

## Seminar on Compositional Data Analysis in Tourism

### Lecturer:

Berta Ferrer-Rosell - University of Lleida ([www.ferrer-rosell.com](http://www.ferrer-rosell.com))

### Objectives:

Compositional data analysis has begun to be used in various fields of the social sciences, which often face similar research questions. For example, in marketing, typical research questions are related either to the distribution of a whole (e.g., quotas or allocations), or to relative importance (e.g., dominance, profile, prevalence). Regarding this, Morais et al. (2018) studies the determinants of market share in the automotive industry, Ferrer-Rosell & Coenders (2018) segment tourists based on how they distribute the travel budget, and Blasco-Duatis et al. (2019) study the relative importance of content in online communication through Twitter.

The motivation for this course is to increase awareness of the peculiarities of compositional data (CoDa) and to help researchers avoid the most common problems in CoDa analysis. The topic is timely, given the large presence of compositional data sets in all areas of the social sciences, from marketing and economics, through urban planning or tourism

The course aims for participants to be able to recognize databases that are compositional in nature and to be able to explore compositional data both numerically and graphically. We will also work on data representation tools in log-ratio coordinates as well as using these coordinates in subsequent multivariate analyses. Finally, the course also seeks to treat irregular data in compositional data and introduce participants to the specialized CoDaPack software.

### Contents:

1. Introduction: definition of composition, compositional research questions and transformations (logarithms of ratios)
2. CoDa in practice: exploratory analysis (center, variation matrix, biplot)
3. Balances and CoDa dendrogram
4. Data process: treatment of zeros and other irregularities in the data
5. Multivariate analysis: cluster and MANOVA
6. Introduction to CoDaPack software

### Methodology:

The seminar will provide participants an introduction to the theoretical, but mainly practical issues of statistical analysis of compositional data, and will be mainly focused to the tourism field.

The seminar will consist of two sessions of 3h30' (7 hours in total), where they will work theoretical and practical contents. The practical work will be carried out with the free software CoDaPack (<http://ima.udg.edu/codapack/>). Participants will be required to install the software beforehand.