**COURSE**

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

**SSD: SECS-S/05**

PHD COURSE: Food Science

ACADEMIC YEAR: 2024/2025

**COURSE DESCRIPTION**

TEACHER: Prof. Alfonso Piscitelli

PHONE: +39 0812537450

EMAIL:  alfonso.piscitelli@unina.it

**GENERAL INFORMATION**

**SUGGESTED YEAR:**  II

**PERIOD**: from September to November

(indicate at least the month)

**CFU**: 4

**EXPECTED OUTCOMES**

The aim of the course is to provide Ph.D. students with the logic and statistical methodologies focused on analysis of high dimensional data trough 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).

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.

**CONTENTS/SYLLABUS**

Introduction to R (4h Laboratory); Principal Component Analysis and application (6h Lectures and Laboratory); Correspondence Analysis and Multiple Correspondence Analysis (6h lectures and Laboratory); Cluster Analysis (6h lectures and Laboratory); Multidimensional Scaling and Multidimensional Unfolding (6h lectures and Laboratory).

**PREREQUISITES**

Matrix Algebra; Intermediate statistics.

**READINGS/BIBLIOGRAFY**

- 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](https://link.springer.com/book/10.1007/978-3-319-73471-2)

- Emmanuel Paradis, R for Beginners. **Available at:** <https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf>

**TEACHING METHODS**

**Lectures hours: 14**

**Laboratory hours: 14**

**Seminars hours: 12**

**Other activities hours: 60**

**EVALUATION CRITERIA**

1. **Methods for acquiring eligibility**
* Written exam
* Oral exam
* Project discussion

X Other

1. **Evaluation pattern**

For the eligibility, a percentage of the attendance at the course of at least 90% is requested.