

COURSE: Computational Statistics (SECS-S/01 – CFU: 5)

TEACHER: Pasquale Sarnacchiaro (pasquale.sarnacchiaro@unitelmasapienza.it)

TUTOR: Carlo Cavicchia (carlo.cavicchia@unitelmasapienza.it)

1. KNOWLEDGE AND SKILLS TO BE ACHIEVED DURING THE COURSE

Statistics is the study of the collection, analysis, interpretation, presentation and organisation of data. Statistical analysis and data analytics is listed as one of the highly desirable skills employers are looking for, and with data becoming an ever increasing part of modern life, the talent to extract information and value from complex data is scarce. The course is designed to train the next generation of statisticians with a focus on the field of data analytics. Employers expect skills in both statistics and computing. This course will provide a unique and coherent blend of modern statistical methods together with the associated computational skills that are essential for handling large quantities of unstructured data. This programme offers training in modern statistical methodology, computational statistics and data analysis from a wide variety of fields.

Statistics is the study of the collection, analysis, interpretation, presentation, and organization of data. The aim of the course is to produce graduates that:

- are equipped with a range of statistical methods and the associated computational skills for handling large quantities of unstructured data;
- have developed a critical awareness of the underlying needs of a wide variety of fields through relevant case studies;
- are able to analyse real-world data and to communicate the output of statistical models in order to inform decision making processes;
- have the necessary computational skills to build and analyse simple/appropriate solutions using statistical Big Data technologies.

Learning outcomes

By the end of the course the student gains an appreciation of the types of problems and questions arising with multivariate data. In particular the student should be able:

- to apply and interpret methods of dimension reduction including principal component analysis,
- to apply and interpret methods for cluster analysis
- to interpret the output of R procedures for multivariate statistics

Knowledge and understanding:

This course will give the students a working knowledge of statistical approaches for handling large quantities of structured and unstructured data.

Applying knowledge and understanding:

On successful completion students will be able to:

- Use statistical software;
- Use Statistical methods
- Report statistical results

Making judgements:

identify the best data/text mining techniques for an assigned task, identify the best data representation form.

Communication skills:

Present data and the results of statistical models in graphs, tables, and orally.

Learning skills:

Undertake basic statistical consultancy.

2. PROGRAM/ CONTENTS

The course will cover the modules listed below.
Quantitative Data Analysis
Research Methods and Case Studies
Methods for simple and multiple regression models, model fitting, variable selection, diagnostic tools
Principal component analysis
Applications of these methods will be illustrated with the statistical packages.

3. TEXT BOOKS

Teaching Materials:
Slides, on line lessons, web seminars, chat, forum.

4. EDUCATIONAL METHOD AND TOOLS

Educational method: On line lessons, web seminars, forum, Other activities available on Moodle.
Tools: R

5. SELF-ASSESSMENT PROCEDURES

On line test in Moodle environment.

6. EVALUATION METHODS (FINAL EXAM)

Questioning.

The practical skills are assessed via a written report of a data analysis project. Each student will be given a data set and corresponding scientific questions. Students will have two week to complete the data analysis and produce a written report following the instructions provided at the time of the exam. Collaboration on the data analysis and report writing is not allowed. Any collaboration will be deemed as cheating and result in a failing grade. Students also present their work orally, in fact each student is required to give a 15-minute oral presentation of the work in their dissertation, at a meeting of all Board of Examiners.

7. AREAS OF APPLICATION OF ACQUIRED KNOWLEDGE

Knowledge analysis and representation.