

Course on compositional data and their analysis

Goals

The course aims to provide attendees with the theoretical and computational elements necessary to develop applications involving compositional data. Compositional data appear frequently in a multitude of applied scientific fields. These fields include biology in all its branches, from microbiology and genetics to biomedicine and ecology; geology, geochemistry, and Earth Sciences in general; Tourism, sociology, psychology, and all branches of engineering. That is, compositional analysis is useful for many applications. Course attendees are expected to present their own problems with compositional data analysis.

Prerequisites

It is recommended to have an elementary background in linear algebra and statistics. The course is partially based on the use of R and some of its libraries. It is recommended to attend with a laptop that has R installed (preferably with R-studio) and CoDa- Pack (free software developed by the group of compositional data analysis in the Universidad de Girona, Spain).



Juan Jose Egozcue, Professor Emeritus ,
Polytechnic University of Catalonia (Universitat
Politècnica de Catalunya)



Vera Pawlowsky-Glahn Professor Emeritus,
University of Girona,
Spain Department of Computer Science, Applied
Mathematics, and Statistics



5/02/2024 to 8/02/2024

Registration (**until 01/02/2024**),
maximum number of 20 participants):

- Internal participants (IPCB) – €100
- External participants – €200

Payment by bank transfer to:

INSTITUTO POLITÉCNICO DE CASTELO BRANCO

NIF:504 152 980 | IBAN: PT50 0781 0112 0000 0002 9912 2

SWIFT/BIC: IGCPPTPL

Proof of payment must be sent to the email teresa@ipcb.pt,
with an indication of the name, address, and Taxpayer
Identification Number (NIF) for receipt issuance.



39.819534 °N
7.512390 °W

Course on compositional data and their analysis

Program

Monday

- 1. Presentation (30min)** - Description and organization.
- 2. Compositional data (CoDa): introduction and representation on the simplex (1:30h)** - Motivation, examples, spurious correlation. Representation of CoDa (using CoDaPack).
- 3. Sample space and Aitchison geometry (2h)** - The simplex as a sample space. Subcompositions. Principles of compositional analysis (CoDA): scale invariance, subcompositional coherence, permutation invariance. Operations: perturbation, powering. Distance, norm, and inner product. Euclidean space of compositions.
- 4. Data presented by attendees (1h)** - These data will be discussed during the course.

Tuesday

- 4. Data presented by attendees Cont. (1h)** - These data will be discussed during the course.
- 5. Coordinate expression (1:30h)** - Centered log-ratio transformation (clr). Orthonormal bases in the simplex and their orthogonal coordinates (ilr/olr). Building ilr coordinates. Straight lines and ellipses in ilr-coordinates (using CoDaPack).
- 6. Tools for descriptive statistics (2:30h)** - Variability, center, and total variance. Variation matrix and Proportionality Index of Parts (PIP). Total variation decomposition. CoDadendrogram (using CoDaPack). Principal coordinates and CoDa-biplot (using CoDaPack).

Wednesday

- 7. Irregular data (1:30h)** - Types of zeros: essential, under detection limit, counting zeros. Zero patterns as factors. Substitution of zeros. Dealing with counting zeros. Missing data.
- 8. Compositional response in regression (1h)** Introduction to compositional software in R. Examples of regression with compositional response (using R).
- 9. Regression with compositional explanatory variables (1h)** - Examples of regression with compositional explanatory variables (Using R).
- 10. Linear models and reduction of variables (1:30h)** - Regression and discriminant analysis with compositional explanatory variables. Methods: R:selbal and R:coda4microbiome.

Thursday

- 11. Case discussion (4h)** - Including cases from attendees.
- 12. Conclusions (Thursday 1h)** - Summary and closing