ADVANCED METHODS FOR AGRI-FOOD DATA ANALYSIS

Prof. Riccardo Negrini
Faculty of Faculty of Agriculture, Food and Environmental Sciences

Course Aims

The course is designed to build upon the foundational concepts introduced in the course "Introduction to Applied Statistics for Agri-Food Data"

The Advanced Course focuses on statistical methods that address complex questions involving multiple variables or groups. The primary aim is to empower students to leverage data and statistical techniques to make informed, evidence-based decisions. Rather than mathematical theory, the course emphasizes understanding the principles of statistics within the context of research and practical applications. Core statistical concepts will be explained using real-world examples and case studies to enhance comprehension and relevance.

The course introduces cutting-edge concepts in data mining and machine learning, situating them within a statistical framework.

This course is essential for those seeking to master advanced analytical tools and their application in the agri-food sector.

Methodology

The course employs a blended teaching methodology that combines frontal lectures, practical examples, and hands-on exercises to ensure a comprehensive learning experience. Theoretical concepts are introduced during lectures, providing students with a strong foundational understanding of advanced statistical methods. These concepts are then reinforced through real-world examples and case studies, illustrating their practical applications in agri-food research and business contexts. To solidify learning, students will engage in practical exercises using statistical software, allowing them to apply the methods to real datasets. This interactive approach ensures that students not only understand the theoretical framework but also develop the skills needed to independently analyze and interpret data in professional scenarios.

Course description

The course key topics include:

Quick refresh of basic concepts

Statistical Inference: Explore methods for comparing populations using both small and large samples. Learn the fundamentals of correlation, regression analysis, and statistical modeling, including ANOVA, Generalized Linear Models (GLM), and Generalized Linear Mixed Models (GLMM).

Non-Parametric Data Analysis: Delve into non-parametric tests such as the Sign test, Kruskal-Wallis test, Wilcoxon test, and rank correlation methods.

High-Dimensional Data and Statistical Learning: Gain an introduction to supervised, unsupervised, and reinforcement learning techniques. Learn to tackle classification problems using advanced tools such as Partial Least Squares Discriminant Analysis (PLS-DA) and Random Forests.

Recommended texts

The range of textbooks on statistical analysis and data mining is vast and varied, making it nearly impossible to propose a one-size-fits-all recommendation. The choice of resources should align with personal interests, toward a more theoretical, mathematics-based approach or a practical, application-oriented perspective. There are advanced and comprehensive texts as well as more introductory ones, depending on your level of expertise and learning goals. My suggestion is to select books that not only cover the statistical concepts but also provide practical guidance on using statistical software or programming tools (R, Python). Additionally, you may find value in specialized books that focus on specific topics (e.g. Multivariate analysis, Time series analysis, Regression and correlation, Machine learning).

Anyway, here a selection of text:

- James, Witten, Hastie, Tibshirani: An Introduction To Statistical Learning With Applications In R (Springer)
- Rasch, Verdooren, Pilz: Applied Statistics Theory and Problem Solutions with R (Wiley);
- Madsen: Statistics for Non-Statisticians (Springler)
- Dalgaard: Introductory statistics with R
- Cohen, Cohen: Statistics and Data with R - An Applied Approach Through Examples (Wiley)