

Module 3: Longitudinal data (or repeated measures data) analyses in R

Trainers: Prof. Romain Glèlè Kakaï, Aubin Amagnidé (PhD Student), Merveille Savi (MSc. Eng.)

Most often, studies in biological sciences are designed using cross-sectional data (type of data collected by observing many subjects (such as individuals, firms, countries, or regions) at the same point of time, or without regard to differences in time) [click here for more details: https://en.wikipedia.org/wiki/Cross-sectional_data]. However, some studies require repeated measurements taken in sequential from the same experimental unit over time. Such data are often referred to as longitudinal or panel data [click here for more additional details: https://en.wikipedia.org/wiki/Panel_data]. For instance, in Agronomy, a research work can be interested in studying whether different seed provenances and pre-treatments affect the germination and growth dynamics of a given plant species. For this purpose, the germination rate of seeds and growth parameters (e.g. total height, collar diameter, number of leaves, etc.) can be measured over a certain period of time based on an appropriate design of experiment (e.g. each week for 6 months). So, efficient methods are designed to analyse this kind of data due to the serial correlation between repeated measures. How to account for the auto-correlation (correlation between observed data) structure is a pivotal issue when analysing such data. Firstly, the module will introduce participants to the principles of longitudinal data analyses. At a second step, main emphasis will be on classical approaches and their limitations. Finally, longitudinal data analysis using modern approaches will be discussed with the participants. The main points to cover include:

- analysis of variance on repeated measures
- Specification of covariance structure
- Multivariate analysis of variance
- Applications in R
- Scientific report of longitudinal data analyses