**APPLIED STATISTICS AND BIG DATA ANALYTICS**

## Prof. Riccardo Negrini

***COURSE AIMS AND INTENDED LEARNING OUTCOMES***

This course aims at introducing the students to tolls and models for quantitative and qualitative data analysis. The course will reinforce students’ previous knowledge in statistics and expand a focus on multivariate statistics, generalised linear models, and analysis of chronological series. The most common Machine Learning approaches, their requirements, applications, and limitations, will be introduced along with some practical examples. The course will also provide the basic knowledge on the informatics and statistical tools involved in the analysis of high-dimensional data (Big Data).

***Knowledge and understanding***

At the end of the course, the student will be able to know and understand:

1. the main statistical methods and experimental design for qualitative, categorical and quantitative data

5. the outputs of the data elaboration, being able to describe and properly interpret the results obtained.

***Applying knowledge and understanding***

At the end of the course, the student will be able to:

1.Use informatics to extract, explore and manipulate information from low- and high-dimensional datasets.

3.Identify the most suited framework for a given hypothesis-testing scenario.

4.Build a coherent and functional analysis pipeline including basic bespoke functions when appropriate.

6.Complement the results with the most appropriate graphical representations.

***Critical thinking***

At the end of the course, the student will be able to:

1. recognize the research questions, critical revise information and data available, create meaningful experimental design and compare hypothesis

2.Formulate precise questions and identify the best statistical tool to test the hypotheses.

***Communication skills***

At the end of the course, the student will be able to:

1. mastering the technical language in interpreting and describing the statistical results

2. argument with adequate confidence the experimental design applied, and the statistical methods adopted.

***Learning skills***

At the end of the course, the student will be able to:

1. autonomously critically assess one own level of knowledge

2. identify adequate sources to deepen and improve their knowledge about statistical techniques applied to agri food research and suitable for qualitative and quantitative data

3. improve their coding ability autonomously exploiting E learning and other open-source training tools

## COURSE CONTENT

|  |  |
| --- | --- |
| **Topics** | **CFU** |
| Recap of exploratory statistics:* Describing data and variability.
* Probability distributions and descriptors (Binomial, Poisson, Chi Squared, Student T, Normal).
* Introduction to Hypothesis testing
* Confidence intervals
 | 1.00 |
| Basic concepts on probability and counting:* Probability rules and applications.
* Permutations and combinations
* Conditional probability
 | 1.00 |
| Introduction to statistical inference:* Comparing populations (small and large samples).
* Correlation and regressions
* Statistical models, ANOVA, GLM and GLMM.
 | 1.00 |
| Non parametric data and beyond:* Non-parametric test (Sign test, Kruskall-Wallis, Wilcoxon, rank correlation.
* Time series analysis
* Introduction to multivariate techniques (MDS, PCOOA)
 | 1.00 |
| High-dimensional data and statistical learning * Introduction on supervised, unsupervised and reinforcement learning.
* Classification problems: PLS-DA and Random Forest.
* Introduction on Neural Networks and deep-learning.
 | 1.00 |
| **Practical** |  |
| * Introduction to R programming.
* Data types and data editing.
* Algorithm thinking and main constructs.
* Plots, graphical representations, and reports in R.
* Introduction to the most common data analysis packages
 | 1.00 |

## READING LIST

Reading material, references, presentations, and code will be made available at the end of each section of the course.

## TEACHING METHOD

The teaching method will combine lectures and practical activities. Lectures will cover the theoretical topics of the course, while providing a set of real-life examples relevant to the professional careers. The practical activities will aim at reinforcing the theoretical information and convert them in practical skills.

Lectures will be delivered through frontal teaching. If the current health emergency prevents frontal teaching, remote teaching can be achieved through synchronous or asynchronous lectures.

## ASSESSMENT METHOD AND CRITERIA

At the end of the course the students will be graded following a written tests (total maximum score is 30/30 consisting in a combination of multiple-choice questions and exercises. Each multiple-choice question correctly answered (in full) will award 2 points, incorrect or partially-correct answers will be awarded proportionally less points. Exercises will be awarded from 2 to 6 points according to the difficulty (exercise value will be reported in the test).

The sum of quiz and exercise will produce the final score.

## NOTES AND PREREQUISITES

Knowledge of elementary statistics concepts will be helpful.

***OFFICE HOURS FOR STUDENTS***

Prof. Riccardo Negrini is available to meet students upon request at the Department of Animal Science, Food and Nutrition (DIANA).