Nonparametric Multiple Comparison Procedures under Heteroscedasticity

Frank Konietschke

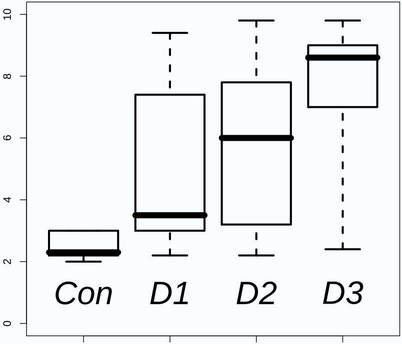
Department of Medical Statistics
University Medical Center Göttingen

Overview

- Example
- Parametric motivation
- Nonparametric model
- Single step and stepwise procedures
- Example evaluation
- References

Example: Reaction Times

- Reaction times [sec] of mice
  - 1 control group and three dose groups
  - 10 animals per group



- Impact of the dose?

Parametric Model

- $X_{ik} \sim N(\mu_i, \sigma_i^2)$; $i = 1, \ldots, a$; $k = 1, \ldots, n_i$; $N = \sum_{i=1}^{a} n_i$
- Reaction times
  - $i =$ control, dose 1, ..., dose 3
  - $k = 1, \ldots, 10$
- $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_a)'$ (expectations)
- Hypothesis:

$$H_0^{\mu} : \mu_1 = \ldots = \mu_a$$

- Global test procedure (ANOVA-based)

$$A = \frac{\text{,,Variance between''}}{\text{,,Variance within''}} \overset{H_0^{\mu}}{\sim} F(f_1, f_2)$$

ANOVA Based Evaluation

- Three steps
  1. $H_0^{\mu} : \mu_1 = \cdots = \mu_a$ (overall hypothesis)
  2. Multiple comparisons after rejecting the global hypothesis
     - $H_0^{(1,2)} : \mu_1 = \mu_2$ / effects: $\mu_{12} = \mu_1 - \mu_2$
     - $H_0^{(1,3)} : \mu_1 = \mu_3$ / effects: $\mu_{13} = \mu_1 - \mu_3$
     - ...
  3. Confidence intervals for the effects
     - Regulatory authorities require confidence intervals (ICH E9)
- Problems
  - Overall result and multiple comparisons may be incompatible

Multiple Contrasts

- Better: Start with multiple comparisons
- Hypotheses $\Rightarrow$ contrast matrix

$$\mathbf{C} = \begin{pmatrix} \mathbf{c}_1' \\ \vdots \\ \mathbf{c}_q' \end{pmatrix} = \begin{pmatrix} c_{11} & \cdots & c_{1a} \\ \vdots & \ldots & \vdots \\ c_{1q} & \cdots & c_{qa} \end{pmatrix}, \ (\mathbf{C1} = \mathbf{0})$$

  - $H_0^\ell : \mathbf{c}_\ell' \boldsymbol{\mu} = 0$ simultaneously
  - $H_0 : \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$

Multiple Comparisons

- Test statistics for $H_0^\ell : \mathbf{c}_\ell'\boldsymbol{\mu} = 0$
  - $\overline{\mathbf{X}}_\cdot = (\overline{X}_{1\cdot}, \dots, \overline{X}_{a\cdot})'$ (vector of means)
  - s_ℓ^2: variance estimator of $Var(\mathbf{c}_\ell'\overline{\mathbf{X}}_\cdot)$
  - T-test type statistic: $T_\ell = \mathbf{c}_\ell'\overline{\mathbf{X}}_\cdot / s_\ell$, $\ell = 1 \dots, q$
- Family of hypotheses and test statistics
  - $\boldsymbol{\Omega} = \{H_0^\ell : \mathbf{c}_\ell'\boldsymbol{\mu} = 0, T_\ell, \ell = 1, \dots, q\}$
- Goal: control the FWER α in the strong sense

$$P\left(\text{reject at least one true } H_0^\ell\right) \leq \alpha$$

- Alternatives H_1^ℓ do not influence the non-rejection of true hypotheses

Multiple Comparisons II

- Single step procedures

$$P\left(\bigcap_{j\in J}\{|T_j| \geq c\}\right) \leq \alpha$$

- Stepwise procedures

$$P\left(\bigcap_{j\in J}\{|T_j| \geq c_j\}\right) \leq \alpha$$

- c and c_j: adequate critical values for all test statistics
- J: set of indexes for true hypotheses
- Difficulty
  - Correlation among the test statistics
  - Needs to be investigated

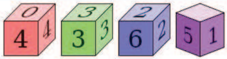
Nonparametric Effects

- Statistical model
  - $X_{ik} \sim F_i;\ i = 1, \ldots, a;\ k = 1, \ldots, n_i;\ N = \sum_i n_i$
  - No parametrization
- Relative effects

$$p_{ij} = \int F_i dF_j = P(X_{i1} < X_{j1}) + 0.5P(X_{i1} = X_{j1})$$

- Pairwise defined relative effects useful?

Nonparametric Effects - Efron's Dice



Wikipedia.de

- Four die

$$D_1 = \{0,0,4,4,4,4\} \qquad D_2 = \{3,3,3,3,3,3\}$$
$$D_3 = \{2,2,2,2,6,6\} \qquad D_4 = \{1,1,1,5,5,5\}$$

  - Effects

$$P\{D_1 \leq D_2\} = \frac{1}{3} \qquad P\{D_2 \leq D_3\} = \frac{1}{3}$$
$$P\{D_3 \leq D_4\} = \frac{1}{3} \qquad P\{D_4 \leq D_1\} = \frac{1}{3}$$

    - Paradox results possible
    - How to overcome the paradox?

Motivation: Nonparametric Procedures

- Statistical model
  - $X_{ik} \sim F_i;\ i = 1, \ldots, a;\ k = 1, \ldots, n_i;\ N = \sum_i n_i$
- Relative effects
  - $H = \frac{1}{N} \sum_{i=1}^{a} n_i F_i$

$$p_i = \int H dF_i = P(Z < X_{i1}) + 0.5P(Z = X_{i1})$$

  - $Z \sim H$ („Mean")
- Interpretation
  - $p_i > 0.5$: X_{i1} tends to result in larger values than Z
  - $p_i = 0.5$: no tendency to smaller or larger values
  - **X_{i1} tends to result in smaller values than X_{j1}, if $p_i < p_j$**
  - **No smaller or larger values, if $p_i = p_j$**

Motivation: Nonparametric Procedures

- Bank-die

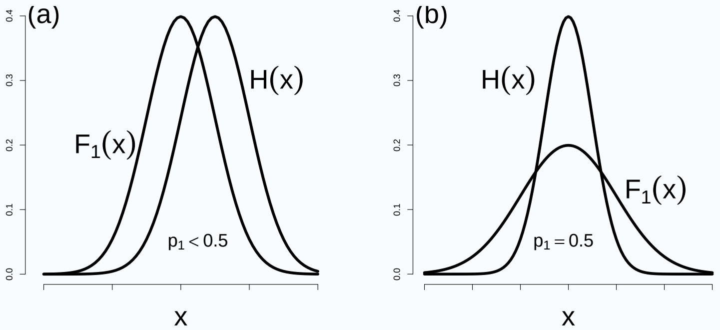
$$B = \{0, 0, 4, 4, 4, 3, 3, 3, 3, 3, 3, 2, 2, 2, 2, 6, 6, 1, 1, 1, 5, 5, 5\}$$

  - Effects

$$P\{B \le D_1\} = \frac{35}{72} \qquad P\{B \le D_2\} = \frac{36}{72}$$
$$P\{B \le D_3\} = \frac{37}{72} \qquad P\{B \le D_4\} = \frac{36}{72}$$

- Meaning
  - Die 1 worse than die 2 and 4, and die 3 is the best

Nonparametric Effects - Demonstration



- Two normal distributions
- $p_i = p_j$ even under heteroscedasticity

Nonparametric Hypotheses

- Goal
  - Multiple comparison procedures for

$$H_0^\ell : \mathbf{c}_\ell' \mathbf{p} = 0$$

  - Simultaneous confidence intervals for $\delta_\ell = \mathbf{c}_\ell' \mathbf{p}$
  - Global hypothesis

$$H_0 : \mathbf{C}\mathbf{p} = \mathbf{0}$$

- $\mathbf{p} = (p_1, \ldots, p_a)'$

Estimators

- Rank estimators
  - Overall ranks (N observations)
    - Assign ranks R_{ij} of X_{ij} among all N observations
  - $\overline{R}_{i\cdot}$: mean of the ranks in sample i
  - $\widehat{p}_i = \frac{1}{N}(\overline{R}_{i\cdot} - 0.5)$
  - $\widehat{\mathbf{p}} = (\widehat{p}_1, \ldots, \widehat{p}_a)'$
- Estimator is strongly consistent and unbiased

Estimators - Distribution

- Shown: $\sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p}) \sim N(\mathbf{0}, \mathbf{V}_N)$
- $\mathbf{V}_N$: rather complicated structure
- For arbitrary contrasts $\sqrt{N}\mathbf{C}(\widehat{\mathbf{p}} - \mathbf{p}) \sim N(\mathbf{0}, \mathbf{C}\mathbf{V}_N\mathbf{C}')$
- Consistent estimator
  - $\widehat{\mathbf{V}}_N$: Different rankings (too technical)

Estimators - Distribution II

- The kind of contrast generates certain kinds of correlations
- For comparisons against a control
  - Distribution of $\sqrt{N}\mathbf{C}(\widehat{\mathbf{p}} - \mathbf{p})$ is multivariate of totally positive order 2 (MTP2)
  - $f(\mathbf{x} \vee \mathbf{y})f(\mathbf{x} \wedge \mathbf{y}) \geq f(\mathbf{x})f(\mathbf{y})$
  - $\mathbf{x} \vee \mathbf{y} = (\max(x_1, y_1), \ldots, \max(x_{a-1}, y_{a-1}))$
  - $\mathbf{x} \wedge \mathbf{y} = (\min(x_1, y_1), \ldots, \min(x_{a-1}, y_{a-1}))$
- Supports the use of step-up procedures

Test Statistics

- Test statistic for $H_0^\ell : \mathbf{c}_\ell' \mathbf{p} = 0$

$$T_\ell = \sqrt{N}(\mathbf{c}_\ell'(\widehat{\mathbf{p}} - \mathbf{p}))/\sqrt{\widehat{v}_\ell} \to N(0,1)$$

- $\widehat{v}_\ell$: consistent variance estimator
- Single step procedures
  - Bonferroni
    - Reject H_0^ℓ if p-Value$_\ell \leq \alpha/q$
  - very conservative, no correlation account

Multiple Contrast Tests

- $\mathbf{T} = (T_1, \ldots, T_q)' \sim N(\mathbf{0}, \mathbf{R})$
- $\mathbf{R}$: correlation matrix (estimator: $\widehat{\mathbf{R}}$)
- Reject $H_0^\ell : \mathbf{c}_\ell'\mathbf{p} = 0$ if

$$|T_\ell| \geq z_{1-\alpha}(\widehat{\mathbf{R}})$$

- Compatible $(1-\alpha)$-simultaneous confidence interval

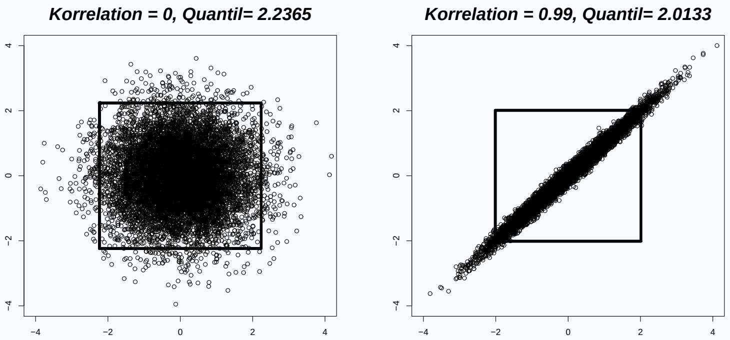
$$CI_\ell = \mathbf{c}_\ell'\widehat{\mathbf{p}} \pm z_{1-\alpha}(\widehat{\mathbf{R}})\widehat{v}_\ell$$

- Reject $H_0 : \mathbf{Cp} = \mathbf{0}$

$$\max\{|T_1|, \ldots, |T_q|\} \geq z_{1-\alpha}(\widehat{\mathbf{R}})$$

- $z_{1-\alpha}(\widehat{\mathbf{R}})$: $(1-\alpha)$-equicoordinate quantile from $N(\mathbf{0}, \widehat{\mathbf{R}})$

Equikoordinate Quantile



- Equicoordinate quantiles from bivariate normal distributions
- Cuboid with quadratic area
- Computation with R software package „mvtnorm"

Stepwise Procedures

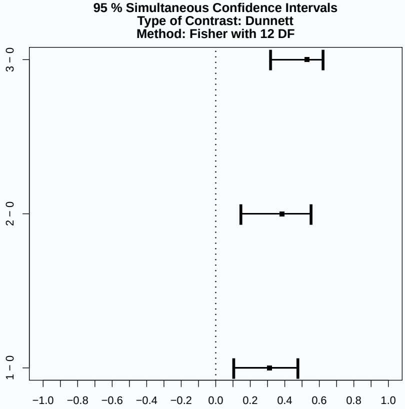
- ordered p-values $pV_{(1)}, \dots, pV_{(q)}$
- Holm - method (step-down)
  - k: minimal index s.t. $pV_{(k)} > \alpha/(q+1-k)$
  - Reject $H_0^{(1)}, \dots, H_0^{(k-1)}$ and do not reject $H_0^{(k)}, \dots, H_0^{(q)}$
  - Known to be more powerful than Bonferroni
- Step-up (Hochberg), only for comparisons against a control
  - Reject all $H_0^{(\ell')}$ $(\ell' \leq \ell)$, if $pV_{(\ell)} \leq \alpha/(q-\ell+1)$, $\ell = q, q-1, \dots, 1$
  - Valid since MTP2 condition is fulfilled
  - Known to be very powerful

Simulations

- Compare control of FWER and power of
  - single -step vs. stepwise procedures
- Results
  - All procedures have similiar power than parametric procedures under normality
  - Significant higher power than paramatric procedures under non-normality
  - Procedures work nicely with $n_i \geq 8$
  - MCTP has slightly higher power than Hochberg adjustment

Example: Evaluation

- R package *nparcomp* (online on CRAN)
- Multiple comparisons against a control



Comp	Est.	Lower	Upper	T	p.Value
1 - 0	0.31	0.10	0.48	3.87	0.006
2 - 0	0.39	0.15	0.55	4.00	0.005
3 - 0	0.53	0.32	0.62	5.84	0.0001

References I

- Bretz F., Genz A., Hothorn LA (2001). On the numerical availability of multiple comparison procedures. *Biometrical Journal* **43**, 645 – 656.
- Munzel, U., Hothorn, L.A. (2001). A Unified Approach to Simultaneous Rank Tests Procedures in the Unbalanced One-Way Layout. *Biometrical Journal* **43**, 553 - 569
- Konietschke, F. (2009) Simultane Konfidenzintervalle für nichtparametrische relative Kontrasteffekte. Ph.D. thesis, Georg-August Universität Göttingen.

References II

- Konietschke, F., Bathke, A. C., Hothorn, L. A., Brunner, E. (2010). Testing and Estimation of purely nonparametric effects in repeated measures designs (2010). *Computational Statistics and Data Analysis* **54**, 1895 – 1905.
- Konietschke, F., Hothorn, L. A., Brunner, E. (2011). Rank-based multiple test procedures and simultaneous confidence intervals. *The Electronic Journal of Statistics* **6**, 737 – 758.