Multiple test procedures using ordered p-values

-- A survey and personal experiences --

G. Hommel
Vienna, 13.05.2013

IMBEI Institut für Medizinische Biometrie, Epidemiologie und Informatik
UNIVERSITÄTSmedizin. MAINZ

Outline

1. Introduction
2. Global tests
3. Multiple tests
   3.1 Control of FWER
   3.2 Control of FDR
   3.3 Other concepts
4. Weighted tests
5. Closing remarks

Introductory example I

- Group 1: diseased
- Group 2: control
- 12 genetic factors (yes/no)
- → 12 (2x2) tables
- Result:
  3 χ^2 values > 3.84
  9 χ^2 values < 3.84

„Is there anything hidden in the data ???“

Introductory example II

- What is the probability that 3 or more p-values are ≤ 0.05 ?
- When independent:
  probability = 0.0196 (binomial distribution)
  Here not realistic!
- Otherwise:
  probability $\leq {}^{12}/_{3} \cdot 0.05 = 20\%$
  (Markov inequality)

" = " is possible

The Rüger test

- n null hypotheses $H_1, \ldots, H_n$
- global hypothesis $H_0 = \cap \{H_i : i = 1,\ldots,n\}$
- n p-values $p_1, \ldots, p_n$
- ordered p-values $p_{(1)} \le p_{(2)} \le \ldots \le p_{(n)}$
- k = number of significances to be achieved

  $P\left(P_{(k)} \le \frac{k \cdot \alpha}{n}\right) \le \alpha$ under H_0 (Rüger, 1978)

  " = " is possible
- Special case: k = 1 → Bonferroni method

Extensions of the Rüger test

$$P\left(P_{(1)} \le \alpha/n \text{ or } P_{(2)} \le \frac{2\alpha}{n} \text{ or} \ldots \text{or } P_{(n)} \le \frac{n \cdot \alpha}{n}\right) \le C_n \cdot \alpha$$

where $C_n = 1 + \frac{1}{2} + \ldots + \frac{1}{n}$ (Hommel, 1978, 1983)

Extensions of the Rüger test

$$P\left(P_{(1)}\leq \alpha/n \text{ or } P_{(2)}\leq \frac{2\alpha}{n} \text{ or} \ldots \text{or } P_{(n)}\leq \frac{n\cdot\alpha}{n}\right)\leq C_n\cdot\alpha$$

where $C_n = 1+\frac{1}{2}+\ldots+\frac{1}{n}$ (Hommel, 1978, 1983)

More general:

$$0\leq c_1\leq c_2 \leq\ldots\leq c_n \text{ with } c_1+\sum_{i=2}^{n}(c_i-c_{i-1})\cdot\frac{n}{i}=\alpha$$

$\Rightarrow P(P_{(1)}\leq c_1 \text{ or } P_{(2)}\leq c_2 \text{ or } \ldots \text{ or } P_{(n)}\leq c_n) \leq \alpha$

" = α " is possible

(Röhmel/Streitberg, 1983, 1987)

Special cases of the Röhmel-Streitberg test

$$0\leq c_1\leq c_2 \leq\ldots\leq c_n \text{ with } c_1+\sum_{i=2}^{n}(c_i-c_{i-1})\cdot\frac{n}{i}=\alpha$$

- $c_1=\ldots=c_n=\alpha/n$ → Bonferroni
- $c_1=\ldots=c_{k-1}=0,\quad c_k=\ldots=c_n=\frac{k\cdot\alpha}{n}$ → Rüger
- $c_1=\frac{\alpha}{n\cdot C_n},\quad c_2=\frac{2\alpha}{n\cdot C_n},\ldots,c_n=\frac{n\cdot\alpha}{n\cdot C_n}$ → Hommel

Practical experiences and simulations:
Bonferroni has highest power! (nearly always)

The Simes test

R.J.Simes (IBC Seattle 1986, Biometrika 1986)

$$P\left(P_{(1)} \le \alpha/n \text{ or } P_{(2)} \le \frac{2\alpha}{n} \text{ or} \ldots \text{or } P_{(n)} \le \frac{n\alpha}{n}\right) \le \alpha$$

The Simes test

R.J.Simes (IBC Seattle 1986, Biometrika 1986)

- p-values independent ⇒

$$P\left(P_{(1)} \le \alpha/n \text{ or } P_{(2)} \le \frac{2\alpha}{n} \text{ or} \ldots \text{or } P_{(n)} \le \frac{n\alpha}{n}\right) \le \alpha$$

- Simulation with multivariate normal distribution or multivariate χ^2 distribution: level α also controlled → numerous successor publications
- theoretical investigations:
  Samuel-Cahn (1996), Sarkar/Chang (1997),
  Sarkar (1998): MTP_2 condition

Multiple tests

Application of closure test principle:

Find for each $I \subseteq \{1, \ldots, n\}$ a „local“ test of $H_I = \cap\{H_i : i \in I\}$ (all H_i with $i \in I$ are true).

Consequence: Control of FWER (of multiple level) α , i.e. the probability of committing a type I error is at most α .

Possible local tests: Bonferroni, Rüger, Hommel, Röhmel-Streitberg

Closure tests with ordered p-values

- Bonferroni global tests: → procedure by Holm (1979)

$$\text{"SD}\left(\alpha/n,\ \alpha/(n-1),\ \alpha/(n-2),\ \ldots,\ \alpha/2,\ \alpha\right)\text{"}$$

- Rüger tests: some problems (see Hommel, 1986)
- Hommel tests (short-cut): Compute

$$j = \max\left\{ i \in \{1, \ldots, n\} : p_{(n-i+k)} > \frac{k\alpha}{i \cdot C_i} \text{ for } k = 1, \ldots, i \right\}$$

  If this set is empty, reject all H_i; otherwise reject all H_i with

$$p_i \leq \frac{\alpha}{j \cdot C_j}$$

  (Hommel, 1986).

- Röhmel-Streitberg tests: Generalisation possible under an additional condition (Bernhard et al., 2004)

Multiple Simes tests

Determine

$$j = \max\left\{ i \in \{1,\ldots,n\} : p_{(n-i+k)} > \frac{k\alpha}{i} \text{ für } k = 1,\ldots,i \right\}$$

If this set is empty, reject all H_i; otherwise all H_i with

$$p_i \leq \alpha / j$$ (Hommel, 1988).

Conservative version:

Step-up $\left(\alpha/n, \ \alpha/n-1, \ldots, \ \alpha/2, \ \alpha \right)$ (Hochberg, 1988)

Observe: Additional conditions on dependence structure needed for all these tests!

Analysis of genetic data

- Genetic-epidemiological studies (many markers)
  - n = 20-100000
- Microarray studies (many genes or oligonucleotides)
  - n = 200-25000

Control of multiple level too restrictive !

Control of FDR I

- Less restrictive control with False Discovery Rate (Benjamini & Hochberg, 1995)

	H_i retained	H_i rejected	
True H_i	U	V	n_0
Wrong H_i	T	S	$n-n_0$
	n-R	R	n

- FDR=E(Q) , with $Q = \begin{cases} V/R & R > 0 \\ 0 & R = 0 \end{cases}$

Control of FDR II

- Application of “explorative” Simes test
- Determine $k = \max \left\{ 1 \leq i \leq n \,\middle|\, p_{(i)} \leq i/n\, \alpha \right\}$.
- Reject all hypotheses belonging to $p_{(1)}$, ... , $p_{(k)}$
- Multiple level is not controlled!

  (Type I error probability can even tend to 1)
- Controls FDR at level α (even at level $n_0/n \cdot \alpha$) for independent p-values

Control of FDR III

Explorative Simes procedure for dependent p-values: see Benjamini/Yekutieli (2001) „positive regression dependence“

Other possibility:

Determine $k = \max \left\{ 1 \le i \le n \,\middle|\, p_{(i)} \le {}^{i}\!\big/\!(n \cdot C_i)\,\alpha \right\}$.

Reject all hypotheses belonging to $p_{(1)}$, ... , $p_{(k)}$.

→ always control of FDR!

Further developments

- FDR $\le {}^{n_0}\!/\!_{n} \cdot \alpha$:
  Estimate n_0 = number of true null hypotheses
- Other concepts of error rates?
  - Control of V/R directly (False exceedance rate)
  - Control of V („k-FWER“)
    see Victor (1982), Hommel/Hoffmann (1988), Korn et al. (2004), van der Laan et al. (2004), Lehmann/Romano (2005)
  - … and others …

Weighted tests

weights $w_1, \dots, w_n \geq 0$ with $\Sigma w_i = 1$

weighted p-values $q_i = p_i / w_i$, $i = 1, \dots, n$

Two types of ordering:

- Type I : $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(n)}$
- Type II : $q_{(1)} \leq q_{(2)} \leq \dots \leq q_{(n)}$

bounds $b_1, \dots, b_n$

Global test type I: Reject H_0 if $p_{(i)} \leq b_i$ for at least one i .

Global test type II: Reject H_0 if $q_{(i)} \leq b_i$ for at least one i .

Step-down tests: Generalization of Holm (1979).

Step-up tests: Generalization of Hochberg (1988).

Weighted Simes tests

Type I: Benjamini/Hochberg (1997)

Similar properties as unweighted Simes test

n=2: level α test under prd condition

(Brannath et al., 2009);

→ consonant

Type II: Hochberg/Liberman (1994)

different rejection region as for type I

less power for high correlations

n=2: → not consonant

Step-down procedures with weights

Type II: Classical weighted Holm procedure
(Holm, 1979)
no conceptual drawbacks

Type I: Benjamini/Hochberg, 1997
2 problems:
- tied p-values
- p-inconsistency, i.e. rejection pattern not monotone in p-values

Step-up procedures with weights

Type I: Tamhane/Liu (2008)
3 problems:
- tied p-values
- p-inconsistency
- α-control for independent p-values: conjectured, but no complete proof

Type II: Does not work at all (with bounds of weighted Holm procedure) (Tamhane/Liu)
Other bounds??

FDR-controlling procedures with weights

- Consider explorative Simes procedure (Benjamini/ Hochberg, 1995) = step-up version of weighted Simes tests → no conceptual problems!
- Control of FDR ≤ α ? (under independence)
  Type II: yes (Genovese et al., Biometrika 2006)
  Type I: control of *W*FDR proven (B/H, 1997)
  control of usual FDR conjectured
- For equal weights: FDR ≤ $n_0/n \cdot \alpha$
  (n_0 = number of true null hypotheses).
  Does not remain true for unequal weights.

Further remarks

Only considered: cut-off tests
(„Schrankentests", Röhmel/Streitberg, 1987)

Other functions of ordered p-values?

Maurer/Mellein (1988): linear minmax tests
e.g. „Reject H_0 if $(1-\alpha)p_{(1)} + \alpha p_{(n)} \leq \alpha$ "
closure test easy
but only valid for independent p-values

Published in

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

Multiple hypotheses testing based on ordered p values--a historical survey with applications to medical research.

J Biopharm Stat. 2011 Jul;21(4):595-609. doi: 10.1080/10543406.2011.552879.