# Advanced statistics

## Prof. Stefano Bertini

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

The course completes the training on statistical methods, presenting the topic of hypothesis estimation and testing and the classic procedures regarding the mean, variance and proportion considering both one and two populations.

After these cases we will move on to the multiple regression model examined under the condition of a homoskedastic error component and not correlated with the explanatory variables. The ANOVA test on the multiple linear model and the main methods of variable selection will then be presented.

At the end of the lessons the student is expected to be able to:

– determine the estimator of a parameter obtained with the

maximum likelihood method;

– calculate confidence intervals and conduct hypothesis tests on a population for the mean and variance of a normal distribution and the proportion of successes in a Bernoulli distribution;

– compare the mean of a character in two populations by applying the student’s t-test in its different versions;

– obtain the estimates of the ordinary least squares for a multiple regression model and calculate the ANOVA test connected to it;

– know the main variable selection mechanisms such as Akaike’s information criterion or stepwise procedures;

– correctly interpret the results of an automatic processing program dedicated to homoskedastic linear regression models.

***COURSE CONTENT***

– Summary of probability

– Random variables and expected value operator. The random variable of Gauss.

– The problems of sampling from a population and the distribution of the sample mean of a character.

– Point and interval estimation - Consistency and non-distortion of an estimator.

– The log-likelihood function and the maximum likelihood estimates.

– Notes on the method of moments.

– The confidence intervals. Hypothesis testing.

– Hypothesis systems and decision-making rules.

– The chi-squared test for contingency tables.

– The significance tests on the mean, variance and proportion. The observed significance level (p - value).

– Student’s t-tests for comparing the means of two populations in the case of normal or Bernoulli distribution.

– Stochastic linear model - The sample distribution of the model parameters.

– Gauss - Markov assumptions and maximum likelihood estimates.

– Confidence intervals and hypothesis testing on individual coefficients.

– The ANOVA test on the overall fit of the model.

– Variable selection methods: Akaike’s information criterion and hints to stepwise procedures.

***READING LIST[[1]](#footnote-1)***

Texts (more details will be given at the beginning of the course)

S. Ross, Introduction to statistics, Apogeo educational - Maggioli editore.

Complementary texts

M. Pelosi-T. Sandifer, Introduction to Statistics, McGraw - Hill Educational.

B.V. Frosini, Regression Analysis, with Appendix on Vectors and Matrices, EDUCatt, Milan.

***TEACHING METHOD***

Lectures and exercises.

***ASSESSMENT METHOD AND CRITERIA***

Written test divided into four sections of exercises only or three exercises and a theoretical question. The topics covered by the exercise or theoretical question are the following:

- the search for estimators obtained with the maximum likelihood method and the calculation of estimates by interval;

- the hypothesis test concerning the parameters of the normal population and Bernoulli (with calculation of the p-value) and the chi-squared test on the contingency tables;

- the student t test for the comparison between the means of two populations;

- the estimation of a multiple regression model, the ANOVA test and the methods of selection of the variables.

Each section contributes equally to the final evaluation.

***NOTES AND PREREQUISITES***

The contents learned in the courses of Statistics I and General Mathematics are to be considered prerequisites.

1. I testi indicati nella bibliografia sono acquistabili presso le librerie di Ateneo; è possibile acquistarli anche presso altri rivenditori. [↑](#footnote-ref-1)