# Statistics

## Prof. Alessandro Recla

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

The knowledge of statistics is becoming more and more important to analyze, understand and make more informed decisions.

In the last decade there has been an unprecedented revolution in the data collection and data accessibility of all kinds: it is estimated that 90% of the data present today has been created in the last two years.

The course aims to provide the first theoretical and applied tools to carry out a rigorous statistical analysis of a set of data. In particular, the student learns how useful information can be extracted and at the same time evaluated.

Students will acquire a good understanding of the statistical tools and techniques related to descriptive statistics and probability covered in the course, as well as the ability to employ them in economic and finance applications.

At the end of the course students will be able to:

* Understand the fundamentals of statistical thinking;
* Reproduce the basics of descriptive statistics to finance data analysis;
* Summarize and visualize information contained in real data sets;
* Study the relationship between relevant variables;
* Choose adequate probabilistic models to represent data and learn from it in a statistical setting.

***COURSE CONTENT***

The detailed contents of the course are divided into two main parts: Descriptive statistics (I part) and Probability and inference (II part).

Descriptive statistics (I part):

* Introduction to the course content, teaching materials.
* Why statistics? Decision making under uncertainty.
* Steps in a statistical analysis: goals, overview of sources, data collection, analysis and presentation.
* Basic notions of descriptive statistics.
* Population and sample. Descriptive and inferential statistics.
* Types of variables and levels of measurement (qualitative and quantitative data).
* Frequency distribution and graphs to describe categorical variables (bar chart, pie chart).
* Frequency distribution and graphs to describe discrete numerical variables (stick chart).
* Line chart to describe time series data.
* Frequency distribution and graphs to describe continuous numerical variables: histogram with equal and unequal class widths. Frequency density.
* Cumulative frequencies and their graphical representation.
* Relationships between variables: scatterplot and two-entry table. Conditional distributions and statistical independence. Simpson's paradox.
* Stacked-bar chart (and side-by-side bar chart) to describe subpopulations.
* Measures of central tendency: algebraic and positional averages. Arithmetic and geometric average. Median. Mode.
* Quartiles and quantiles.
* The Boxplot and the shape of the distribution. Mean vs median comparison: robustness.
* Measures of variability: range, interquartile range, variance, standard deviation.
* Alternative formula to compute the variance. Coefficient of variation.
* Measures of relationships between variables: covariance and linear correlation coefficient. Analysis of variance.
* OLS regression.
* Independence.

Probability and inference (II part):

* Definition of probability.
* Random experiments, conditional probability.
* Events and random variables.
* Discrete random variables.
* Parametric probability distributions of discrete random variables: Bernoulli Binomial.
* Continuous random variables.
* Parametric probability distributions of continuous random variables: Normal, Standard Normal.
* The use of tables of the Normal distribution.
* Introduction to Statistical Inference: population, sample, statistics and parameters.
* The sample mean and its properties: in particular, expected value and variance.
* Sampling distribution of the sample mean: normal and arbitrary populations (Central Limit Theorem).
* Bernoulli population: application of the Central Limit Theorem to the sample proportion.
* Introduction to confidence intervals and hypothesis testing.

***READING LIST***

Textbook: P. Newbold, W.L.Carlson, B. Thorne (2020). Statistics for Business and Economics, Pearson. (Ninth Edition – Global Edition).

***TEACHING METHOD***

The course is structured as follows:

70 hours of lectures which cover all topics of the syllabus. These lectures cover theory, examples and the analysis of a “real” data-set (as indicated below); data analysis will be done using the software Microsoft Excel, introduced during the lectures.

***ASSESSMENT METHOD AND CRITERIA***

The exam consists of a written general exam. Alternatively, the general written exam can be replaced by two partial exams.

*Partial exams*

The first partial exam and the second partial exam are traditional written exams with the possibility to have Excel outputs to be interpreted and commented correctly. The first partial exam covers the first sessions of the syllabus (please refer to the detailed program of the course above). The second partial exam covers mainly the remaining topics.

Each of the two written partial exams is considered passed only if each of their scores is greater than or equal to 18/30.

If a passing grade is obtained on both written partial tests, the final grade is determined by the average of the two scores not rounded. The final score will be then rounded.

The exam is considered passed only if the final grade is greater than or equal to 18/30. The score 31/30 indicates 30/30 “cum laude.”

*Written general exam*

The general written exam consists of exercises, theory questions, and optionally questions on Excel.

The general written exam has a maximum grade of 31/30.

The general written exam is passed only if the score is greater than or equal to 18/30 (on the usual scale with 31/30 as maximum).

***NOTES AND PREREQUISITES***

This course considers as a prerequisite topics of mathematics developed in the course Mathematics.

If the restrictions due to the Covid-19 pandemic should persist, distance learning will be guaranteed, with modalities that will be communicated in advance to students.

*Office hours: time and place*

A weekly office hour will be held during the course.