



WEBINAR

“Transformation Models: Pushing the Boundaries”

Prof. Torsten Hothorn

*Epidemiology, Biostatistics and Prevention Institute
Department Biostatistics, University of Zurich*

Chair: Prof. Paola Zuccolotto

Dott.ssa Ambra Macis

University of Brescia

September 27th, 2022

15.30 hrs CEST (UTC+2)

Sala Biblioteca, S. Faustino building

Live presentation on: Google Meet

<https://meet.google.com/dco-wixi-vwg>

ABSTRACT

Transformation models have been around since the seminar papers by John Tukey, George Box and David Cox published 60 years ago. The core idea is to transform a distribution of interest, which typically is a rather messy thing, into a nicely behaving distribution, such as the normal or logistic, prior to analysis. The literature mostly followed two distinct paradigms: Either, an appropriate transformation is somehow guesstimated without the actual analysis being even aware of such a data transformation having been applied in the first place or the transformation is treated as a nuisance parameter. Log-transforming count data or "Box-Cox-Transformations" are typical of the former approach and the partial likelihood estimation in Cox models sparked "semi-parametric" inference in similar models. These developments had tremendous success in many applied disciplines, yet there are limits of what can be done in this line of thinking. More recently, it was proposed to actually estimate the necessary data transformation explicitly. Thus, the actual model needs to be aware of data transformations and the uncertainty associated with them being estimated from data simultaneously with other model parameters. While there are some technical costs associated with such a procedure, it allows many previously hard problems to be solved rather conveniently. This talk will discuss some areas where fully parameterised transformation models are attractive alternatives to established statistical instruments, such as regression for discrete, skewed, bounded, or otherwise "difficult" responses, for count regression, in multivariate regression, in penalised regression, and in situations where observations are correlated in some way, most importantly for clustered data. Special emphasis will be given to trees and forests with a focus on the survival setting.