Psychological Test and Assessment Modeling, Volume 61, 2019 (4), 363-364

Editorial

Tree-based methods for regression and classification – Statistical methods at the interface of graphics and statistics

Mark Stemmler[1] *& Alexander von Eye*[2]

The editors of this Special Issue invited scholars to contribute new formal and statistical developments as well as interesting applications in the field of tree-based methods in psychological research. Among the many interesting properties of tree-based methods is that they enable researchers to investigate effects of multiple independent variables on one or several dependent variables without any restriction on level of measurement.

The contributions included in this issue contain theoretical considerations, new methods for tree-based analysis, simulation studies, and sample analyses of interesting psychological data sets. Also included are software developments and applications (R packages such as confreq, rpart, dHISC and partykit). The reader will find articles that encompass classification and regression tree methods (CART), CHAID, and bootstrap regression tree methods. This special issue first presents a selection of new developments and theoretical contributions, and, second, a selection of applications of tree-based methods.

Section I: Theory and Methods

In the first manuscript of the theory and methods section, Michael P. van Wie, Xintong Li, and Wolfgang Wiedermann propose and discuss methods for the *Identification of confounded subgroups using linear model-based recursive partitioning*. The authors aim at detecting confounded subgroups in linear regression models by way of combining a confounder detection approach, based on kernel-based independence testing, with model-based recursive partitioning.

[1] *Correspondence concerning this article should be addressed to:* Prof. Dr. Mark Stemmler, Institut für Psychologie, Universität Erlangen-Nürnberg, Nägelsbachstraße 49c, 91052 Erlangen, Germany; email: mark.stemmler@fau.de

[2] Michigan State University, USA

The second article, contributed by Markus Fritsch, Harry Haupt, Friedrich Lösel, and Mark Stemmler, discusses data science alternatives to ordinary least squares: Investigating the risk factors for corporal punishment with decision trees and random forests. The authors demonstrate that decision trees and random forests enable data-driven modeling. These methods are valuable tools in psychological research to gain new insights and to validate existing results. In an application, Fritsch and colleagues examine the behavior of father's aggressiveness, dysfunctional parent-child interactions, and other risk factors for corporal punishment of children by their fathers.

This article is followed by *Analyzing Tree Structures with Configural Frequency Analysis and the R-package confreq* by Mark Stemmler, Jörg Henrik-Heine and Susanne Wallner. This article describes the use of Configural Frequency Analysis (CFA) for detecting a tree structure in data that were analyzed with the SPSS module Answer Tree. In addition, the application of the R package *confreq* is demonstrated. The data example is taken from a longitudinal study on deviant and delinquent behavior in juveniles.

In the last article of the theory and methods section, which is entitled *Log-linear and Configural Analysis of Tree Structures*, Alexander von Eye, Wolfgang Wiedermann and Stefan von Weber propose two methods for the analysis of existing tree structures. These methods are log-linear modeling and Configural Frequency Analysis (CFA). In a data example, students' decisions and life satisfaction are examined.

Section II: Applications

The first article of the application section is contributed by Cody Ding and Yuyang Zhao. The authors present on *Using tree-based regression to examine factors related to math ability among 15-year old students* to predict students' mathematics ability, particularly subgroups of students who share characteristics that are associated with different levels of math ability. Based on PISA 2012 data from the United States and China, the authors used regression tree analysis to select the most salient predictors of math ability, and identify subgroups of 15-year-old students who were likely to be proficient in math ability.

The second article of the application section, co-authored by Eun-Young Mun and Feng Geng is entitled *Predicting post-experiment fatigue among healthy young adults: Random forest regression analysis*. In this paper, a random forest regression analysis is used to predict post-experiment fatigue in a sample of 212 participants between the ages of 18 and 30 following a mildly stressful experiment. The random forest regression analysis is used together with a built-in cross-validation function.

We hope that the readers will enjoy this Special Issue.

Mark Stemmler & Alexander von Eye