UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Advanced Statistics

2425-1-F7702M034-F7702M117M

Learning objectives

The course aims at introducing the main concepts of statistical reasoning. When analysing a dataset, students will be able to formalise a statistical problem and to identify probabilistic tools and suitable statistical methods for the data. The focus of the module will be on the rationale underlying basic statistical methods so that students might develop an autonomous critical approach and be able to understand and apply statistical techniques, even if not covered by the syllabus of this module.

Contents

- Using statistics to summarise datasets
- Essential review of probability theory
- Distribution of sampling statistics
- Estimation
- Testing statistical hypothesis
- Linear regression
- Data analysis with Stata

Detailed program

Using statistics to summarise datasets

- Populations and samples
- Sample mean
- Linearity of sample mean

- Deviations
- Sample median
- Sample percentile
- Sample quartiles
- Sample mode
- Sample variance and standard deviation
- Property of sample variance
- Sample correlation coefficient and interpretation

Essential review of probability theory

- random experiment, sample space, events
- defining properties of probability
- frequentist interpretation of probability
- probability of the complement and addition rule
- equally likely outcomes
- independent events
- random variables (RV)
- discrete RVs
- distribution of a RV
- expected value of a discrete RV
- properties of the expectation
- variance of a discrete RV (equivalent definitions)
- Bernoulli RV
- Binomail distribution, with derivation
- Factiorial numbers and binomail coefficient
- continuous RVs
- probability density function of X and P(a<X<b)
- uniform random variable
- normal random variable
- standard normal random variable
- computing normal probabilities
- additivity of normal RVs

Distribution of sampling statistics

- sample from a distribution
- sample mean as a RV
- expected value and variance of sample mean
- simulation study with binomial sample
- approximate areas under the normal curve
- central limit theorem
- normal approximation for the sample mean (for large n)

Estimation

- estimator and estimate
- unbiased estimator of a parameter
- point estimator of population mean
- margins of error via normal approximation
- point estimator of the population variance
- sampling proportions from a finite population
- random sample of size n from a population of size N
- approximate Binomial distribution for the sample when N is much larger than n
- sample proportion as estimator for population proportion

- interval estimator
- point and interval estimate for the sample proportion

Testing statistical hypothesis

- introduction to hypothesis testing
- null and alternative hypothesis
- test statistic
- critical region
- type I and type II errors
- level of significance of a test
- two-sided and one-sided tests
- p-value
- one-sample Z-test (derivation from the distribution of the test statistic)
- T-test

Linear regression

- simple linear regression
- independent and dependent variables
- least square estimates for the parameters of a linear regression model
- estimated regression line
- prediction via the estimated regression line
- confidence intervals for the predicted values
- coefficient of determination

Data analysis with Stata

- Interface of Stata
- Menu system, command windows and Do-files
- Importing and editing datasets
- computing probabilities with Stata
- examples with binomial and normal random variables
- Z-test: practical examples with Stata
- T-test: practical examples with Stata
- test for proportions: practical examples with Stata
- simple linear regression with Stata
- introduction to residual analysis
- multiple linear regression with Stata

Prerequisites

Basic calculus and descriptive statistics

Teaching methods

The course will consist of a total of 35 classes, of which 27 regular classes and 8 lab sessions (with Stata)

Assessment methods

The exam is written and consists of questions about theory and exercises. The former verify students' knowledge and understanding of the main concepts of the subject; the latter measure students' ability in the application of such concepts to solve simple practical problems.

Textbooks and Reading Materials

Textbook: Sheldon Ross, Introductory Statistics (4th edition)

Other material is provided during the course

Semester

First semester (second cycle)

Teaching language

English

Sustainable Development Goals