As Different as Chalk and Cheese: Meta-statistics for Biomarker Detection

Ron Wehrens Pietro Franceschi

Biostatistics and Data Management http://cri.fmach.eu/BDM

StatSeq Workshop, Verona April 18–19, 2012



프 🖌 🛪 프 🕨

Biomarker identification (2-class situation)





< ∃→

Biomarker identification (2-class situation)





> < ≣ >

Ron Wehrens Meta-statistics for Biomarker Detection

Context	Tools	Challenges
Univariate	statistical tests	multiple testing correlations
Multivariate	PCLDA, PLSDA, VIP, variable selection	model validation optimization criterion local optima



æ

臣 🕨 🔸 臣 🛌

Context	Tools	Challenges
Univariate	statistical tests	multiple testing correlations
Multivariate	PCLDA, PLSDA, VIP, variable selection	model validation optimization criterion local optima



æ

臣 🕨 🔸 臣 🛌

Spike-in apple data



- 10 control apples
- 10 treated apples spiked
- 9 chemical compounds
- 3 sets of concentrations
- ESI+ and ESI-
- In each: 22 "true" biomarkers

Total: six comparisons between treated and control

Franceschi et al. - J. Chemom. (2012)



One apple sample – the result





.⊒⇒

One apple sample – the result





э

- "God loves 0.05 just as much as 0.06" selecting optimal thresholds
- Second-level of significance testing (Tukey, 1976)
- "Real differences are rare and weak"
- Here: extension to multivariate methods
- Compare to "standard" thresholds: $\alpha = 0.05$, *VIP* = 1

Dohono and Jin, PNAS (2008) Wehrens and Franceschi – submitted.





() < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < ()



Higher Criticism: theory

$$HC = \max_{i} \frac{\sqrt{N}(i/N - p_i)}{\sqrt{i/N(1 - i/N)}}$$





ъ

Results for apple data: selection efficiency



Ron Wehrens Meta-statistics for Biomarker Detection

Central idea: good biomarkers are stable

- Perturb data by subsampling (samples as well as variables)
- Fit many models on many perturbed data sets
- Which variables come up consistently at the top?
- Parameters: defining subsampling, consistency

Wehrens *et al.*, Anal. Chim. Acta (2011) Meinshausen and Bühlmann, J. Roy. Stat. Soc. B (2010)



★ E > < E >

Simulations based on real apple data



⇒ < ⇒ >

Biomarker selection – conclusions

- Both SBBS and HC form viable alternatives
- Much better than "standard" thresholds
- HC: very few parameters
- ... but time-consuming
- SBBS: faster
- ... but more settings
- ... and you need > 10 replicates
- Both implemented in R package BioMark, available from http://cran.r-project.org



프 🖌 🛪 프 🕨

Acknowledgements



BDM:

Pietro Franceschi Matthias Scholz Nir Shahaf Yonghui Dong Nikola Dordevic

FQAN:

Fulvio Mattivi Urska Vrhovsek Domenico Masuero



< ∃→