UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Probabilità e Statistica per l'Informatica (blended)

2021-2-E3101Q127

Aims

To provide core skills in the fields of **probability theory** and (**descriptive and inferential**) **mathematical statistics**. At the end of the course, students will be able:

- to understand language, definitions and statements of some of the core results in probability theory and mathematical statistics
- to take adavantage of the concepts learned in the course to perform autonomous reasoning under uncertainty
- to analyze and describe data sets

- to estimate parameters for models describing random phenomena and to perfom hypothesis testing
- to implement the above mentioned points in the computer using the R programming language

Contents

The first part of the course, devoted to **probability theory**, starts with an introduction to the so-called **descriptive statistics** which allows to summarize the salient features of some data set. Subsequently we present **probability spaces**, a key concept in the modeling of random phenomena, along with **random variables** which form its "operating language". Finally, we focus on **limit theorems** (law of large numbers and central limit theorem) which describe the behavior of random phenomena with a large number of components.

The second part of the course, devoted to **mathematical statistics**, starts with a presentation of the main **parameter estimation** techniques for probabilistic models. Then we discuss **hypothesis testing** and the corresponding statistical tests, concerning both unknown parameters in a random model (*parametric tests*) and the comparison between distributions (*non-parametric tests*). The last part of the course is devoted to the **linear regression**, a powerful technique to investigate the link that might exist between some variable (output) and a

given set of variables (input) under uncertainty.

Detailed program

0. Descriptive Statistics

- Introduction to data analysis
- Sample statistics (mean, median, quantiles, variance, correlation)
- Graphical representations

1. Probability spaces

- Random phenomena, probability spaces and events
- Elements of combinatorial calculus
- Independence of events

2. Random variables

- Random variables and distributions
- Expectation, moments, variance and covariance
- Important discrete random variables (Bernoulli, Binomial, Geometric, Poisson)
- Important continuous random variables (Uniform, Esponential, Gamma, Normal)

3. Convergence Theorems

- Convergence of random variables and distribuitons (hints)
- Law of Large Numbers
- Central Limit Theorem

4. Parameters Estimation

- Samples and statistics
- Estimators (sample mean and variance)
- Confidence intervals

5. Hypothesis Testing

- Hypothesis testing, I type and II type errors
- Parametric hypothesis testing for mean and variance
- Non parametric hypothesis testing for goodness of fit and independence

6. Linear Regression

- Introduction to linear regression
- Statistical inference for parameters
- Residual analysis

Prerequisites

The knowledge, competences and skills taught in previous mathematical courses, in particular concerning Mathematical Analysis.

Teaching form

Lectures divided into:

- theoretical lectures, focused on the knoledge of definitions, results, proofs and relevant examples;
- recitations, focused on the skills necessary to apply the theoretical knoledge and competencies to the solution of exercises.

If the Covid-19 emergency will allow for it, lectures will be given in the classroom. Otherwise they will be given in a mixed mode (partially in the classroom, partially on-line) or entirely on-line, according to the safety rules that we will receive. All lectures will be recorded and made available on the e-learning website. In order to help students take an active part in the course, on-line lectures will be integrated with discussion events that will be held in real time.

The course is given in *blended-learning* in Italian.

Textbook and teaching resource

Reference book:

- S. M. Ross (2014), Introduzione alla Statistica, seconda edicione, Apogeo Editore

Moreover, the following teaching material is made avaliable:

- Videorecording of the lectures (theory and exercise classes)
- Theacher's notes
- Exercise sheets

Semester

Spring term (Second semester)

Assessment method

The exam consists of a written part and of a possible oral part. It receives a mark out of 30.

The written part consists of two parts:

- a first part of closed-ended questions, which contributes one third to the final mark, where theoretical skills are tested;
- a second part of open questions, which contributes two thirds to the final mark, where practical skills are tested.

Two midterms will be held, halfway and at the end of the course, with the same structure as the written part. Passing both midterms is equivalent to passing the written part with the mean of the marks.

The oral part is optional (or at the request of the teacher) and can increase or decrease the final mark.

The examinations will be held in presence if the Covid-19 emergency will make it possible, otherwise they will be held on-line, according the the safetly rules that we will receive.

Office hours

By appointment