

# Logic regression model versus classical linear and logistic regression models

Magdalena Malina<sup>1</sup>

## SUMMARY

We consider a biological problem of locating multiple interacting genes. Logic regression ([1]) is a regression method, specifically aimed at detecting high order interactions in genetic data. The method attempts to construct predictors as Boolean combinations of dummy variables coding marker genotypes. The method searches for these combinations of predictors in the entire space of such combinations, which are represented by logic binary trees. When there are many genes however, then the multiple testing problem arises and in a classical model selection criteria we need an additional penalty for model dimension ([2]). In Bayesian versions the problem of the penalty choice is replaced by a problem of proper selection of prior distributions.

We describe theoretically constraints under which the logic regression is more effective in detection of complex interactions of variables than the classical linear regression. We perform a comparative analysis of powers of these methods with the level of FWER fixed. Considering the problem of detection of interacting genes as a problem of multiple testing, we show theoretically that logic regression may lead to a larger power than the standard methods based on a linear or logistic regression model.

**Keywords:** logic regression, model selection, multiple testing

**AMS Classification:** 62J05, 62J15

## References

- [1] I. RUCZINSKI, C. KOOPERBERG AND M. LEBLANC (2003). Logic regression. *J. Comput. Graphical Statist.* **12**(3), 474–511.
- [2] M. BOGDAN, J. K. GOSH AND M. ŻAK-SZATKOWSKA (2008). Selecting explanatory variables with the modified version of Bayesian Information Criterion. *Quality and Reliability Engineering International.* **24** 627–641.

<sup>1</sup>Mathematical Institute, Wrocław University,  
pl. Grunwaldzki 2/4, 50-385 Wrocław  
Magdalena.Malina@math.uni.wroc.pl