

## Please join the **Biometric Colloquium**

### **IRANTZU BARRIO**

University of the Basque Country

#### **A NEW GENERAL AND MULTIVARIABLE APPROACH TO CATEGORIZE PREDICTOR VARIABLES. APPLICATION IN COPD PATIENTS**

**March 22<sup>th</sup>, 2023 at 9:00 am**

Seminarraum Center for Medical Data Science (previously CeMSIIS),  
Spitalgasse 23, Room 88.03.513  
Medical University of Vienna, 1090 Wien

**HOST:** Florian Frommlet

#### **ABSTRACT**

The use of discretized variables in the development of prediction models is a common practice, partly because the decision-making process is more natural when based on rules created from segmented models. Although this practice is perhaps most common in medicine, it is extensible to any area of knowledge in which a prediction model supports the decision-making process. Therefore, providing researchers with a useful and valid categorization method may be a relevant issue when developing predictive models.

A methodology to select the optimal cut-off points to categorize a continuous covariate has been previously proposed in the context of logistic and Cox regression models, which was based on the maximization of the discrimination ability of the model measured by the c-index (Barrio et al. 2017a, Barrio et al. 2017b).

In this work, we propose a new general methodology, which can be applied to categorize a predictor variable in any regression model where the response variable belongs to the exponential family distribution. This new methodology can be applied in any multivariate contexts, moreover, allows to estimate cut-off points differentially for the different levels of a factor variable with which the variable to be categorized has a significant interaction. Furthermore, a computationally very efficient method is proposed to obtain the optimal number of categories. Several simulation studies have been conducted in which the efficiency of the method with respect to both the location and the number of estimated cut-off points is shown.

Finally, we applied this proposal to a real data set of patients with stable chronic obstructive pulmonary disease (COPD) in order to obtain optimal categorization proposals for continuous predictor variables related to patient's physical activity considering mortality and number of hospitalizations as outcome variables.

Barrio, I., Arostegui, I., Rodríguez-Álvarez, M.-X., and Quintana, J.-M. (2017a). A new approach to categorising continuous variables in prediction models: Proposal and validation. *Statistical Methods in Medical Research*, 26(6):2586–2602.

Barrio, I., Rodríguez-Álvarez, M. X., Meira-Machado, L., Esteban, C., and Arostegui, I. (2017b). Comparison of two discrimination indexes in the categorisation of continuous predictors in time-to-event studies. *SORT*, 1:73–92.