



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



# Seminar

## STRATEGIES TO HANDLE MANDATORY COVARIATES USING MODEL- AND LIKELIHOOD-BASED BOOSTING

**Riccardo De Bin**

**University of Oslo**

**December 17, 2018 | 12.30 p.m. | Aula Benvenuti  
Campus S. Caterina**

**Abstract :** [www.stat.unipd.it/fare-ricerca/seminari](http://www.stat.unipd.it/fare-ricerca/seminari)

# STRATEGIES TO HANDLE MANDATORY COVARIATES USING MODEL- AND LIKELIHOOD-BASED BOOSTING

## Riccardo De Bin

Associate Professor of Statistics and Biostatistics | Department of Mathematics | University of Oslo

Among the iterative methods exploited during recent years in statistical practice, particular attention has been focused on boosting. Originally developed in the machine learning community to handle classification problems, boosting has been successfully translated into the statistical field and extended to many statistical problems, including regression and survival analysis. In a parametric framework, the basic idea of boosting is to provide estimates of the parameters by updating their values iteratively: at each step, a weak estimator is fitted on a modified version of the data, with the goal of minimizing a loss function. Thanks to its resistance to overfitting, boosting is particularly useful in the construction of prediction models. Its iterative nature, moreover, allows straightforward adaptations to cope with high-dimensional data. In this talk, we first review and contrast two well-known boosting techniques, model-based boosting and likelihood-based boosting. We note that in the simple linear regression case they lead to the same results, provided there is a specific choice for their tuning parameters. This is not the case for more complex situations. As an example, we show the differences in survival analysis under the proportional hazards assumption. As a main contribution of the talk, we analyze strategies to include mandatory variables, i.e. those variables that for some reasons must enter in the final model, in a statistical model using the two boosting techniques. In particular, we examine solutions currently only considered for one and explore the possibility of extending them to the other. We show the importance of a good handling of mandatory variables in a real data example.