Fuzzy tools for constructing robust statistical models of dynamic systems in environmental epidemiology

Michael Friger, Arkady Bolotin

Department of Epidemiology, Ben-Gurion University of the Negev, P.O.B. 653 Beersheba, Israel 84105
friger@bgumail.bgu.ac.il; arkadyv@bgumail.bgu.ac.il

Modeling biological and health systems plays extremely important role in health and life sciences. One of the important aspects in the construction of statistical model is its robustness. There exist different definitions of this term.
We understand robustness of the generalized linear regression model as a resilience of the model to the bifurcation when the cut-off point changes (by bifurcation we mean dramatic abrupt changes in the model structure, strength and sign of the coefficients) .So, robustness of the regression coefficients and the stability of the model structure are of our particular interest. Practically this means that any cut-point is given in some “natural” interval, within which the model behavior is suppose to be persistent. We will call this interval of choice.
Such aspect of the model robustness is quite important for the regression models with the binomial (multinomial) outcomes resulted from categorized continuous latent variables. When categorizing a continuous dependent variable, a set of cut-points is used. Thus, the resulted regression model may affect (and it does depend) on these cut points as in the structure, strength and sign of the predictors.
The aim of our paper is to propose, analyze and discuss a new fuzzy approach to the construction of the regression models of dynamic and hierarchical systems that are robust against instability in categorization of the continuous latent dependent variables.
The main tool for the proposed approach is Stability Theorem proved by the authors. The theorem states that for the given latent dependent variable and interval of choice there exists some "actual" interval of choice of cutpoints such that for any cut point from this interval a generalized linear model with a symmetric fuzzy ramp membership function as a dependent variable is robust.
This means that for any cut point from the certain interval, the models keep the same structure of significant coefficients; the corresponded coefficients have the identical sign and near the same value. Extensive computer experiments show that this Theorem is correct for others families of the fuzzy membership functions (triangular, exponential, asymmetrical “ramp”, S-shape functions).
So, the proposed approach includes the following 3 stages:
1. Building of an appropriate fuzzy membership function for the given latent dependent variable and interval of choice.
2. Substitution of the given binomial dependent variable with the constructed fuzzy membership function in a generalized linear model.
3. Solution of the obtained model and determination of the "actual" interval of choice of cut-points.
We applied this approach for modeling health effects of air pollution on the real-life dataset.
We constructed two types of the model depending on the type of the outcome variable we dealt with: the logistic and linear regression of the fuzzy outcome. The obtained logistic regressions are structurally instable toward the cut-point location. The models with fuzzy outcome are robust for each cut point from "actual" interval of choice and for each type of tested membership functions.
The found results demonstrate that the proposed approach is effective and resulted models are stable.