ROC2025 - International Workshop on ROC Analysis and Related Topics
Vigo, January 23 and 24, 2025

Prediction in the presence of missing values: no credible alternative to imputation-based use of the predictive density

(Predictive averaging, Complete Cases, Indicator Methods and Pattern Submodels)

Bart J. A. Mertens[1]

[1]Department of Biomedical Data Sciences, Leiden University Medical Centre, The Netherlands

Abstract

Prediction in the presence of missing values is a complex and still poorly understood problem, particularly when future records also contain missing values. Mertens, *et al.* (2020) demonstrate that with non-linear models (such as logistic regression or Cox survival) and when using imputations, averaging of multiple predictions obtained from distinct models fitted on imputed data should be preferred to pooled models. In this talk we contrast predictive averaging with complete-case-based model calibration (CC) as well as use of missing-indicator (IDX) and Pattern Submodel (PS) approaches. We demonstrate that only predictive averaging guarantees required coverage levels in prediction. Scoring and class-separation measures (such as Brier or AUC) strongly favour IDX and PS methods however. We show this is due to the biased nature of these methods, which (Brier) scoring or AUC measures do not correct for.

References

Mertens, Banzato and de Wreede (2020). Construction and assessment of prediction rules for binary outcome in the presence of missing predictor data using multiple imputation and cross-validation: methodological approach and data-based evaluation. *Biometrical Journal*, 62, 724-741.

Mertens and de Wreede. Calibration of prediction rules for life-time outcomes using prognostic Cox regression survival models and multiple imputations. Technical report. *ArXiv:2105.01733*