

## Heidelberger Kolloquium Medizinische Biometrie, Informatik und Epidemiologie

Sehr geehrte Damen und Herren,

wir laden Sie herzlich ein zu dem Vortrag:

### “Quantification of model stability after variable selection“

von

**Christine Wallisch**

Medizinische Universität Wien

**am Montag, 17.12.2018, 16.15 Uhr**

im **ATV**-Raum des DKFZ, Im Neuenheimer Feld 242, 69120 Heidelberg

---

Data-driven variable selection methods, e.g., backward elimination or Lasso, are commonly used to identify relevant explanatory variables and to reduce the number of explanatory variables in multivariable statistical models. Applying these algorithms may lead to false inclusion (resulting in increased variance) or exclusion (resulting in a possible bias) of variables. Hence, stability investigations are needed to assess the robustness of a model with regard to slight modifications in the data.

Stability investigations can be carried out by using resampling methods to assess variable selection frequencies, model selection frequencies, and distributions of selected regression coefficients. Recently, we proposed to compute two additional measures (BiomJ, 2018):

- the relative conditional bias RCB, which quantifies the bias that is induced by variable selection relative to the coefficient in a global model, and
- the root mean squared difference ratio RMSDR, which expresses the additional uncertainty induced by variable selection by comparing the variance of selected regression coefficients with the variance of the regression coefficients in a global model.

We conducted a comprehensive simulation study to investigate whether bootstrap or subsampling is better suited to estimate these quantities. Bootstrap-based estimation was found to be suboptimal for variable and model selection frequencies, whereas subsampling-based estimation was inappropriate for the RMSDR. Thus, both resampling approaches have to be employed to calculate reliable stability measures.

Stability investigations supply important information for data analysts conducting data-driven variable selection. It should become a routine step whenever a statistical model is developed.

Heinze G, Wallisch C and Dunkler D. Variable selection – A review and recommendations for the practicing statistician. *Biometrical Journal*. 2018:1-19.

---

Alle Interessenten sind herzlich eingeladen!

Gezeichnet: Dickhaus, Kieser, Knaup, Kopp-Schneider, Wellek

#### Organisation: Birgit Schleweis

Institut für Medizinische Biometrie und Informatik, Im Neuenheimer Feld 130.3, 69120 Heidelberg Tel. 06221/56-4142. Bitte registrieren Sie sich hier für die Ankündigung der Vorträge per E-Mail: <https://web.imbi.uni-heidelberg.de/newsletter/>