

## Applied Explorative Data Analysis: Statistical Tools for Dimensionality Reduction

Dr. Alfonso Piscitelli	(e-mail): alfonso.piscitelli@unina.it (phone): +39 0812537450 (address): Department of Agricultural Sciences Division of 'Food Science and Technology' Via Università 100, Parco Gussone (Building 84) ZIP code 80055 - Portici (Napoli), Italy
------------------------	--

Number of CFU:  <b>5</b>	<b>Activities</b>	<b>Lectures</b>	(hours) <b>15</b>
		<b>Laboratory</b>	(hours) <b>20</b>
		<b>Seminars</b>	(hours) <b>10</b>
		<b>Other activities</b> (please indicate the activity): <i>Individual study of the course topics and data analysis using appropriate R packages</i>	(hours) <b>80</b>

### Objectives

The aim of the course is to provide Ph.D. students with the logic and statistical methodologies focused on analysis of high dimensional data through tools of dimensionality reduction. Ph.D. students will be able to extract useful and value-added information content from the data to produce knowledge directed at decision-making processes in the field of Food Science.

“Vast amounts of data are being generated in many fields, and the statisticians job is to make sense of it all: to extract important patterns and trends, and understand what the data says. We call this learning from data” (Hastie et al., 2009. The elements of Statistical Learning).

### Learning outcome

Ph.D. students will acquire application skills, that is they will be able to transform a real problem in the food and technological field into a statistical problem, detecting and / or selecting the data appropriately, identifying one or more appropriate methods for data analysis, correctly interpreting the results of statistical analysis, developing the reporting necessary for decision-making purposes.

Applications will be made with R software.

### Topics (with the indication of the n of hours/topic)

Introduction to R (2h Laboratory); Principal Component Analysis and application (7h Lectures and Laboratory); Correspondence Analysis and Multiple Correspondence Analysis (12h lectures and Laboratory); Multidimensional Scaling (7h lectures and Laboratory); Multidimensional Unfolding (7h lectures and Laboratory).

### Evaluation

Research paper – containing: research questions, data description, data analysis and conclusions – analysing Ph.D. students' own data

### Prerequisites

Matrix Algebra; Intermediate statistics.

### Recommended readings

- Husson, F., Lê, S., and Pagès, J. (2017). *Exploratory multivariate analysis by example using R*. CRC press.
- Borg, I., Groenen, P. J., and Mair, P. (2012). *Applied multidimensional scaling*. Springer Science & Business Media. **Available at:** <https://link.springer.com/book/10.1007%2F978-3-319-73471-2>
- Emmanuel Paradis, R for Beginners. **Available at:** [https://cran.r-project.org/doc/contrib/Paradis-rdebuts\\_en.pdf](https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf)