A Multivariate Dunnett Procedure for Correlated Endpoints

Mario Hasler

Institute of Biostatistics, Leibniz University Hannover
hasler@biostat.uni-hannover.de

Lübeck, July 2008

Motivation

Data from Schulte et al. [7]: measurements for three liver enzymes selected from an interlaboratory immunotoxicity study for females in one center; ten animals randomized to a control and three dose groups

Group	ASAT	ALAT	ALP
0	86.944	57.0752	461.496
1	80.080	49.2674	391.304
2	81.536	51.9610	308.976
3	81.536	46.3008	281.260

Table: Sample means per group and enzyme.

Question: For which enzymes do the doses have significantly smaller values than the control?

Motivation

Objective: A single step procedure for ...

- comparing several groups with a control
- having multiple endpoints without order restriction
- by simultaneous confidence intervals
- for ratios of means (and differences as well)

Why ratios:

- easier interpretation because, e.g., in per cent
- independence from the scale of the endpoints

Motivation

Objective: A single step procedure for . . .

- comparing several groups with a control
- having multiple endpoints without order restriction
- by simultaneous confidence intervals
- for ratios of means (and differences as well)

Why ratios:

- easier interpretation because, e.g., in per cent
- independence from the scale of the endpoints

Test procedure

Assumptions:

- $l = 0, \ldots, q$ groups ($l = 0$ for the control)
- $i = 1, \ldots, k$ endpoints
- $j = 1, \ldots, n_l$ observations and $\sum_{l=0}^{q}(n_l - 1) \geq k$
- observations X_{lij} with

$$\{X_{lij} : i = 1, \ldots, k\} \sim \perp \mathrm{N}_k(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}), \; l = 0, \ldots, q, \, j = 1, \ldots, n_l \quad (1)$$

- mean vectors $\boldsymbol{\mu}_l = (\mu_{l1}, \ldots \mu_{lk})$ with estimators $\bar{\mathbf{X}}_l = (\bar{X}_{l1}, \ldots, \bar{X}_{lk})$
- estimator $\hat{\boldsymbol{\Sigma}} = (\hat{\sigma}_{ih})_{1 \leq i,h \leq k}$ for the covariance matrix $\boldsymbol{\Sigma}$ where

$$\hat{\boldsymbol{\Sigma}} = \frac{\sum_{l=0}^{q}(n_l - 1)\hat{\boldsymbol{\Sigma}}_l}{\sum_{l=0}^{q}(n_l - 1)} \quad (2)$$

- pooled sample variances for the endpoints $\hat{\sigma}_{ii} = S_i^2$ on the diagonal

Test procedure

Testing problem:

- hypotheses for general non-inferiority or superiority testing:

$$H_0: \quad \gamma_{li} \geq \theta_i \quad \forall i = 1, \ldots, k, \quad \forall l = 1, \ldots, q \quad \text{vs.} \tag{3}$$
$$H_A: \quad \gamma_{li} < \theta_i \quad \text{for at least one endpoint and one dose}$$

  with $\gamma_{li} = \mu_{li}/\mu_{0i}$ and endpoint-specific relative thresholds $\theta_i > 0$
- $\Rightarrow$ union-intersection test $\rightarrow$ "max-test" over all doses and endpoints
- for $\theta_i = 1$ for all $i = 1, \ldots, k$, the problem reduces to the difference based one
- test statistics:

$$T_{li} = \frac{\bar{X}_{li} - \theta_i \bar{X}_{0i}}{S_i \sqrt{\frac{1}{n_l} + \frac{\theta_i^2}{n_0}}} \quad (l = 1, \ldots, q, \ i = 1, \ldots, k). \tag{4}$$

Test procedure

Distribution:

- $\mathbf{T}_l = (T_{l1}, \ldots, T_{lk})$ for a fixed comparison (dose l with control) can be formed as

$$\mathbf{T}_l = \left(\frac{Y_{l1}}{\sqrt{U_1/\nu}}, \ldots, \frac{Y_{lk}}{\sqrt{U_k/\nu}} \right) \quad (l = 1, \ldots, q) \tag{5}$$

where $(Y_{l1}, \ldots, Y_{lk}) \sim \mathrm{N}_k(\boldsymbol{\delta}_l, \mathbf{R}_l)$ and $U_1, \ldots, U_k$ are dependent χ^2 variables with degrees of freedom

$$\nu = \sum_{l=0}^{q} (n_l - 1) \tag{6}$$

Test procedure

- approximation: under ∂H_0

$$\mathbf{T}_l \overset{appr.}{\sim} t_k(\nu, \mathbf{R}_l) \quad (7)$$

with ν degrees of freedom and correlation matrix $\mathbf{R}_l$ with elements

$$\rho_{l,ih} = \rho_{ih} \frac{\frac{n_0}{n_l} + \theta_i \theta_h}{\sqrt{\frac{n_0}{n_l} + \theta_i^2}\sqrt{\frac{n_0}{n_l} + \theta_h^2}} \quad (l = 1, \ldots, q, \quad i, h = 1, \ldots, k) \quad (8)$$

and the endpoints' correlations $\mathbf{R} = (\rho_{ih})_{1 \leq i,h \leq k}$

- not the classical definition but see Tong [8] (page 202f)
- for equal thresholds for all endpoints ($\theta_i = \theta \; \forall \, 1 \leq i \leq k$), we get the identity $\rho_{l,ih} = \rho_{ih}$ and $\mathbf{R}_l = \mathbf{R}$

Test procedure

- conclusion: under ∂H_0

$$\mathbf{T} = (\mathbf{T}_1, \ldots, \mathbf{T}_q) \overset{appr.}{\sim} t_{qk}(\nu, \tilde{\mathbf{R}}) \tag{9}$$

with ν degrees of freedom and correlation matrix

$$\tilde{\mathbf{R}} = \begin{pmatrix} \mathbf{R}_1 & \mathbf{R}_{12} & \ldots & \mathbf{R}_{1q} \\ \mathbf{R}_{12} & \mathbf{R}_2 & \ldots & \mathbf{R}_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{1q} & \mathbf{R}_{2q} & \ldots & \mathbf{R}_q \end{pmatrix} \tag{10}$$

and off-diagonal sub-matrices $\mathbf{R}_{lr}$ with elements

$$\rho_{lr,ih} = \rho_{ih} \frac{\theta_i \theta_h}{\sqrt{\frac{n_0}{n_l} + \theta_i^2}\sqrt{\frac{n_0}{n_r} + \theta_h^2}} \quad (l, r = 1, \ldots, q, \quad i, h = 1, \ldots, k) \tag{11}$$

Test procedure

- for $i = h$, we get the correlations of a Dunnett test for ratios of means (see Dilba et al. [1])

$\Rightarrow$ **Dunnett test ($k = 1$ endpoint) simply incorporated**

- $\tilde{\mathbf{R}}$ no product moment structure
- correlations ρ_{ih} unknown $\rightarrow$ estimator $\hat{\tilde{\mathbf{R}}}$

Decision rule: reject H_0 if

$$T_{li} < -t_{qk,1-\alpha}(\nu, \hat{\tilde{\mathbf{R}}}) \quad (12)$$

for at least one endpoint and one dose

Test procedure

- for $i = h$, we get the correlations of a Dunnett test for ratios of means (see Dilba et al. [1])

$\Rightarrow$ **Dunnett test ($k = 1$ endpoint) simply incorporated**

- $\tilde{\mathbf{R}}$ no product moment structure
- correlations ρ_{ih} unknown $\rightarrow$ estimator $\hat{\tilde{\mathbf{R}}}$

Decision rule: reject H_0 if

$$T_{li} < -t_{qk,1-\alpha}(\nu, \hat{\tilde{\mathbf{R}}}) \qquad (12)$$

for at least one endpoint and one dose

Test procedure

- for $i = h$, we get the correlations of a Dunnett test for ratios of means (see Dilba et al. [1])

$\Rightarrow$ **Dunnett test ($k = 1$ endpoint) simply incorporated**

- $\tilde{\mathbf{R}}$ no product moment structure
- correlations ρ_{ih} unknown $\rightarrow$ estimator $\hat{\tilde{\mathbf{R}}}$

Decision rule: reject H_0 if

$$T_{li} < -t_{qk,1-\alpha}(\nu, \hat{\tilde{\mathbf{R}}}) \qquad (12)$$

for at least one endpoint and one dose

Test procedure

- for $i = h$, we get the correlations of a Dunnett test for ratios of means (see Dilba et al. [1])

$\Rightarrow$ **Dunnett test ($k = 1$ endpoint) simply incorporated**

- $\tilde{\mathbf{R}}$ no product moment structure
- correlations ρ_{ih} unknown $\rightarrow$ estimator $\hat{\tilde{\mathbf{R}}}$

Decision rule: reject H_0 if

$$T_{li} < -t_{qk,1-\alpha}(\nu, \hat{\tilde{\mathbf{R}}}) \quad (12)$$

for at least one endpoint and one dose

α-simulations

Settings:

- 4 doses and a control
- 2, 4, 8 endpoints with means and thresholds:
  - $\boldsymbol{\mu_0} = (10, 100)$, $\boldsymbol{\theta} = (0.8, 1.25)$
  - $\boldsymbol{\mu_0} = (0.1, 1, 10, 100)$, $\boldsymbol{\theta} = (0.8, 0.8, 1.25, 1.25)$
  - $\boldsymbol{\mu_0} = (0.05, 0.1, 0.5, 1, 5, 10, 50, 100)$, $\boldsymbol{\theta} = (0.8, \ldots, 0.8, 1.25, \ldots, 1.25)$
- equicorrelations: ρ^{min}, 0, 0.5, 1
- fix sample size 20 for each endpoint of each group
- 10000 simulation runs using the statistical software *R* [6], package *mvtnorm* [4, 5]

α-simulations

All dose groups under H_0 (for weak control)

Endpoints	Method	Correlations			
		ρ^{min}	0	0.5	1
2	new	0.051	0.052	0.052	0.052
	Bonf.	0.043	0.045	0.040	0.028
4	new	0.052	0.053	0.052	0.049
	Bonf.	0.045	0.046	0.038	0.013
8	new	0.049	0.051	0.050	0.050
	Bonf.	0.043	0.044	0.036	0.008

Table: Simulated global FWER of the one-sided test ("greater"); groups $X_0, \ldots, X_4$, $\alpha = 0.05$.

$\Rightarrow$ **size-α test**

α-simulations

All dose groups under H_0 (for weak control)

Endpoints	Method	Correlations			
		ρ^{min}	0	0.5	1
2	new	0.051	0.052	0.052	0.052
	Bonf.	0.043	0.045	0.040	0.028
4	new	0.052	0.053	0.052	0.049
	Bonf.	0.045	0.046	0.038	0.013
8	new	0.049	0.051	0.050	0.050
	Bonf.	0.043	0.044	0.036	0.008

Table: Simulated global FWER of the one-sided test ("greater"); groups $X_0, \ldots, X_4$, $\alpha = 0.05$.

$\Rightarrow$ **size-α test**

α-simulations

First endpoint of each dose group under H_{Ai} (for strong control)

Endpoints	Method	Correlations			
		ρ^{min}	0	0.5	1
2	new	0.024	0.022	0.025	0.039
	Bonf.	0.021	0.018	0.020	0.018
4	new	0.037	0.038	0.040	0.049
	Bonf.	0.032	0.033	0.030	0.013
8	new	0.043	0.045	0.044	0.050
	Bonf.	0.037	0.039	0.032	0.008

Table: Simulated local FWER of the one-sided test ("greater"); groups $X_0, \dots, X_4$, $\alpha = 0.05$.

$\Rightarrow$ **keeps the α-level in strong sense**

α-simulations

First endpoint of each dose group under H_{Ai} (for strong control)

Endpoints	Method	Correlations			
		ρ^{min}	0	0.5	1
2	new	0.024	0.022	0.025	0.039
	Bonf.	0.021	0.018	0.020	0.018
4	new	0.037	0.038	0.040	0.049
	Bonf.	0.032	0.033	0.030	0.013
8	new	0.043	0.045	0.044	0.050
	Bonf.	0.037	0.039	0.032	0.008

Table: Simulated local FWER of the one-sided test ("greater"); groups $X_0, \ldots, X_4$, $\alpha = 0.05$.

⇒ **keeps the α-level in strong sense**

Simultaneous confidence intervals

...for testing but also estimating the γ_{li} ($l = 1, \ldots, q,\ i = 1, \ldots, k$)

Derivation:

- according to Fieller [3] and Dilba et al. [1]
- correlation matrix $\tilde{\mathbf{R}}$ here depends on the unknown ratios γ_{li}
- plug in method of Dilba et al. [1] $\rightarrow$ estimators $\hat{\gamma}_{li} = \frac{\bar{X}_{li}}{\bar{X}_{0i}}$ instead of θ_i in Equations (8) and (11) (equivalently for index h)

Simultaneous confidence intervals

Upper limits of the approximate $(1-\alpha)100\%$ SCI for the γ_{li}:

$$\hat{\theta}_{li}^{upper} = \frac{\hat{\gamma}_{li} + \sqrt{g_i \left(\hat{\gamma}_{li}^2 + (1-g_i)\frac{n_0}{n_l} \right)}}{1-g_i} \quad (l = 1, \ldots, q,\ i = 1, \ldots, k) \quad (13)$$

where

$$g_i = \frac{1}{n_0 \bar{X}_{0i}^2} t_{qk,1-\alpha}^2(\nu, \hat{\hat{\mathbf{R}}}) S_i^2 \overset{!}{<} 1 \quad (14)$$

Evaluation of the example

Question: For which enzymes do the doses have significantly smaller values than the control?

Correlation matrix of the data:

$$\mathbf{R} = \begin{pmatrix} 1.000 & 0.361 & 0.335 \\ 0.361 & 1.000 & 0.135 \\ 0.335 & 0.135 & 1.000 \end{pmatrix}$$

Correlation matrix of the comparisons for all endpoints:

$$\hat{\hat{\mathbf{R}}} = \begin{pmatrix} 1.000 & 0.361 & 0.334 & 0.463 & 0.165 & 0.126 & 0.463 & 0.154 & 0.118 \\ 0.361 & 1.000 & 0.135 & 0.162 & 0.440 & 0.049 & 0.162 & 0.412 & 0.046 \\ 0.334 & 0.135 & 1.000 & 0.148 & 0.059 & 0.360 & 0.148 & 0.055 & 0.337 \\ 0.463 & 0.162 & 0.148 & 1.000 & 0.361 & 0.330 & 0.468 & 0.156 & 0.119 \\ 0.165 & 0.440 & 0.059 & 0.361 & 1.000 & 0.134 & 0.166 & 0.424 & 0.047 \\ 0.126 & 0.049 & 0.360 & 0.330 & 0.134 & 1.000 & 0.127 & 0.047 & 0.290 \\ 0.463 & 0.162 & 0.148 & 0.468 & 0.166 & 0.127 & 1.000 & 0.361 & 0.327 \\ 0.154 & 0.412 & 0.055 & 0.156 & 0.424 & 0.047 & 0.361 & 1.000 & 0.134 \\ 0.118 & 0.046 & 0.337 & 0.119 & 0.047 & 0.290 & 0.327 & 0.134 & 1.000 \end{pmatrix}$$

Evaluation of the example

Question: For which enzymes do the doses have significantly smaller values than the control?

Correlation matrix of the data:

$$\mathbf{R} = \begin{pmatrix} 1.000 & 0.361 & 0.335 \\ 0.361 & 1.000 & 0.135 \\ 0.335 & 0.135 & 1.000 \end{pmatrix}$$

Correlation matrix of the comparisons for all endpoints:

$$\hat{\hat{\mathbf{R}}} = \begin{pmatrix} 1.000 & 0.361 & 0.334 & 0.463 & 0.165 & 0.126 & 0.463 & 0.154 & 0.118 \\ 0.361 & 1.000 & 0.135 & 0.162 & 0.440 & 0.049 & 0.162 & 0.412 & 0.046 \\ 0.334 & 0.135 & 1.000 & 0.148 & 0.059 & 0.360 & 0.148 & 0.055 & 0.337 \\ 0.463 & 0.162 & 0.148 & 1.000 & 0.361 & 0.330 & 0.468 & 0.156 & 0.119 \\ 0.165 & 0.440 & 0.059 & 0.361 & 1.000 & 0.134 & 0.166 & 0.424 & 0.047 \\ 0.126 & 0.049 & 0.360 & 0.330 & 0.134 & 1.000 & 0.127 & 0.047 & 0.290 \\ 0.463 & 0.162 & 0.148 & 0.468 & 0.166 & 0.127 & 1.000 & 0.361 & 0.327 \\ 0.154 & 0.412 & 0.055 & 0.156 & 0.424 & 0.047 & 0.361 & 1.000 & 0.134 \\ 0.118 & 0.046 & 0.337 & 0.119 & 0.047 & 0.290 & 0.327 & 0.134 & 1.000 \end{pmatrix}$$

Evaluation of the example

Question: For which enzymes do the doses have significantly smaller values than the control?

Correlation matrix of the data:

$$\mathbf{R} = \begin{pmatrix} 1.000 & 0.361 & 0.335 \\ 0.361 & 1.000 & 0.135 \\ 0.335 & 0.135 & 1.000 \end{pmatrix}$$

Correlation matrix of the comparisons for all endpoints:

$$\hat{\tilde{\mathbf{R}}} = \begin{pmatrix} 1.000 & 0.361 & 0.334 & 0.463 & 0.165 & 0.126 & 0.463 & 0.154 & 0.118 \\ 0.361 & 1.000 & 0.135 & 0.162 & 0.440 & 0.049 & 0.162 & 0.412 & 0.046 \\ 0.334 & 0.135 & 1.000 & 0.148 & 0.059 & 0.360 & 0.148 & 0.055 & 0.337 \\ 0.463 & 0.162 & 0.148 & 1.000 & 0.361 & 0.330 & 0.468 & 0.156 & 0.119 \\ 0.165 & 0.440 & 0.059 & 0.361 & 1.000 & 0.134 & 0.166 & 0.424 & 0.047 \\ 0.126 & 0.049 & 0.360 & 0.330 & 0.134 & 1.000 & 0.127 & 0.047 & 0.290 \\ 0.463 & 0.162 & 0.148 & 0.468 & 0.166 & 0.127 & 1.000 & 0.361 & 0.327 \\ 0.154 & 0.412 & 0.055 & 0.156 & 0.424 & 0.047 & 0.361 & 1.000 & 0.134 \\ 0.118 & 0.046 & 0.337 & 0.119 & 0.047 & 0.290 & 0.327 & 0.134 & 1.000 \end{pmatrix}$$

Evaluation of the example

Results:

Comparison	ASAT	ALAT	ALP
1/0	1.100 (0.921)	1.052 (0.863)	0.953 (0.848)
2/0	1.118 (0.938)	1.104 (0.910)	0.765 (0.670)
3/0	1.118 (0.938)	0.995 (0.811)	0.702 (0.609)

Table: Upper limits (and estimates) per comparison and enzyme.

- e.g., for $\theta_i = 0.9$ for all endpoints $(1 \leq i \leq k)$, doses 2 and 3 show relevantly smaller values than the control for endpoint ALP
- moreover: e.g., dose 3 yields a reduction to 0.702 times the values for ALP or less compared with the control (with simultaneous coverage probability 0.95)

Summary

New method: ...

- uses a multi-t distribution involving the endpoints' estimated correlations
- is a size-α test
- gives adjusted p-values and SCI available for each endpoint and comparison
- is for ratios as well as differences of means
- implies the "normal" Dunnett test as a special case (see [2, 1])
- is also checked for coverage probability and power (not shown here)
- is already extended for general multiple contrasts and possibly heterogeneous covariance matrices (not shown here)

References

[1] G. Dilba, F. Bretz, V. Guiard, and L. A. Hothorn. Simultaneous confidence intervals for ratios with applications to the comparison of several treatments with a control. *Methods Of Information In Medicine*, 43(5):465–469, 2004.

[2] C. W. Dunnett. A multiple comparison procedure for comparing several treatments with a control. *Journal Of The American Statistical Association*, 50(272):1096–1121, 1955.

[3] E. C. Fieller. Some problems in interval estimation. *Journal of the Royal Statistical Society, Series B*, 16:175–185, 1954.

[4] A. Genz, F. Bretz, and R port by T. Hothorn. *mvtnorm: Multivariate Normal and T Distribution*, 2008. R package version 0.8-3.

[5] T. Hothorn, F. Bretz, and A. Genz. On multivariate t and Gauß probabilities in R. *R News*, 1(2):27–29, 2001.

[6] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2007. ISBN 3-900051-07-0.

[7] A. Schulte, J. Althoff, S. Ewe, and H. B. Richter-Reichhelm. Two immunotoxicity ring studies according to OECD TG 407 - Comparison of data on cyclosporin A and hexachlorobenzene. *Regulatory Toxicology And Pharmacology*, 36(1):12–21, 2002.

[8] Y. L. Tong. *The multivariate normal distribution*. Springer: New York; 191-3., 1990.