g_{\ell k} + g_{\ell j} g_{jk}}{1 - g_{\ell j} g_{j\ell}}, & \ell, k \in I, \ell \neq k, g_{\ell j} g_{j\ell} < 1 \\ 0, & \text{otherwise} \end{cases}$$

4. Go to step 1.

Note: The order in which we select elements from $I \setminus J$ does not matter, the resulting weights are the same.

General weighting algorithm

To obtain the weights for some subset J of $\{1, \dots, m\}$ set $I = \{1, \dots, m\}$.

1. If $I = J$ stop.
2. Let j be any $j \in I \setminus J$
3. Update the graph:

$$I \rightarrow I / \{j\}$$
$$w_\ell \rightarrow \begin{cases} w_\ell + w_j g_{j\ell}, & \ell \in I \\ 0, & \text{otherwise} \end{cases}$$
$$g_{\ell k} \rightarrow \begin{cases} \frac{g_{\ell k} + g_{\ell j} g_{jk}}{1 - g_{\ell j} g_{j\ell}}, & \ell, k \in I, \ell \neq k, g_{\ell j} g_{j\ell} < 1 \\ 0, & \text{otherwise} \end{cases}$$

4. Go to step 1.

Note: The order in which we select elements from $I \setminus J$ does not matter, the resulting weights are the same.

General weighting algorithm

To obtain the weights for some subset J of $\{1, \dots, m\}$ set $I = \{1, \dots, m\}$.

1. If $I = J$ stop.
2. Let j be any $j \in I \setminus J$
3. Update the graph:

$$I \rightarrow I / \{j\}$$
$$w_\ell \rightarrow \begin{cases} w_\ell + w_j g_{j\ell}, & \ell \in I \\ 0, & \text{otherwise} \end{cases}$$
$$g_{\ell k} \rightarrow \begin{cases} \frac{g_{\ell k} + g_{\ell j} g_{j k}}{1 - g_{\ell j} g_{j \ell}}, & \ell, k \in I, \ell \neq k, g_{\ell j} g_{j \ell} < 1 \\ 0, & \text{otherwise} \end{cases}$$

4. Go to step 1.

Note: The order in which we select elements from $I \setminus J$ does not matter, the resulting weights are the same.

Case Study I: 2 primary, 2 secondary hypotheses

New drug for the treatment of multiple sclerosis

- ▶ Two active treatment arms (high dose given once per day, low dose given 3 times per day), one placebo control arm
- ▶ Primary endpoint annualized relapse rate: H_1, H_2
- ▶ Secondary endpoint number of lesions in the brain: H_3, H_4

Testing Strategy

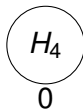
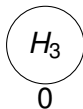
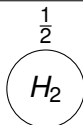
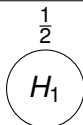
- ▶ Rejection of secondary hypotheses is only of interest if at least one of the primary hypotheses can be rejected
- ▶ Assuming equal efficacy the two treatments should have same probability of success.

Example: Tailoring the procedure



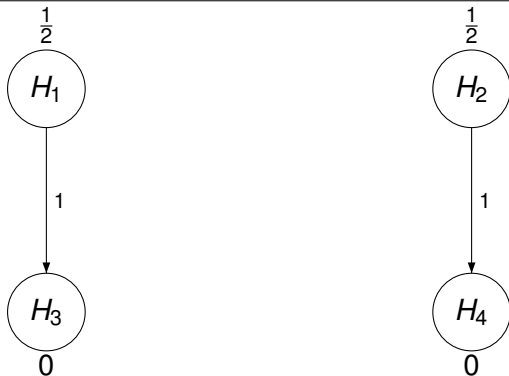
1. Split α equally between primary hypotheses
2. Give no α to secondary hypotheses
3. Reallocate significance levels to secondary hypotheses
4. Reallocate significance levels between treatment arms

Example: Tailoring the procedure



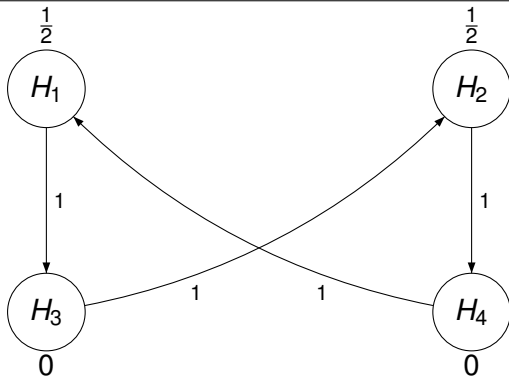
1. Split α equally between primary hypotheses
2. Give no α to secondary hypotheses
3. Reallocate significance levels to secondary hypotheses
4. Reallocate significance levels between treatment arms

Example: Tailoring the procedure



1. Split α equally between primary hypotheses
2. Give no α to secondary hypotheses
3. Reallocate significance levels to secondary hypotheses
4. Reallocate significance levels between treatment arms

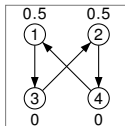
Example: Tailoring the procedure



1. Split α equally between primary hypotheses
2. Give no α to secondary hypotheses
3. Reallocate significance levels to secondary hypotheses
4. Reallocate significance levels between treatment arms

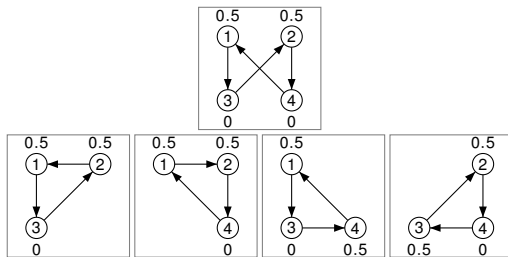
Case Study I: Graphical weighting algorithm

Compute weights for all subsets J of $\{1, \dots, 4\}$



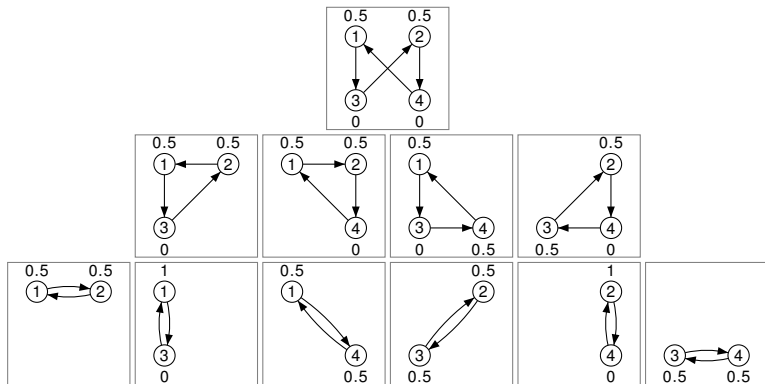
Case Study I: Graphical weighting algorithm

Compute weights for all subsets J of $\{1, \dots, 4\}$



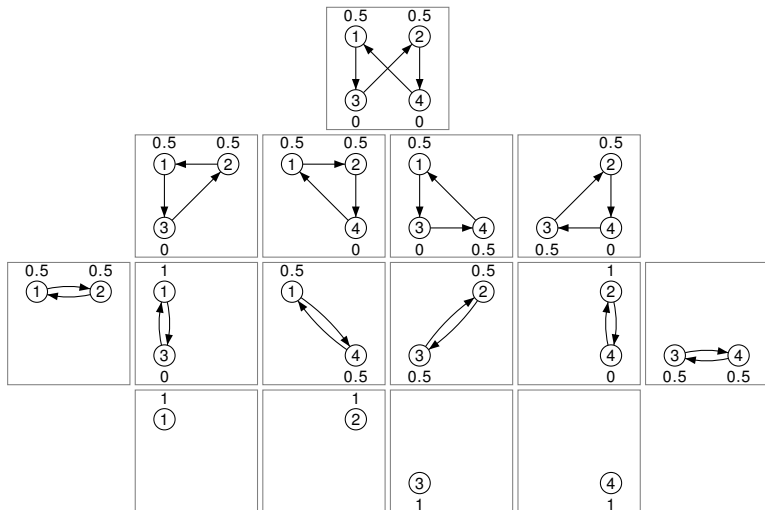
Case Study I: Graphical weighting algorithm

Compute weights for all subsets J of $\{1, \dots, 4\}$



Case Study I: Graphical weighting algorithm

Compute weights for all subsets J of $\{1, \dots, 4\}$



Graph-based multiple testing algorithms

Putting it all together

- ▶ The graph defines weights for each intersection H_J
 $J \subseteq \{1, \dots, m\}$
(BRETZ, POSCH, GLIMM, KLINGLMUELLER, MAURER, ROHMEYER '11)
- ▶ Parametric min- p or nonparametric combination tests provide weighted level α tests for all intersection hypotheses

The Closure Principle (MARCUS ET AL. 1976)

- ▶ Consider m elementary Hypotheses $H_j, j \in \{1, \dots, m\}$
- ▶ For $J \subseteq \{1, \dots, m\}$ let $H_J = \bigcap_{i \in J} H_i$.
- ▶ For each H_J define a level α test
- ▶ Reject H_i if all H_J for which $i \in J$ can be rejected at level α

Putting it all together

- ▶ The graph defines weights for each intersection H_J
 $J \subseteq \{1, \dots, m\}$
(BRETZ, POSCH, GLIMM, KLINGLMUELLER, MAURER, ROHMEYER '11)
- ▶ Parametric min- p or nonparametric combination tests provide weighted level α tests for all intersection hypotheses

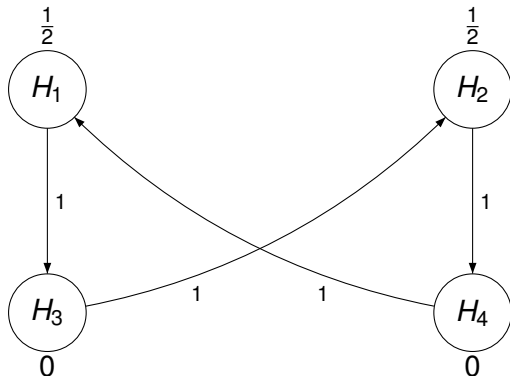
The Closure Principle (MARCUS ET AL. 1976)

- ▶ Consider m elementary Hypotheses $H_j, j \in \{1, \dots, m\}$
- ▶ For $J \subseteq \{1, \dots, m\}$ let $H_J = \bigcap_{i \in J} H_i$.
- ▶ For each H_J define a level α test
- ▶ Reject H_i if **all H_J for which $i \in J$ can be rejected** at level α

Closed test of weighted intersection tests

- ▶ Provides strong FWE control at level α
- ▶ However, in the worst case $2^m - 1$ hypotheses have to be tested
- ▶ Consonance property of the weighted Bonferroni test (*i.e.*, rejecting an intersection hypothesis implies that also an individual hypothesis can be rejected) provides a shortcut so that at most m hypotheses have to be tested (HOMMEL '07)
- ▶ For weighted parametric and multivariate permutation tests we have no shortcut - computationally hard for more than 20 hypothesis (~ 1 million intersection hypotheses)

Case Study I (ctd.)



1. Two treatments vs. control, parallel fixed sequence procedure

Case Study I: Parametric approach

- ▶ Assuming equal numbers of patients in the control and the two treatment groups under the null hypothesis of no treatment effect the mean differences between control and either of the two treatments have correlation 1/2 for both endpoints.
- ▶ The correlation between mean differences across endpoints is not known.
- ▶ Assuming normal observations with known (unknown) common standard deviations the z (t) statistics follow a multivariate normal (t) distribution with correlation matrix:

$$\begin{bmatrix} 1 & 1/2 & \rho & 1/2\rho \\ 1/2 & 1 & 1/2\rho & \rho \\ \rho & 1/2\rho & 1 & 1/2 \\ 1/2\rho & \rho & 1/2 & 1 \end{bmatrix}$$

where ρ is unknown.

Case Study I: package gMCP

The screenshot displays the gMCP GUI interface. On the left, the 'Graph' tab shows a directed graph with four nodes: H1 (top-left), H2 (top-right), H3 (bottom-left), and H4 (bottom-right). Edges connect H1 to H3, H1 to H4, H2 to H3, and H2 to H4, with weights of 1 on each edge. Below the graph is the 'Description' and 'Analysis' section, containing literature references.

On the right, the 'Transition Matrix' tab shows a 4x4 matrix with all values set to 0. Below this, the 'Hypothesis' section lists H1, H2, H3, and H4 with their respective weights (0.5, 0.5, 0, 0) and p-values (0.0126, 0.0126, 0.2, 0.01). The 'Sum of weights' is 1, and the 'Total α ' is 0.025. Buttons for 'Reject and pass α ' are provided for each hypothesis. A 'Load p-values from R' button is also present.

At the bottom right, there is a section for correlation matrices with a dropdown menu set to 'corr. sim', and buttons for 'Refresh' and 'Create Matrix'.

R-package **gMCP** with GUI (`graphGUI()`) ROHMEYER & KLINGLMUELLER '14

Case Study I: package gMCP

The screenshot displays the gMCP software interface, which is used for multiple comparison procedures. It is divided into several panels:

- Graph Panel:** Shows a directed graph with four nodes labeled H1, H2, H3, and H4. Each node has a weight of 1. The graph structure is as follows:


```

      graph TD
        H1((H1  
1)) --> H3((H3  
0))
        H1 --> H4((H4  
0))
        H2((H2  
1)) --> H3
        H2 --> H4
      
```
- Transition Matrix Panel:** A table showing the transition matrix for the hypotheses. The rows are labeled H1, H2, H3, and H4, and the columns are labeled H1, H2, H3, and H4. The matrix is:

	H1	H2	H3	H4
H1	0	0	0	0
H2	0	0	0	0
H3	0	0	0	0
H4	0	0	0	0
- Hypothesis Settings Panel:** Allows setting weights and p-values for each hypothesis.

Hypothesis	Weights	P-Value	Action
H1	0.5	0.0126	Reject and pass α
H2	0.5	0.0126	Reject and pass α
H3	0	0.2	Reject and pass α
H4	0	0.01	Reject and pass α

 The sum of weights is 1, and the total α is 0.025.
- Description Panel:** Contains a description of the graph and a list of references.

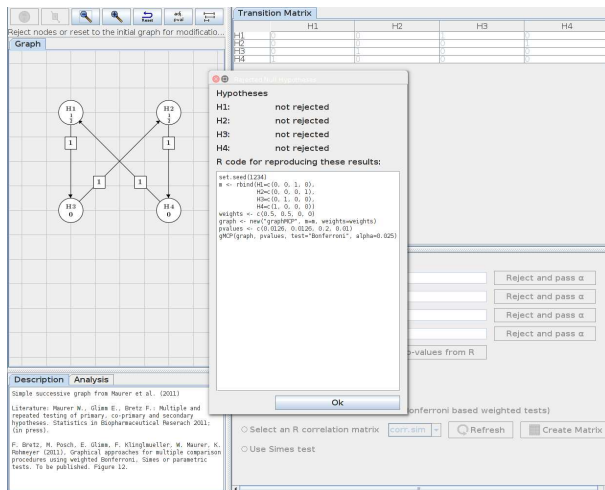
Simple successive graph from Maurer et al. (2011)

Literature: Maurer W., Glm E., Bretz F.: Multiple and repeated testing of primary, co-primary and secondary hypotheses. Statistics in Biopharmaceutical Research 2011: (in press).

F. Bretz, M. Posch, E. Glm, F. Klinglmaier, W. Maurer, K. Rehmeyer (2011). Graphical approaches for multiparameter comparison procedures using weighted Bonferroni, Simes or parametric tests. To be published. Figure 12.
- Analysis Panel:** Contains options for selecting an R correlation matrix (currently set to 'corr.slm'), a Refresh button, and a Create Matrix button. It also includes a note: "No information about correlations (Bonferroni based weighted tests)".

Paint graph, set weights, set p-values

Case Study I: package gMCP



Bonferroni based graphical procedure

Case Study I: package gMCP

GeneralBlock DiagonalTreatments and Endpoints

Save matrix as:

Reset to identity matrix

Change 0 in matrix to NA

You can reorder the hypotheses by drag'n'drop:

H1
H2
H3
H4

Apply reordering

	H1	H2	H3	H4
H1	1	0.5	NA	NA
H2	0.5	1	NA	NA
H3	NA	NA	1	0.5
H4	NA	NA	0.5	1

Save matrix to R

Define correlation matrix

Case Study I: package gMCP

The screenshot displays the gMCP software interface, which is used for graphical procedures based on parametric min- p tests. The interface is divided into several panels:

- Graph Panel:** Shows a simple successive graph with four nodes (H1, H2, H3, H4) and four edges. The nodes are arranged in a 2x2 grid, with H1 and H2 at the top and H3 and H4 at the bottom. The edges are labeled with '1'.
- Transition Matrix Panel:** Displays a 4x4 matrix with columns H1, H2, H3, and H4. The matrix values are:

	H1	H2	H3	H4
H1	0	0	1	0
H2	0	0	0	1
H3	0	1	0	0
H4	1	0	0	0
- Hypotheses Panel:** Shows the results of the tests:
 - H1: rejected
 - H2: rejected
 - H3: not rejected
 - H4: rejected
- R code for reproducing these results:**

```
set.seed(1234)
a <- rbind(c(0, 0, 1, 0),
           c(0, 0, 0, 1),
           c(0, 1, 0, 0),
           c(1, 0, 0, 0))
weights <- c(0.5, 0.5, 0, 0)
graph <- new("graphMCP", a=a, weights=weights)
pvalues <- c(0.026, 0.026, 0.2, 0.0)
cr <- rbind(c(1, 0.5, 0.3, 0.15),
            c(0.5, 1, 0.15, 0.3),
            c(0.3, 0.15, 1, 0.5),
            c(0.15, 0.3, 0.5, 1))
gMCP(graph, pvalues, test="parametric", correlation=cr, alpha=0.025)
```
- Description Panel:** Contains text about the literature and the test procedure.

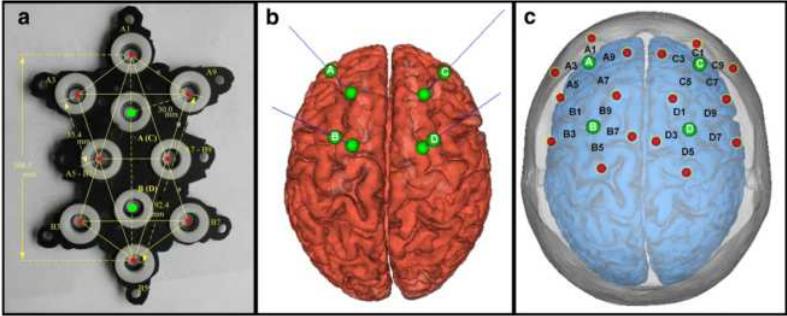
Literature: Maurer W., Glim E., Bretz F.: Multiple and repeated testing of primary, co-primary and secondary hypotheses. Statistics in Biopharmaceutical Research 2011. (in press).

F. Bretz, M. Pisch, E. Glim, F. Klinglmaier, W. Maurer, K. Rohmeyer (2011). Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes or parametric tests. To be published. Figure 12.

Graphical procedure based on parametric min- p test

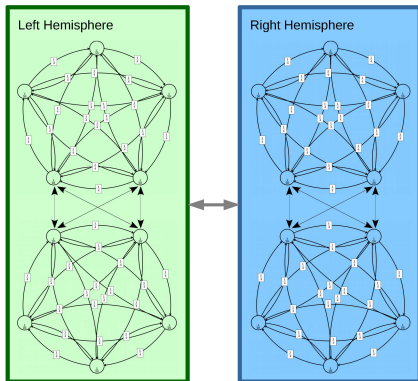
Case Study II: fNIRS data CUTINI ET AL. '13

- ▶ 11 participants had to perform a sequence of enumeration tasks
- ▶ They were shown images with a number of dots ranging from 2 to 8 - quickly had to enter the number of dots they counted
- ▶ Hypothesis: up to 4 dots brain counts, over 4 dots combination of counting and estimation
- ▶ Measurement device: multi-channel near-infrared spectoscopy with 20 channels (10 left and 10 right hemisphere).
- ▶ Data: paired (from 11 patients) average (across repeated tasks) differences (between counting tasks above and below 4 dots) of peak response in 20 fNIRS channels



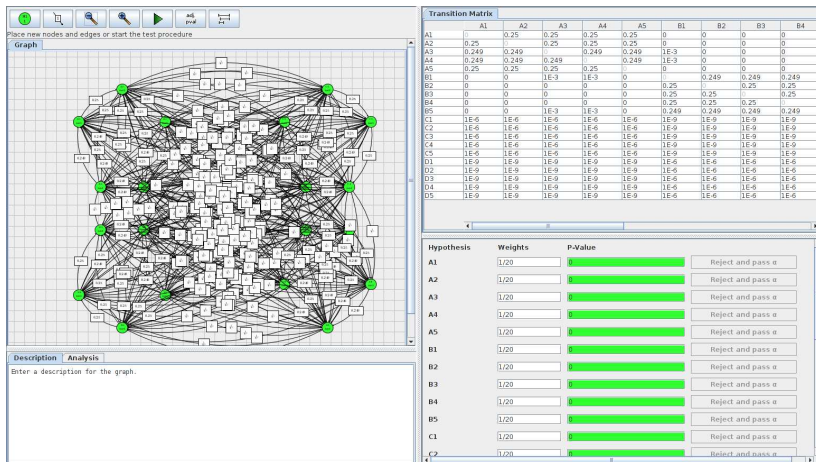
- ▶ 2x2 Light sources
- ▶ 2x8 Sensors
- ▶ 2 Sensors in the middle provide readings for each light source

Case Study II: Graph



- ▶ Split α between hemispheres
- ▶ Equal weight for each hypothesis
- ▶ Bonferroni-Holm type within sensors of each light source
- ▶ Middle sensors exchange weights
- ▶ Hemispheres exchange weights only if all are rejected in one hemisphere

Case Study II: Implementation



- ▶ R-package **gMCP** for graphical weighting (with some help from the command-line) ROHMEYER & KLINGLMUELLER '14
- ▶ R-package **flip** for permutation tests and NPC FINOS '13

Case Study II: Test procedures

One sided paired permutation tests for each of the 20 channels using the t -statistic, FWER to be controlled at level 5%

1. Test unadjusted permutation p -values using Bonferroni based gMCP
2. Test unadjusted permutation p -values using Simes based gMCP
3. Unweighted permutation based adjustment using Westfall & Young maxT
4. Graph-based multivariate permutation test using Fisher combination function

Case Study II: Results

- ▶ No significant results in left hemisphere
- ▶ Test right hemisphere only at $\alpha/2$ (0.025)

	Bonferroni	Simes	maxT	fisher
C1	0.214	0.344	0.261	0.151
C2	0.049	0.059	0.040	0.018
C3	0.053	0.059	0.066	0.020
C7	0.140	0.140	0.105	0.065
C8	0.010	0.010	0.001	0.008
D3	0.214	0.344	0.261	0.214
D4	0.059	0.063	0.053	0.025
D5	0.144	0.144	0.152	0.050
D6	0.094	0.094	0.053	0.019
D7	0.211	0.245	0.248	0.095

Summary & Conclusions

- ▶ The weighted directed graph completely defines the multiple testing procedure.
- ▶ Parametric tests exploit knowledge of the joint distribution of test statistics.
- ▶ Permutation tests use correlation of test statistics even if it is not known.
- ▶ Multiplicity from different sources can be adjusted for.

Further Work

- ▶ Possible shortcuts
- ▶ Extension to adaptive designs
- ▶ Confidence intervals



C.E. Bonferroni.

Il calcolo delle assicurazioni su gruppi di teste.

Tipografia del Senato, 1935.



F. Bretz, M. Posch, E. Glimm, F. Klinglmueller, W. Maurer, and K. Rohmeyer.

Graphical approaches for multiple comparison procedures using weighted bonferroni, simes, or parametric tests.

Biometrical Journal, 2011.



Simone Cutini, Pietro Scatturin, Sara Basso Moro, and Marco Zorzi.

Are the neural correlates of subitizing and estimation dissociable? an fnirs investigation.

Neuroimage, 85:391–399, 2014.



Livio Finos.

flip: Multivariate Permutation Tests, r package version 2.4.3 edition, 2014.



G. Hommel, F. Bretz, and W. Maurer.

Powerful short-cuts for multiple testing procedures with special reference to gatekeeping strategies.

Statistics in Medicine, 26(22):4063–4073, 2007.



Nicolai Meinshausen, Marloes H Maathuis, Peter Bühlmann, et al.

Asymptotic optimality of the westfall–young permutation procedure for multiple testing under dependence.

The Annals of Statistics, 39(6):3369–3391, 2011.



R. Marcus, E. Peritz, and K.R. Gabriel.

On closed testing procedures with special reference to ordered analysis of variance.

Biometrika, 63(3):655–660, 1976.



Fortunato Pesarin.

Multivariate permutation tests: with applications in biostatistics.
Wiley, Chichester, 2001.



K. Rohmeyer, F. Klinglmueller, and B. Bornkamp.

gmcp: Graph based multiple comparison procedures, 2014.

GNU R-package, Version 0.8-6.



P.H. Westfall and S.S. Young.

Resampling-based multiple testing: Examples and methods for p-value adjustment, volume 279.

Wiley-Interscience, 1993.