Please join the virtual Biometric Colloquium

WILLI SAUERBREI AND EDWIN KIPRUTO

University of Freiburg, Germany

MULTIVARIABLE REGRESSION MODELLING - ISSUES IN SELECTION OF VARIABLES AND FUNCTIONAL FORMS OF CONTINUOUS VARIABLES

October 13th, 2021 at 9:00 am

JOIN WEBEX MEETING
https://meduniwien.webex.com/meduniwien/j.php?MTID=mc9c51afeaa2e7f804f11d648de6e342a

HOSTS: Georg Heinze

ABSTRACT:

In deriving a multivariable regression model, selection of variables and determination of functional forms for continuous variables are central issues. In a recent review of methods, selection bias and the role of shrinkage to correct for it were identified as important issues which require better understanding, further research and more comparisons of methods proposed. The nonnegative garrote (NNG) was the first proposed method that combine variable selection and shrinkage, however, it is often ignored in practice because of its dependence on ordinary least squares estimates from the full model which performs poorly in high degree of multicollinearity and cannot be used for the analysis of high-dimensional data. Using real datasets with different structures both in low and high-dimensional settings we will show that recent developments makes it possible to extend its potential applications. We will compare the performance of NNG with the lasso, adaptive lasso, relaxed lasso and best subset with and without post-estimation shrinkage.

Although severe weaknesses are known, categorisation or the assumption of a linear effects are still the most popular approaches to determine the functional form for continuous variables. The multivariable fractional polynomial (MFP) approach combines variable selection using backward elimination and function selection for continuous variables using fractional polynomial functions. Regression diagnostics, such as detection of influential points and plots of residuals may reveal lack of fit or other peculiarities of a selected model. Single observations

Wiener Biometrische Sektion
http://www.meduniwien.ac.at/wbs/

Vorstand:
Susanne Strohmaier, Martin Wolfsegger
Kontakt:
susanne.strohmaier@meduniwien.ac.at

may be responsible for the selection of a linear, a monotonic non-linear or even a non-monotonic function. Using simulated data we will discuss approaches to identify influential points, consider model reproducibility and investigate the effect of sample size on MFP models. We will show that with a sufficiently large sample size and a carefully check on regression diagnostics, MFP is a suitable approach to identify variables with a stronger effect and suitable functional forms for continuous variables.