Flexible ROC regression analysis

María Xosé Rodríguez-Álvarez[1], Javier Roca-Pardiñas[2], Ignacio López-de-Ullibarri[3] and Carmen Cadarso-Suárez[1]

[1] Department of Statistics and Operations Research, University of Santiago de Compostela, Spain
[2] Department of Statistics and Operations Research, University of Vigo, Spain
[3] Department of Mathematics, University of A Coruña, Spain

Continuous diagnostic tests are often used to discriminate between diseased and healthy populations. The receiver operating characteristic (ROC) curve is a widely used tool that provides a graphical visualization of the effectiveness of such tests. The potential performance of the tests in distinguishing diseased and healthy people may be strongly influenced by a set of continuous and factor-type covariates and a variety of regression methods for adjusting ROC curves has been developed. The induced methodology [1], [2] is based on the use of regression models for the result of the test, both in the healthy and disease populations. Covariate effects on the associated ROC curve can be then computed by deriving the induced form of the ROC curve. So far, this methodology assumes parametric forms for the covariate effects. In this work we extend the induced methodology by allowing for arbitrary nonparametric effects of continuous covariates and their possible interactions with factors. On this purpose, generalized additive models (GAM) including factor-by-continuous interactions are used [3], [4]. Our method allows for covariate effects not only on the mean but also on the variance of the test. From the induced ROC curve, summary measures of the accuracy, such as the area under the curve (AUC) and the generalized Youden index (YI), are obtained. Bootstrap-based methods are used for constructing confidence intervals for these measures, and also for testing significant covariate effects on the ROC curve. To illustrate the method, endocrine data was analyzed with the aim of assessing the performance of anthropometric measures in predicting clusters of cardiovascular risk factors in a population of adult Galician people (NW Spain), adjusting for age and gender.

References

[1] Pepe, M.S. (1998). A regression modelling framework for receiver operating characteristic curves in medical diagnostic testing. *Biometrika*, **84**, 595-608.

[2] Faraggi, D. (2003). Adjusting receiver operating characteristic curves and related indices for covariates. *The Statistician*, **52**, 179-192.

[3] Hastie, T.J. and Tibshirani ,R.J. (1990). *Generalized Additive Models*. London: Chapman and Hall.

[4] Roca-Pardiñas, J., Cadarso-Suárez, C., Nácher, V. and Acuña, C. (2006). Bootstrap-based methods for testing factor-by-curve interactions in Generalized Additive Models: assessing prefrontal cortex neural activity related to decision-making. *Statistics in Medicine*, **25**, 2483-2501.