

Short Course and Open Seminar on Compositional Data Analysis

UPC-Barcelona,
Spain
5 - 9 July, 2010

Contents:

Compositional data (CoDa) are defined as vectors of positive components and constant sum, usually 100% or 1. These conditions render most classical statistical techniques useless on compositions, as they were devised for unbounded real vectors. However, there are many more types of data having the same limitations: as soon as the variables of a data set show the relative importance of some parts of a whole, data must be considered compositional. Examples of disguised compositions are data presented in ppm, ppb, molarities, or any other concentration units. Aitchison introduced the log-ratio approach to analyse CoDa back in the eighties. His solution was based on transforming the data vector with some log-ratio transformations, and applying classical techniques to the scores so obtained. This became the foundation of modern compositional data analysis, nowadays based on an own geometric structure for the simplex, the sample space of compositional data. This course will present the current state of the art in this field of active research, covering these topics:

- 1.- Hypothesis underlying statistical data analysis (sample space, scale). The spurious correlation. The relative information of compositions. Posing right questions. Principles of CoDa analysis: scale invariance, permutation invariance and sub-compositional coherence.
- 2.- The Aitchison geometry of the simplex. Perturbation, powering, scalar product, distance. Angles and shapes.
- 3.- Coordinate representation. Orthonormal basis, balances, and subcompositions. CoDa-principal component analysis. Normal distribution on the simplex.
- 4.- Exploratory analysis of CoDa. Variation array, biplot and CoDa-dendrogram. Kinds of zero, and their treatment in practice.
- 5.- Processes in the simplex: decay/growth evolution vs. mixture. Linear regression of a compositional response on external covariates, comparison of centers of populations (ANOVA)
- 6.- CoDa discriminant analysis and logistic-regression. Looking for the best coordinates to separate existing sub-groups.

Exercises are done with freeware CoDaPack3D on Excel®, and package "compositions" from the open source statistical environment R. It is strongly recommended that attendants have undergone some first semester courses on statistics, algebra and calculus. Basic knowledge about multivariate statistics may also be handy.

Practical information:

The course costs 300 euros, including course materials and coffee breaks. Inscription starts on Dec. 1, 2009 and will be first come, first served. Funding needs can be partly covered by a grant programme. The course will be held at the North Campus of the Technical University of Catalonia (UPC) in Barcelona, from Monday 5 to Friday 9 of July 2009, 15:30 to 20:00h.

Thanks to its cultural attractions and mediterranean climate, Barcelona is a world-class touristic center: it is well communicated with most important cities and airports in Europe, and offers a wide accommodation selection for all kinds of budget. However, the city is very busy in July: we recommend not to leave flight and accommodation issues for the last moment!

Visit our website for more information on any of these subjects.

Instructors:

Prof. Dr. Juan José Egozcue is professor of applied mathematics at the UPC. He works on statistics and probabilistic modelling in earth and environmental sciences.

Dr. Raimon Tolosana-Delgado is fellow researcher at the UPC. His expertise include geostatistics and statistical modelling in earth and environmental sciences.

Prof. Dr. Vera Pawlowsky-Glahn is professor of statistics at the University of Girona (Spain). She is an expert on statistics for compositional data and variables on constrained sample spaces.

CoDaWork 2011: 4th Workshop on Compositional Data Analysis
Girona, May 2011

Visit the website regularly for updated information!



Dpt. Matemàtica Aplicada III
c/Jordi Girona 1-3, E-08034
Barcelona (Spain)



Technical secretariat: CIMNE
International Center for Numerical Methods in Engineering
c/Jordi Girona 1/3 E-08034 Barcelona (Spain)

codawork@cimne.upc.edu
<http://congress.cimne.com/codawork>



Dpt. Informàtica i
Matemàtica Aplicada
c/Universitat de Girona s/n
E-17071 Girona (Spain)