

# Clusterwise Regression using Mixtures of Regression Models

Kristýna Vaňkátová<sup>1</sup> and Eva Fišerová<sup>1</sup>

<sup>1</sup>*Palacký University, Czech Republic*

## Abstract

The classical linear regression approach is reasonable under the assumption of a homogeneous population. Nevertheless, when we suspect that there are several heterogeneous groups in the population represented by a given sample, then mixtures of regression models can be applied. The method acquires parameters estimates by modelling the mixture conditional distribution of the response given the explanatory variable.

Provided the mixture consists of  $c$  components, the mixture distribution is given by the weighted sum over all  $c$  components. Parameters of a mixture of linear regression models are estimated by maximum likelihood using the expectation maximization (EM) algorithm. In order to improve regression parameter estimates and data classification, mixture model can be extended to include concomitant variables. These additional variables influence the weights of a mixture regression model so they are no longer deterministic but they operate as functions of one or more concomitant variables.

Recently, mixture models are used more and more in a various fields, including the economics. The methodology will be illustrated on an analyses of the relationship between an old age pension and income of EU countries residents older than 65 years.

## Keywords

Cluster analysis. Mixture regression models. Linear regression. EM algorithm. Old age pension. Income.

## References

- [1] Bengalia, T., Chauveau, D., Hunter, D. R., Young, D. S.: Mixtools: An R Package for Analyzing Finite Mixture Models, *Journal of Statistical Software* **32(6)** (2009)
- [2] Grün, B., Leisch, F.: FlexMix Version 2: Finite Mixtures with Concomitant Variables and Varying and Constant Parameters, *Journal of Statistical Software* **28(4)** (2008)
- [3] McLachlan, G., Peel, D.: *Finite mixture models*. New York: John Wiley & Sons, 2000.