# Statistics

## Professor Enrico Fabrizi

***COURSE AIMS AND INTENDED LEARNING OUTCOMES***

The course introduces the students to the basic concepts and methods of statistical data analysis. The aim is to provide essential understanding of statistical thinking, its role in scientific research and business decisions and to provide the students tools to conduct simple data analyses and to understand content and limitations of analyses conducted by others. The course provides a preparation for more advanced statistical techniques students will face during the rest of their degree course.

In terms of skills, at the end of the course the students should be able to:

-apply fundamental concepts in exploratory data analysis;

- design studies for obtaining data whilst avoiding common design flaws that incur bias, inefficiency and confounding;

- demonstrate a basic understanding of probability and random variables;

- understand the concept of the sampling distribution of a statistic as well as the foundations of classical inference related to the concepts of confidence intervals and hypothesis testing;

- apply inferential methods relating to the means of Normal distribution

- apply exploratory and modelling techniques to the analysis of bi-variate data and use inferential methods in the context of simple linear models;

***COURSE CONTENT***

* Introduction to statistical methodology and its role in the social sciences. Computers and software in statistics.
* Variables and their measurement. The principle of randomization in sampling.
* Tables and graphs in the description of data. Means and variability measures.
* Bivariate descriptive statistics: two way tables, association, correlation.
* Sample statistics and population parameters. A concise introduction to probability calculus. Probability distributions. The Normal probability distribution. Sampling distributions.
* Point and interval estimation. Confidence intervals for means and proportions. Choice of sample size.
* Statistical significance tests for means and proportions. Decisions and error types in tests. Limitations of testing procedures. Comparison of two groups: proportions and means. Non parametric hypothesis testing for comparing two groups. The one-way ANOVA for comparing more than two groups.
* Analysis of associations between variables. Chi-square test of independence. Other measures of association in contingency tables.
* Linear relationships. The linear regression model. Inferences for the linear regression model parameters and the correlation coefficient. Model assumptions and violations.
* A short introduction to the software BlueSky for statistical analysis

***READING LIST***

Reference text

##### Alan Agresti: Statistical Methods for the Social Sciences, 5th Edition

Complementary notes, lesson handouts, exercises will be released by the teacher during the course and published online on the course blackboard page.

***TEACHING METHOD***

The course is based on lectures. Most of them will be theoretical lectures, with a minority dedicated to an introduction to softwares for statistical data analysis.

The lectures will be integrated with exercises development and case examples. Practice exercises taken from previous exams are suggested to students every second week (approximately), and solutions distributed via Blackboard***.***

The course includes one tutorial hour every week focused on the introduction to the BlueSky software. Tutorial lecturers are held by a tutor.

The course assumes full attendance to class and active participation in class. Should you have problems in attending, please contact the instructor.

***ASSESSMENT METHOD AND CRITERIA***

The exam includes a written test (mandatory) and an oral interview (optional).

The written exam consists of practical exercises and theoretical questions. By solving the exercises the students should demonstrate their ability to apply the techniques of analysis treated in the course to small data sets. The theoretical questions are designed to test their ability to use concepts to solve simple problems of data analysis and to interpret results obtained using statistical softwares.

The written exam is made of 16 questions / exercises, each evaluated from 0 up to 2 points. If the total exceeds 30, “cum laude” is added to the maximum mark.

The oral test verifies that students have understood the basics of statistical thinking, are able to illustrate their economic and business applications of statistics and are able to work with basic concepts of mathematical statistics. Only those students who have achieved a positive result in the written test (at least 18/30) are admitted to oral interview; the grade obtained in the oral test may change the grade obtained in the written test of at most 4 points (in both directions).

According to the decisions taken in this regard by the faculty, the written exam can be passed by getting a positive result in two written test: a first mid-term test in the (unique) date approved for this purpose by the faculty, and a second test in exam sessions immediately following the end of the teaching period.

The two partial test are made of 16 questions / exercises, each evaluated from 0 up to 2 points.

The average marks obtained in intermediate examinations defines the written test grade. If the total exceeds 30, “cum laude” is added to the maximum mark.

***NOTES AND PREREQUISITES***

No previous exposure to statistics is required. It is assumed that students are with basic mathematical concepts and notation to the level characterizing the Mathematics course included in the first year of their degree program.

***OFFICE HOURS***

Information on office hours available on the teacher's personal page at <http://docenti.unicatt.it/>. The teacher’s office in room 386, third floor, DISES building. During office hours, the students can talk to the teacher using 8Microsoft Teams.