SYLLABUS DEL CORSO

Foundations of Probability and Statistics

2122-1-F9101Q002

Learning objectives

The course aims to introduce the concepts and methods of descriptive statistics, probability and statistical inference (estimation, tests, models) both from a theoretical and an application point of view through the use of software (R), with particular attention to the topics relevant for the most advanced datamining and machine learning courses.

At the end of the course the student has the opportunity to understand the statistical induction and the implications on the population deriving from the study of a data sample, being able to experiment and apply the knowledge acquired on real datasets.

Contents

Descriptive statistics, probability and statistical inference (estimation, tests, models)

Detailed program

- Introduction to data analysis with R
- Descriptive analysis: distributions, graphical representations, position and variability indices
- Probability: probabilistic conceptions, probability on events, Bayes theorem, random variables and probability distributions, large distributions, LLN and CLT statements
- Statistical inference: the logic of probabilistic sampling. Estimators and their properties. Point estimate (Average, variance and proportion). Notes on maximum likelihood estimators.

- Interval estimation: concept of confidence, confidence intervals, particular cases on the mean and variance
- Hypothesis testing: The concept of test statistics. The significance and power of the test. Test on average, variance, proportion, on the difference between averages, independence test.
- Simple linear regression: least squares estimation method, model adequacy measures, sampling distribution of OLS estimators, hypothesis tests and confidence intervals for the regression coefficients, analysis of variance, outliers and influential observations

Prerequisites

None.

Teaching methods

Lectures and computer lab.

Assessment methods

WRITTEN EXAM: it is a multiple choice and open question exam about the theoretical issues proposed in class. (max grade 28)

PROJECT WORK: (in group - max 3 students - or individually) involving a data analysis (with R) on a dataset chosen by the student to replicate arguments and analyses discussed during lab sessions. The project must be send at least 10 days before the final test (written exam) to the teacher (max grade 6)

There is no mid term evaluation

Textbooks and Reading Materials

Teacher material available for the students on e-learning site

Text of your choice:

- A.M. Mood, F.A. Graybill, D.C. Boes, Introduzione alla statistica
- G. Cicchitelli, P. D'Urso M. Minozzo, Statistica: principi e metodi
- P.S. Mann, Introductory Statistics
- M. Lavine, Introduction to Statistical Thought

Semester

I semester (September-November)

Teaching language

Italian