UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Probability and Statistics For Computer Science (blended)

1920-2-E3101Q127

Aims

To develop basic skills on **descriptive statistics**, **inferential statistics** and **probability calculus**.
The student will be able to

- analyze and summarize quantitative data, uni-variate and multi-variate
- reason under uncertainty
- apply parametric distributions to describe real world phenomena
- compute and use point and interval estimates
- apply hypothesis testing to make decisions about parameters and distriutions
- discover the link between independent variables and a target variable
- develop computer programs using the R open source programming language

Contents

The course introduces basics about **descriptive statistics**, uni-variate, bi-variate and multi-variate. We'll present basic concepts of **probability calculus:** continuous and discrete random variables, uni-variate, bi-variate and multi-variate. The main discrete and continuous **probability distributions** will be described. The **Central Limit Theorem** together with the **Law of Large Numbers** will be presented.. The course provides the basics of **point**

and interval estimates, as well as the basics of **Parametric and Non Parametric Hypothesis Testing**. Finally, **Linear Regression Analysis** is introduced to develop simple linear regression models and multi-variate linear regression models. The course will introduce the **R programming language** with specific reference to those packages related with the course topics.

Detailed program

1. Descriptive Statistics

1.1. Graphical and numerical representation of data
1.2. central and variability tendency indices
1.3. Bi-dimensional representation

2. Probability Calculus

2.1. Sample space, probability and partition function
2.2. Conditional probability and stochastic independence
2.3. Chain rule and Bayes theorem
2.4. Continuous and discrete random variables
2.5. Uni-dimensional and multi-dimensional random variables
2.6 Central tendency and variability indices

3. Parametric Probability Distributions

3.1. Discrete: Bernoulli, Binomial, Poisson, Geometric.
3.2. Continuous: uniform, triangular, normal, beta, exponential, Student t, F, Chi-squared

4. Convergence Theorems

4.1. Convergence in distribution
4.2. Law of Large Numbers
4.3. Central Limit Theorem

5. Parameters Estimation

5.1. Samples and sampling

5.2. Main sampling distributions
5.3. Point and interval estimates
5.4. Interval estimates for mean and variance

6. Parametric Hypothesis Testing

6.1. Introduction to hypothesis testing
6.2 Type I and II errors
6.3. Hypothesis on mean value and variance for a single population
6.4. Hypothesis for the difference in mean and in variance for two populations

7. Non Parametric Hypothesis Testing

7.1. Goodess of Fit tests;: Kolmogorov-Smirnov and Chi-square
7.2. Comparing distribution of two populations; sign test, rank test
7.3. Independence test

8. Linear Regression

8.1. Introduction to linear regression analysis
8.2. Estimation of constants, confidence intervals for individuals
8.3. Model fitting, Residuals analysis
8.4. Multiple linear regression, fitting, and variable's relevance

Prerequisites

Mathematical Analysis, Programming skills.

Teaching form

The course runs in **blended-learning mode**. and is *teached in Italian*. Each chapter consists of **learning modules**.

A **learning module consists of** the following sessions:

- **introduction and discussion** of methodology and/or theoretical contents
- **exercises** paired to methodology and/or theoretical contents

- **R programming language** design and development laboratory
- **Self-evaluation** with automatic **feedback from the teacher**

The course **stimulates teacher - online tutor - student interaction** by using forum, two for each chapter. All the course material is available before the first class and thus the student is encouraged to come to classes after having watched video-lectures and have practiced with self-evaluation modules. It is worthwhile to mention that about 70% of the course topics are covered by **video lectures designed and recorded by the teacher**. Therefore, the student can make the decision to come to classes or not.

Textbook and teaching resource

The textbook is **Franco Pellerey (2007)**. *Elementi di statistica per le applicazioni. con esercizi*, CELID.
Furthermore, the following teaching material from the teacher is made available;

- Video-lectures from the teacher
- Slides from the teacher
- Numerical exercises fully explained and commented, slides and video-lectures.
- Contextual forum, two for each chapter.
- Self-evaluation quizzes, Simulation of exams.

Semester

Spring Semester

Assessment method

- **Mid-term**; the student attending classes and/or studying during the running of the course is offered to undergo two mid-term exams. The first mid-term exam is about chapters 1, 2, 3 and 4 contents while the second mid-term exam is about the remaining chapters, 5, 6, 7, and 8. Each mid-term exam consists of **10 quizzes** (*each quiz is about notions presented in the course and gives 1 point*), **2 numerical exercises** (*each exercise requires to develop a solution procedure to a given problem, each exercise gives a maximum of 9 points*), **1 open ended question** (*the question is about the notions presented in the course and gives a maximum of 5 points*) and optionally the student can ask for solving an **R language exercise** (*commenting a given R language code, it gives a maximum of 3 points*) and/or to undergo the **oral examination** (*oral examination is about reasoning and deduction and gives a maximum of 2 points*). The signed grade is the rounded up arithmetic mean of the grades of the two mid-term exams.

- **Standard**; the exam is about all chapters of the course, from 1 to 8. The exam consists of **10 quizzes** (*each quiz is about notions presented in the course and gives 1 point*), **3 numerical exercises** (*each exercise requires to develop a solution procedure to a given problem, each exercise gives a maximum of 6 points*), **2 open ended questions** (*each question is about notions presented in the course and gives a maximum of 2 points*) and optionally the student can ask for solving an **R language exercise** (*commenting a given R language code which gives a maximum of 1 point*) and/or to undergo the **oral examination** (*oral examination is about reasoning and deduction and gives a maximum of 2 points*)

Office hours

On dating