Einladung zum Biometrischen Kolloquium

ARIS PERPEROGLOU

AstraZeneca, UK

A (ROUGH) GUIDE TO USING SPLINES IN MULTIVARIABLE REGRESSION

19. Februar 2020, 9:15h

Seminarraum CeMSIIS, Raumnummer 88.03.513

Medizinischen Universität Wien, Spitalgasse 23, 1090 Wien

HOST: Georg Heinze

ABSTRACT:

Splines are a powerful tool used to describe non-linear relationships in regression models. The splendour of splines relies on their flexibility and their ability to model a variety of cases, from simple curves to rather complex patterns. That flexibility comes at a price: fitting a spline can be a daunting exercise for analysts with limited previous experience. First, there are several different spline bases to choose from, many of which with a complicated mathematical form. Second, even with a specific spline base there are several parameters that must be optimised by the user, like knots, positioning of knots, degree of the polynomial, or degrees of freedom for the spline and in some cases the weight of the penalty. This talk will attempt to demystify spline fitting and illustrate that at their heart, splines are polynomials that can be fitted with little extra work and limited experience. We will illustrate how to fit common splines bases and discuss some of their properties. In the second part we will focus on penalised splines and how penalties can be utilised to obtain a useful model. When working in a multivariable setting, penalties can also be utilised for variable selection. We will introduce the problem of variable selection when fitting splines in a multivariable setting and compare backward selection methods with double penalisation in real and simulated data.