

Module 3. Generalized Linear Models: Principles and Applications in R Software

- Date: May 21st to 25th, 2018
- Time: 9h-12h A.M.
- Venue: LABEF Conference Room
- Trainers: Prof. Romain Glèlè Kakaï, Wilfried Bonou (PhD Student), Bruno Lokonon (PhD-student)

In statistical modelling, the general linear model (e.g., linear regression, ANOVA, ANCOVA, etc.) has long been used in many situations. But this model is based on a strong hypothesis: the error term follows a normal distribution with the same variance. However, there are many phenomena (biological, physical, social, etc.) where this hypothesis is generally not satisfied. A good example is the counting of rare events. For this example, considering the variable to be explained (and its error) as following a normal distribution assumes a symmetric distribution around the mean, which is probably not the case, since the majority of the data collected can for example have more frequently zero or one. In addition, for the general linear model, it is assumed that the predictive variables have a linear effect on the measured variable (this effect is reflected by the regression coefficients), and some of these effects may not be linear in reality. For example, the number of eggs laid by a chicken may not change linearly with the chicken age. In order to analyze such data, several solutions are available. The most common is to find a mathematical transformation of the variable to be explained to make it normal. These transformations are not all effective, and their normalizing effect is sometimes difficult to quantify. Moreover, it is obviously preferable to use the original data rather than their transformed values to make interpretation of the results easier. It is for these multiple reasons that generalized linear models were developed by Nelder and Wedderburn (1972) to overcome the shortcomings of the general linear model. Many recent developments were made, many of which are already available under the free software R for efficient data analysis.

Under the supervision of Prof. Romain GLELE KAKAÏ (Full Professor of Biostatistics and Forestry, Director of LABEF), this statistical training module of LABEF will introduce you to generalized linear models. During the module, you will:

- learn how to identify the factors in models as well as the distribution of the response variable;
- be introduced to the principles of generalized linear models;
- write R-scripts for generalized linear models of the Poisson and Binomial family;
- interpret and report the results of a data analysis using generalized linear models.