Weierstrass Institute for
Applied Analysis and Stochastics

Simultaneous Statistical Inference

Thorsten Dickhaus

WIAS Berlin, Research Group “Stochastic Algorithms and Nonparametric Statistics“

Content

We assume a statistical model (statistical experiment)
$(\Omega, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$

More concrete scenario (chosen for exemplary purposes):

Balanced ANOVA1 model:

$X = (X_{ij})_{i=1,\dots,k,\ j=1,\dots,n}, X_{ij} = \mathcal{N}(\mu_i, \sigma^2)$,
X_{ij} stochastically independent random variables on $\mathbb{R}$,

$\mu_i \in \mathbb{R}\ \forall 1 \le i \le k, \sigma^2 > 0$ (known or unknown) variance
$k \ge 3, n \ge 2, \nu = k(n-1)$ (degrees of freedom)

$\Omega = \mathbb{R}^{k\cdot n}, \mathcal{F} = \mathbb{B}^{k\cdot n}$
$\vartheta = (\mu_1, \dots, \mu_k, \sigma^2) \in \mathbb{R}^k \times (0, \infty) = \Theta$

$\mathcal{H}_m = (H_i)_{i=1,\ldots,m}$ Family of null hypotheses with $\emptyset \neq H_i \subset \Theta$ and alternatives $K_i = \Theta \setminus H_i$

$(\Omega, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}, \mathcal{H}_m)$ multiple test problem

$\varphi = (\varphi_i : i = 1, \ldots, m)$ multiple test for $\mathcal{H}_m$

	Test decision		
Hypotheses	0	1	
true	U_m	V_m	m_0
false	T_m	S_m	m_1
	W_m	R_m	m

Wanted features of φ:

No (or only minor) contradictions among the individual test decisions

Control of the probability of erroneous decisions

Concepts of Contradiction: Let $I = \{1, \ldots, m\}$.

- The multiple test φ is coherent if $\forall i, j \in I$ with $H_i \subseteq H_j$:
  $\{\varphi_j = 1\} \subseteq \{\varphi_i = 1\}$.
- The multiple test φ is consonant if $\forall i \in I$ with $\exists j \in I : H_i \subset H_j$:
  $\{\varphi_i = 1\} \subseteq \bigcup_{j: H_j \supset H_i} \{\varphi_j = 1\}$.

(Type I) error measures / concepts:

$\text{FWER}_m(\varphi) = \mathbb{P}_\vartheta(V_m > 0) \overset{!}{\leq} \alpha \quad \forall \vartheta \in \Theta$

(Strong) control of the Family-Wise Error Rate (FWER)

$\text{FDR}_m(\varphi) = \mathbb{E}_\vartheta \left[\frac{V_m}{R_m \vee 1} \right] \overset{!}{\leq} \alpha \quad \forall \vartheta \in \Theta$

Control of the False Discovery Rate (FDR)

References to multiple testing

FWER-controlling multiple testing principles:

- Historical single-step tests (Bonferroni, Šidák, Tukey, Scheffé, Dunnett)
- Closed test principle (Marcus, Peritz, Gabriel (1976))
- Intersection-union principle (generalized closed testing)
- Partitioning principle (Finner and Straßburger 2001, Hsu 1996)
- Projection methods under asymptotic normality (Bretz, Hothorn and Westfall)
- Multivariate methods based on copula modeling (Dickhaus et al. 2013/2014)
- Resampling-based multiple testing (Westfall & Young 1993, Dudoit & van der Laan 2008)

FDR-controlling multiple tests:

- Linear step-up (under positive dependency, Benjamini and Hochberg 1995)
- Adaptive LSU (BH 2000, Storey 2002, 2003, 2004, Benjamini, Krieger, Yekutieli 2006, Sarkar 2008, Blanchard and Roquain (2008, 2009), ...)
- Asymptotically optimal rejection curve
  (under independence; Finner, Dickhaus, Roters 2009)
- Resampling-based multiple testing (Dudoit & van der Laan 2008)

Multiplicity of current applications

Due to rapid technical developments in many scientific fields, the number m of hypotheses to be tested simultaneously can nowadays become **almost arbitrary large:**

- Genetics, microarrays: $m \sim 30{,}000$ genes / hypotheses
- Genetics, SNPs: $m \sim 500{,}000$ SNPs / hypotheses
- Proteomics: $m \sim 5000$ proteine spots per gele sheet
- Cosmology: Signal detection, $m \sim 10^6$ pixels / hypotheses
- Neurology: Identification of active brain loci, $m \sim 10^3$ voxels
- Biometry: pairwise comparisons of many means, tests for correlations

Analyses have (in a first step) typically **explorative character.**

$\Longrightarrow$ Control of the FWER much too conservative goal!

Θ	Parameter space
$H_1, \ldots, H_m$	Null hypotheses
$\varphi = (\varphi_1, \ldots, \varphi_m)$	Multiple test procedure
$V_m = \lvert\{i : \varphi_i = 1 \text{ and } H_i \text{ true }\}\rvert$	Number of falsely rejected, true nulls
$R_m = \lvert\{i : \varphi_i = 1\}\rvert$	Total number of rejections

$$\mathrm{FDR}_\vartheta(\varphi) = \mathbb{E}_\vartheta\left[\frac{V_m}{R_m \vee 1}\right]$$

False Discovery Rate (FDR) given $\vartheta \in \Theta$

Definition: Let $\alpha \in (0, 1)$ fixed.

The multiple test φ controls the FDR at level α if

$$\mathrm{FDR}(\varphi) = \sup_{\vartheta \in \Theta} \mathrm{FDR}_\vartheta(\varphi) \leq \alpha.$$

$p_1, \ldots, p_m$ (marginal) p-values for hypotheses $H_1, \ldots, H_m$

H_i true for $i \in I_{m,0}$, H_i false for $i \in I_{m,1}$

$I_{m,0} + I_{m,1} = \mathbb{N}_m = \{1, \ldots, m\}$, $m_0 = |I_{m,0}|$

$p_i \sim \text{UNI}[0,1],\ i \in I_{m,0}$	stochastically independent	(I1)
$(p_i : i \in I_{m,0}), (p_i : i \in I_{m,1})$	stoch. independent vectors	(I2)
$p_{1:m} \leq \cdots \leq p_{m:m}$	ordered p-values	

Linear step-up procedure φ^{LSU} with Simes' crit. values $\alpha_{i:m} = i\alpha/m$:

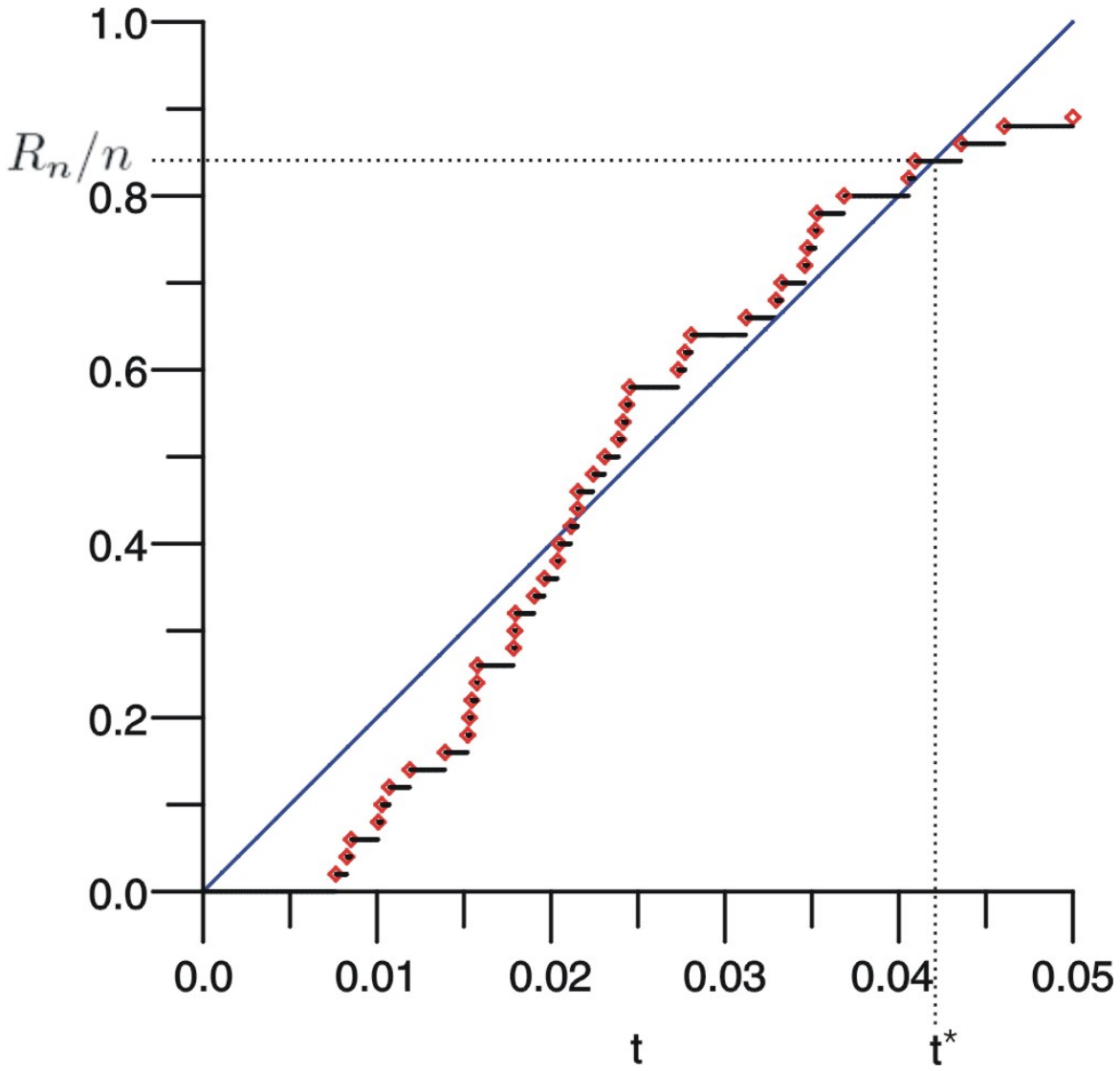
Reject all H_i with $p_i \leq \alpha_{\overline{m}:m}$, where $\overline{m} = \max\{j : p_{j:m} \leq j\alpha/m\}$.

Then it holds:

$$\forall \vartheta \in \Theta : \text{FDR}_\vartheta(\varphi^{\text{LSU}}) = \mathbb{E}_\vartheta \left[\frac{V_m}{R_m \vee 1} \right] = \frac{m_0}{m} \alpha.$$

Google Scholar, 22.04.2014, 11am: Paper cited 21,614 times!

Linear step-up in terms of Simes' rejection line



Benjamini, Y. & Yekutieli, D. (2001) / Sarkar, S. K. (2002):

Proofs for FDR control in presence of special dependency structures:

$$\forall \vartheta \in \Theta : \mathrm{FDR}_\vartheta(\varphi) \leq \frac{m_0}{m}\alpha.$$

for stepwise test procedures employing Simes' critical values.

Model assumptions: **MTP_2 oder PRDS**

Examples:

Multivariate normal distributions with non-negative correlations,

certain multivariate (absolute) t-distributions

(Finner, Dickhaus, Roters (2007), The Annals of Statistics **35**, 1432-1455)

Since under positive dependency the FDR of the LSU- procedure is bounded by

$$\frac{m_0}{m}\alpha$$

for any $m > 1$ and given $\alpha \in (0, 1)$, φ^{LSU} does not exhaust the FDR level α in case of $m_0 < m$.

It may be asked:

Is it possible to derive a better rejection curve

circumventing the factor m_0/m appearing in

the FDR of φ^{LSU} ?

Assumptions:

Independent p-values $p_1, \ldots, p_m$;

$m_0 = m_0(m)$ null hypotheses true with

$$\lim_{m \to \infty} \frac{m_0(m)}{m} = \zeta \in (0, 1].$$

m_0 of the p-values	UNI$[0, 1]$-distributed	(corresp. hypotheses true)
$m_1 = m - m_0$ of the p-values	δ_0-distributed	(corresp. hypotheses false)

Then the ecdf of the p-values F_m (say) converges (Glivenko-Cantelli) for $m \to \infty$ to

$$G_\zeta(x) = (1 - \zeta) + \zeta x \text{ for all } x \in [0, 1].$$

Assume we reject all H_i with $p_i \leq x$ for some $x \in (0,1)$.

Then the FDR (depending on ζ and x) under DU(ζ) is asymptotically given by

$$\mathrm{FDR}_\zeta(x) = \frac{\zeta x}{(1-\zeta)+\zeta x}.$$

Aim: Find an optimal threshold x_ζ (say), such that

$$\mathrm{FDR}_\zeta(x_\zeta) \equiv \alpha \text{ for all } \zeta \in (\alpha, 1).$$

We obtain:

$$\mathrm{FDR}_\zeta(x_\zeta) = \alpha \quad \Longleftrightarrow \quad x_\zeta = \frac{\alpha(1-\zeta)}{\zeta(1-\alpha)}.$$

Ansatz: Rejection curve f_α and G_ζ shall cross each other in x_ζ, i.e.,

$$f_\alpha(x_\zeta) = G_\zeta(x_\zeta).$$

Plugging in x_ζ derived above yields

$$f_\alpha\left(\frac{\alpha(1-\zeta)}{\zeta(1-\alpha)}\right) \stackrel{!}{=} \frac{1-\zeta}{1-\alpha}.$$

Substituting $t = \dfrac{\alpha(1-\zeta)}{\zeta(1-\alpha)} \iff \zeta = \dfrac{\alpha}{(1-\alpha)t+\alpha}$,

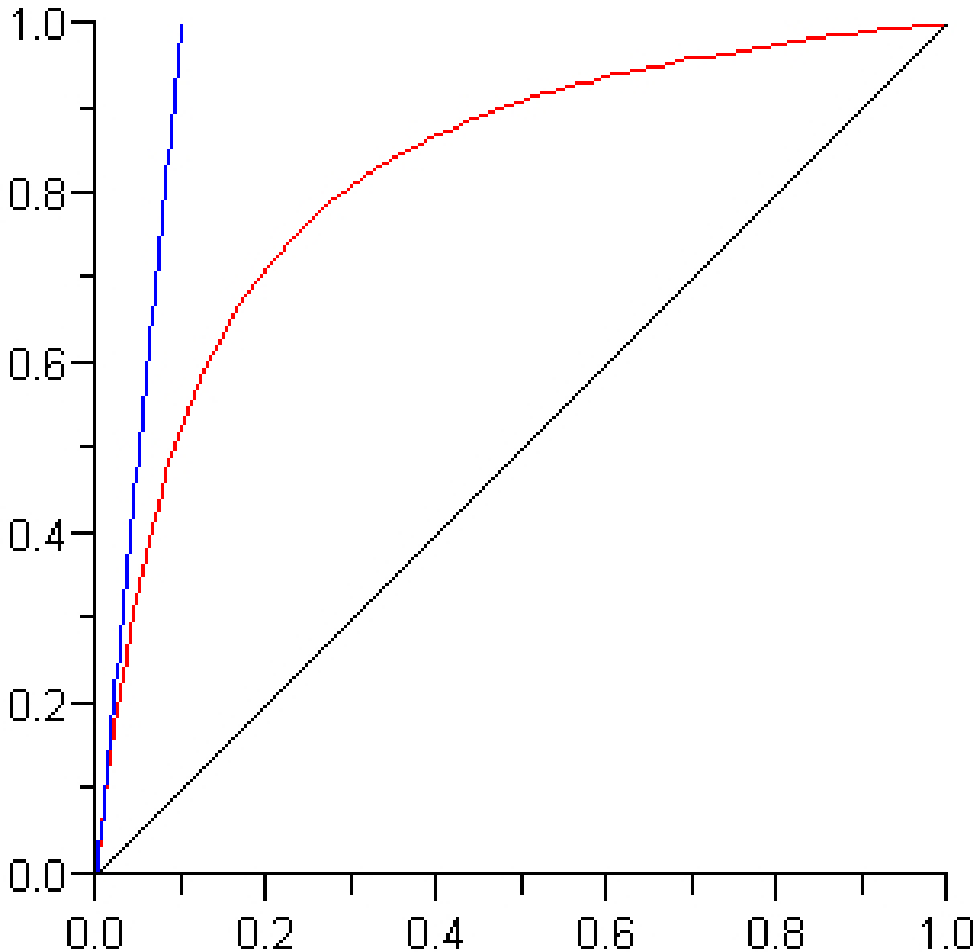
we get that

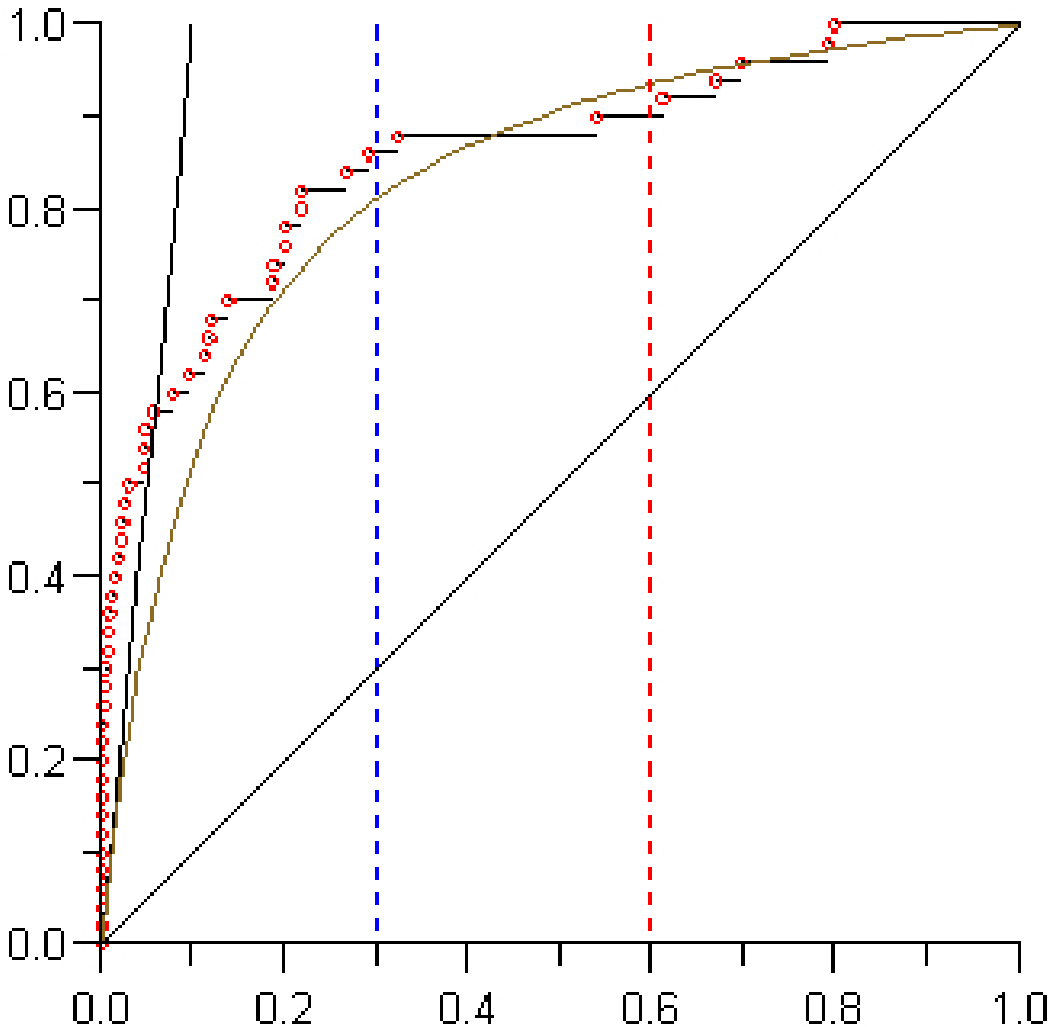
$$f_\alpha(t) = \frac{t}{(1-\alpha)t+\alpha},\ t \in [0,1],$$

is the curve solving the problem!

AORC: Finner, Dickhaus, Roters (2009), Annals of Statistics 37, 596-618

Asymptotically optimal rejection curve for $\alpha = 0.1$




1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

Consider the ANOVA1 model from above.

Define $\theta_{ij} = \mu_i - \mu_j, 1 \leq i < j \leq k$ (pairwise mean differences).

A simultaneous confidence region $C_{ij}(X)$ is a subset of $\mathbb{R}$ with the property ($\alpha \in (0,1)$ fixed) that

$$\forall \mu \in \mathbb{R}^k : \forall \sigma^2 > 0 : \mathbb{P}_{\mu,\sigma^2}(C_{ij}(X) \ni \theta_{ij}\ \forall 1 \leq i < j \leq k) \geq 1 - \alpha. \qquad (1)$$

Duality of multiple tests and simultaneous confidence regions:

Tukey's test $\Longleftrightarrow$ Tukey confidence intervals,

Scheffé's simultaneous coverage thm. $\Longleftrightarrow$ Scheffé test, etc.

Remark. Equation (1) is an analog of the FWER.

Let $\bar{Y}_i = \bar{X}_i - \mu_i \Rightarrow \bar{X}_i - \bar{X}_j - \theta_{ij} = \bar{Y}_i - \bar{Y}_j, i = 1, \ldots, k$. Setting

$$T_{ij}(X) = \sqrt{n}\frac{|\bar{Y}_i - \bar{Y}_j|}{S}, 1 \leq i < j \leq k,$$

the statistic $\max_{1 \leq i < j \leq k} T_{ij}(X)$ has a known distribution, namely, the

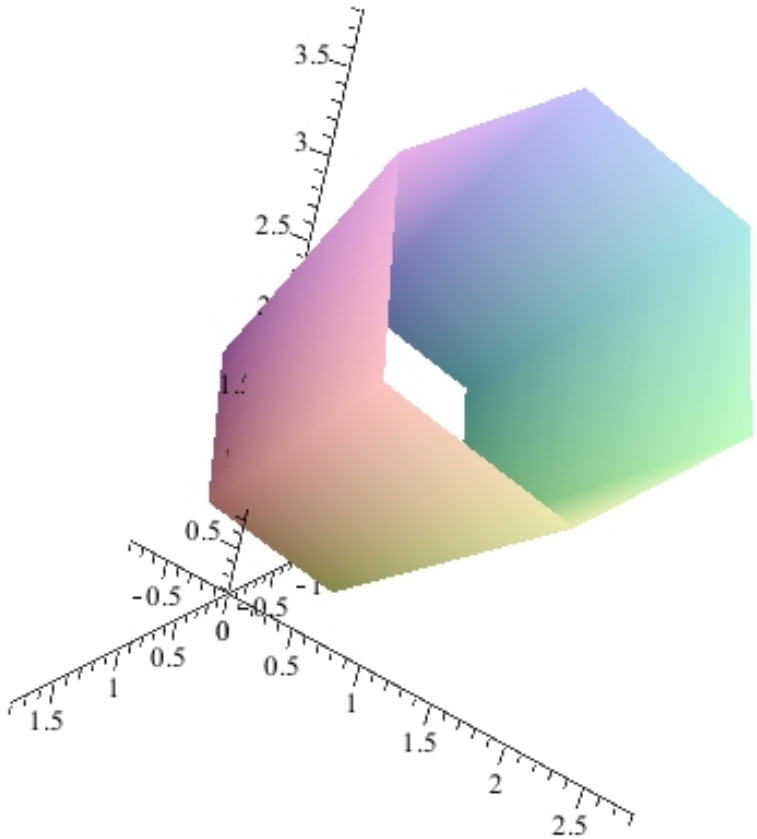
Studentized range distribution $q_{k,k(n-1)}$ with k groups and $k(n-1)$ degrees of freedom.

The simultaneous multiple test based on the $T_{ij}(X)$ and the quantiles of $q_{k,k(n-1)}$ is due to Tukey (1953).

Noticing that

$$\sqrt{n}\frac{|\bar{y}_i - \bar{y}_j|}{s} \leq q_{k,k(n-1);\alpha} \Longleftrightarrow \theta_{ij} \in \left[(\bar{x}_i - \bar{x}_j) \pm \frac{s}{\sqrt{n}} q_{k,k(n-1);\alpha}\right],$$

we obtain simultaneous confidence intervals for the pairwise mean differences and, consequently, a simultaneous confidence region for the mean vector in $\mathbb{R}^k$!



(Maple code taken from Jason C. Hsu, 1996)

- Abramovich et al. (2006), Bogdan et al. (2011), Neuvial and Roquain (2012):

  FDR-thresholding (with φ^{LSU}) asymptotically achieves Bayes-optimal classification risk under sparsity

- Storey (2003), Dickhaus et al. (2013):

  Weighted average of pFDR and pFNR is a proxy for cost-weighted Bayes classification risk in non-sparse models

- Donoho and Jin (2008):

  Higher criticism thresholding leads to optimal feature selection for classification in rare/weak models

- Bauer et al. (1988), Bunea et al. (2006), Furmańczyk (2013):

  Model selection consistency of multiple test procedures

- Zuber and Strimmer (2011):

  Optimization of information criteria AIC, BIC, RIC, ... is equivalent to single-step multiple testing

- Abramovich et al. (2006), Benjamini and Gavrilov (2009):

  Model selection performance of stepwise rejective multiple tests

- Berk et al. (2013):

  Simultaneous inference implies valid post-selection inference

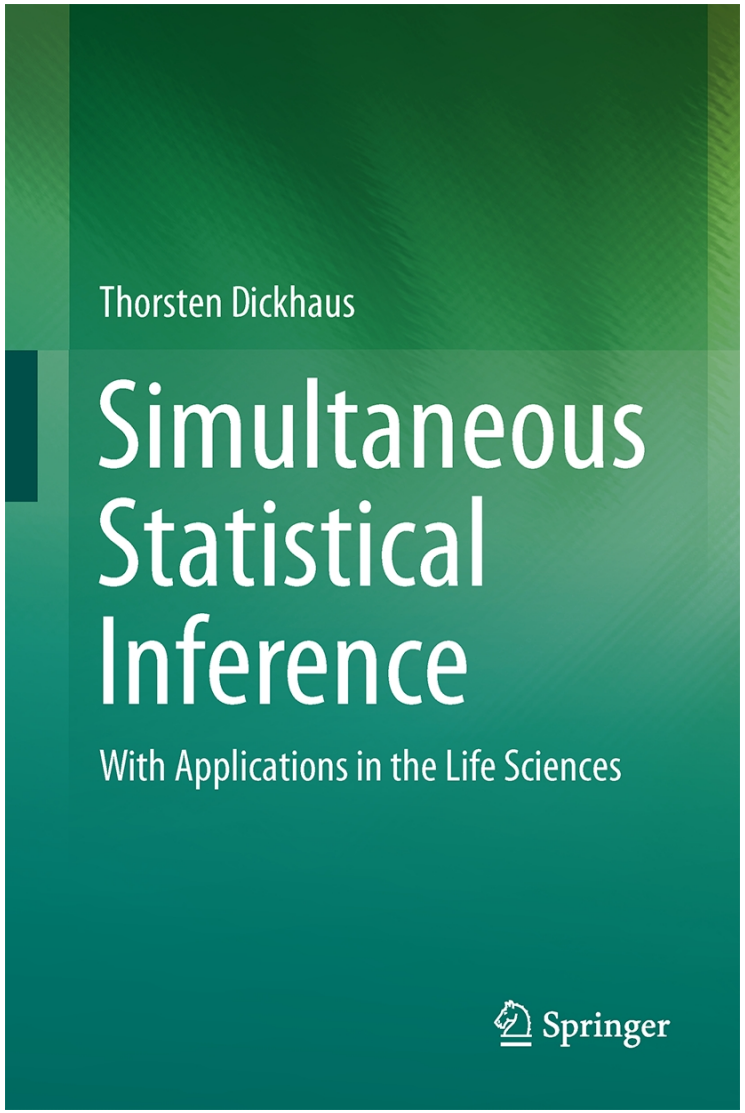
- Qiu and Hwang (2007), Benjamini and Yekutieli (2005), ...

  Valid inference for selected parameters based on FWER / FDR control

- **Jens Stange**, Nina Loginova:
  Multiple testing for behavioural genetics (DFG grant DI 1723/3-1)
- **Konstantin Schildknecht**, Tina Baumann:
  Multiple testing for epigenetic methylation data
  (BMBF project EPILYZE with Epiontis GmbH)
- Simultaneous statistical inference in multivariate time series models
  (Project A3 in IRTG 1792 “High Dimensional Non Stationary Time Series“)

Still at Humboldt-University Berlin:

- **Taras Bodnar**, Yuriy Kopanskyy:
  Multiple testing under unspecified dependency structure
  (DFG Research Unit FOR 1735)


Thorsten Dickhaus
Simultaneous Statistical Inference
With Applications in the Life Sciences
Springer