UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Statistica per la Sicurezza e Analisi dei Dati

2627-1-F8805N030

Learning objectives

The course adopts a practical and application-oriented approach and aims to provide students with the fundamental knowledge and skills required for the statistical analysis of security-related data, from data acquisition to predictive analytics.

Knowledge and Understanding
Upon successful completion of the course, students will have acquired:

- knowledge of the fundamental principles of descriptive, inferential and predictive statistics applied to the analysis of security-related data;
- an understanding of the main methodologies for collecting, organizing and analyzing data from official sources;
- an understanding of the assumptions, strengths and limitations of the statistical methods and tools employed.

Applying Knowledge and Understanding
Upon successful completion of the course, students will be able to:

- acquire, organize and manage datasets obtained from official sources;
- use the R programming language and the RStudio environment for data manipulation and analysis;
- describe and summarize data through statistical indicators and graphical representations;
- apply inferential methods to the study of security-related phenomena;
- build and interpret regression models and machine learning models;
- apply appropriate statistical procedures to the analysis of real-world data.

Making Judgements
Upon successful completion of the course, students will be able to:

- critically assess the quality and reliability of security-related data;

- select the most appropriate statistical tools according to specific problems;
- critically interpret the results obtained from statistical analyses.

Communication Skills
Upon successful completion of the course, students will be able to:

- present and communicate the results of statistical analyses through tables, graphs and summary indicators;
- clearly describe the methodologies adopted and the results obtained using appropriate statistical terminology.

Learning Skills
Upon successful completion of the course, students will have developed the ability to:

- identify, evaluate and critically use new data related to security phenomena;
- independently address new analytical problems by applying the knowledge and skills acquired during the course;
- update and further develop their competencies in data analysis.

Contents

The course aims to provide students with the fundamental knowledge and skills required for statistical analysis in the field of security. The approach combines theoretical concepts and practical applications through the analysis of real-world security-related datasets, supported by the R statistical software and the RStudio environment.
The course is organized into four main topics.
The first topic introduces the fundamental concepts of statistics applied to security, with particular attention to data sources (such as ISTAT and the Italian Ministry of Justice), data quality and data structure. The use of R for data import and management is also introduced.
The second topic focuses on exploratory data analysis and covers key concepts such as frequency distributions, summary measures, graphical representations and the analysis of relationships between variables.
The third topic addresses statistical inference and the main tools of probabilistic and inferential analysis. Topics include random variables, probability distributions, point and interval estimation, hypothesis testing, group comparisons and analysis of variance.
The fourth topic is devoted to predictive analytics, introducing linear and logistic regression as well as the main machine learning methods.
All techniques and tools presented throughout the course are implemented using R and RStudio, with particular emphasis on the critical interpretation of analytical results.

Detailed program

Topic 1 Introduction to statistics for security and the R environment

- Introduction to statistics for security: objectives and tools
- Objective and subjective security; official data sources
- Populations, samples and statistical units
- Data quality: errors and accuracy
- Introduction to the R statistical software and the RStudio environment

Topic 2 Exploratory data analysis

- Frequency distributions
- Measures of central tendency and dispersion
- Contingency tables
- Relationships between pairs of variables
- Graphical representations

Topic 3 Statistical inference

- Discrete and continuous random variables
- Probability distributions
- Point and interval estimation
- One-sample statistical tests
- Two-sample statistical tests
- Group comparisons and analysis of variance (ANOVA)

Topic 4 Predictive analysis

- Simple and multiple linear regression
- Logistic regression
- Introduction to machine learning

Prerequisites

Basic knowledge of mathematics and numerical computation.

Teaching methods

The course consists of 56 hours of in-person teaching, delivered primarily through lecture-based (erogative) teaching (DE). Classes combine the presentation of theoretical concepts through slides with practical activities carried out using R and RStudio.
Interactive teaching (DI) is integrated into the lectures and is designed to encourage active student participation through the real-time application of the procedures presented during class using personal devices. Interactive teaching activities mainly consist of guided exercises in R, discussion of results, and the solution of practical analytical problems.
Students will also be able to independently replicate the analyses presented during the course using the code provided by the instructor.
The course e-learning platform (https://elearning.unimib.it) is regularly updated with teaching materials, including lecture slides, R scripts, datasets used during the course, and additional learning resources.

Assessment methods

Student learning is assessed through a written exam designed to verify the achievement of the course learning outcomes.
The exam consists of multiple-choice questions and open questions requiring students to report the R code used to identify the correct answer and to provide a critical interpretation of the results obtained.
Further information will be provided during the course and will be available on the course e-learning page.

Assessment will be based on the following criteria:

- completeness and accuracy of the knowledge acquired;
- ability to use R and RStudio for the analysis of security-related data;
- ability to critically interpret the results of statistical analyses;
- ability to apply appropriate statistical tools to the solution of specific problems;
- ability to synthesize information effectively;
- appropriate use of statistical language and terminology.

Textbooks and Reading Materials

Lecture slides, handouts, and additional teaching materials will be available on the course e-learning page.

Suggested references:

- Mecatti, F. (2022). Statistica di Base. Come, Quando e Perché (3rd ed.). McGraw Hill.
- Wooditch, A., Johnson, N. J., Solymosi, R., Medina Ariza, J., & Langton, S. (2021). A Beginner's Guide to Statistics for Criminology and Criminal Justice Using R. Springer International Publishing.
- Weisburd, D., Wilson, D. B., Wooditch, A., & Britt, C. (2021). Advanced Statistics in Criminology and Criminal Justice. Springer.

Sustainable Development Goals

REDUCED INEQUALITIES | SUSTAINABLE CITIES AND COMMUNITIES | PEACE, JUSTICE AND STRONG INSTITUTIONS