Description

Objectives

In this course, we aim to introduce methods that will enable students to study a "high-dimensional" dataset (here, "high-dimensional" is taken in the sense that we can't simply graph all the variables for all the observations) without resorting to a probabilistic model. The techniques we teach are used to reduce the dimension of the data, identify certain correlation between variables, visualize the data or divide the dataset into groups/classes.

Without neglecting the theory, the emphasis will be on the practical aspect of data analysis and the use of a programming language, whether R, Python or other.

Place of the course in the program

This course is generally taken by students in their second year of a bachelor's degree in statistics. It is also an elective in actuarial science, mathematics and some engineering and business programs.

All students should ensure that they have taken at least one linear algebra course (e.g. MAT-1200) and one basic statistics course (e.g. STT-1000), as most data analysis methods are based on these concepts.

Students should also have some familiarity with basic algorithmic concepts, as well as with (at least) one programming language.

Specific objectives

By the end of the course, students should be able to:

- understand and briefly describe the theoretical foundations of the data analysis methods studied;
- correctly identify situations where the use of these methods is indicated;
- use a programming language effectively to implement these methods;

- analyze and interpret the results of analysis;
- formulate the conclusions of the analysis in writing, within the limits of the methodology.

Computer equipment

You may need a computer, speakers or headphones, a microphone, a webcam and a wired broadband or wireless Internet connection. To check the minimum configuration parameters for your operating system, please visit this page.

In addition, this course may require specific software requirements, which will be described in other sections of the course outline as appropriate.

How it works

Classes and exams will take place face-to-face on Tuesday and Friday mornings, but materials will be available online. In general, the Friday session will be a lecture session, and the Tuesday session will be dedicated to practical exercises carried out independently by students with the support of the teacher.

Pedagogical approaches

The preferred teaching approach is interactive lectures (Fridays) alternating with laboratory periods (Tuesdays). The pedagogical approach is very much geared to active learning and requires sustained commitment from students throughout the session.