Association between different-scaled multiple metabolomics endpoints and clinical predictors

Ludwig A. Hothorn

(retired) Institute of Biostatistics, Leibniz University Hannover, Germany, hothorn@biostat.uni-hannover.de

Joint work with P. Ferrario, Max-Rubner-Institute, Karlsruhe and T. Hothorn, University Zurich, Switzerland

THE PROBLEM

Metabolic analytes: multiple correlated endpoints, $p > n$ problem, **unlikely to be similarly distributed**. Some are left-censored because a lower detection limit

Association translated into finding simple function between clinical covariate and some metabolomics marker

Toy example Covariate Age (X) with: 1) GUDCA left-censored at value 25 (Y_1, 2) C10.0 no censoring but many ties (Y_2) (N=291)

Find the arbitrarily shaped $Y_i = f(X)$ relationships for arbitrarily distributed endpoints Y_i

Most likely transformation

- Unrealistic: assume the same distribution for each of the many analytes
- Metabolomic analytes: a **mixture of complete measured and left-censored** analytes
- **mlt** $\Rightarrow$ maximum likelihood estimators in the class of conditional transformation models [1]- providing a **comparable** analysis of such quite differently-scaled analytes
- I.e. both complete and censored analytes can be analysed by an unique approach AND different distributed analytes (left-skewed, right-skewed, curtosis, tied) can be analysed jointly

```
library(mlt)
yvar2 <- numeric_var("C10.0", support =
 quantile(dd$C10.0, prob = c(.1, .9)))
# Bernstein polynomial
yb2 <- Bernstein_basis(yvar2, ui = "increasing", order = 5)
# Conditional transformation model
ma2 <- ctm(yb2, shifting = ~ Age, todistr = "Normal",
        data = dd)
fma2 <- mlt(ma2, data = dd) # mlt
```

COLR

- **Continuous outcome logistic regression** [2]: association between a continuous covariate and an arbitrarily distributed analyte, independently of scale and of certain cut-offs
- It provides the effect size **odds ratios** with confidence intervals
- Dimensionless OR: comparable over different-scaled analytes

```
library(mlt)
ma2OR <- ctm(yb2, shifting = ~ Age, todistr = "Logistic",
     data = dd)
```

Tukey trend test

- **Maximum test on three quasi-linear regression models for the arithmetic, ordinal, and logarithmic dose** metameters [5]
- Covers a wide range of dose-response patterns
- Recent GLMM-generalization and CRAN-library(tukeytrend) [4]
- Basic principle: multiple marginal models (mmm) [3] estimating the correlations between GLM-models
- ξ multiple linear regression models for the ξ dose transformation functions $\psi^{\xi}(D_j)$ for a vector of response variables y_{ijk} with $i = 1, ..., I$ multiple endpoints in $j = 0, ..., J$ dose levels with k_j unbalanced replicates $y^{\xi}_{ijk} = \alpha_{i\xi} + \beta_{i\xi}(\psi^{\xi}(D_{jk})) + \epsilon_{\xi ijk}$

CRAN libraries

```
library("mlt"); library("tukeytrend"); library("multcomp")
(dd$GUDCAsurv <- with(dd, Surv(GUDCA, event = GUDCA > 25, type = "left")))
yvar <- numeric_var("GUDCAsurv", support = quantile(dd$GUDCA, prob = c(.1, .9)))
yb <- Bernstein_basis(yvar, ui = "increasing", order = 5)
yvar2 <- numeric_var("C10.0", support = quantile(dd$C10.0, prob = c(.1, .9)))
yb2 <- Bernstein_basis(yvar2, ui = "increasing", order = 5)
TukeyMetam<-dosescalett(data=dd, dose="Age", scaling = c("ari", "ord","log"))
m_dat<-TukeyMetam$data # Tukeys dose metameters
lma <- ctm(yb, shifting = ~ Ageari, todistr = "Logistic", data = m_dat)
lmO <- ctm(yb, shifting = ~ Ageord, todistr = "Logistic", data = m_dat)
lmA <- ctm(yb, shifting = ~ Agelog, todistr = "Logistic", data = m_dat)
lfma <- mlt(lma, data = m_dat)
lfmO <- mlt(lmO, data=m_dat)
lfmA <- mlt(lmA, data=m_dat)
....
lfma2 <- mlt(lma2, data = m_dat)
lfmO2 <- mlt(lmO2, data=m_dat)
lfmA2 <- mlt(lmA2, data=m_dat)
####### bivariate Multiple endpoints test CORL
BBl <- glht(mmm(ari=lfma, log=lfmO, ord=lfmA, ari2=lfma2, log2=lfmO2, ord2=lfmA2),
           mlf(ari="Ageari=0", log="Ageord=0", ord="Agelog=0",
               ari2="Ageari=0", log2="Ageord=0", ord2="Agelog=0"))
multEndpointl<-summary(BBl)
```

Result

	OR	simul. lowerCI	adj p-value
GUDCA:ari	1.028	1.014	0.0000022
GUDCA:ord	1.005	1.003	0.0000023
GUDCA:log	3.436	1.864	0.00000057
C10.0:ari	0.993		0.49
C10.0:ord	0.999		0.50
C10.0:log	0.769		0.54

References

[1] T. Hothorn, L. Most, and P. Buhlmann. Most likely transformations. *Scandinavian Journal of Statistics*, 45(1):110–134, March 2018.

[2] T Lohse, S Rohrmann, D Faeh, and T Hothorn. Continuous outcome logistic regression for analyzing body mass index distributions. *F1000Research*, 2017, 6:1933 (doi: 10.12688/f1000research.12934.1).

[3] C. B. Pipper, Ch. Ritz, and H. Bisgaard. A versatile method for confirmatory evaluation of the effects of a covariate in multiple models. *JRSS-C*, 61:315–326, 2012.

[4] F. Schaarschmidt and L.A. Hothorn. library(tukeytrend). 2018.

[5] J. W. Tukey, J. L. Ciminera, and J. F. Heyse. Testing the statistical certainty of a response to increasing doses of a drug. *Biometrics*, 41(1):295–301, 1985.