STATISTICAL LABORATORY
3 CFU - 1° semestre

Docente titolare dell’insegnamento

ALESSANDRO ORTIS
Email: ortis@dmi.unict.it
Edificio / Indirizzo: Dipartimento di Matematica e Informatica, I blocco
Telefono: 3426454328
Orario ricevimento: 10-12

OBIETTIVI FORMATIVI
AIMS AND SCOPE

The aim of the course is introduce the knowledge of the R language for statistical data analysis with special focus on descriptive statistics, probability distributions and statistical inference.

LEARNING OBJECTIVES

1. Knowledge and understanding (Conoscenza e capacità di comprensione). The objectives aim at introducing the knowledge of the R language for statistical data analysis with special focus on descriptive statistics, probability distributions and statistical inference.

2. Applying knowledge and understanding (Capacità di applicare conoscenza e comprensione). On completion. Students will be able to utilize the R language for: i) providing basic statistical analyses of data; ii) simulating data according to given probability distributions; iii) applying main methods of statistical inference.

3. Making judgements (Autonomia di giudizio). On completion, students will able to extract knowledge from data through statistical analyses in R.

4. Communication skills (Abilità comunicative). On completion, students will be able how to present the results from the statistical analyses, based on the use of the statistical software R.

5. Learning skills (Capacità di apprendimento). On completion, students will able how to utilize the statistical software R for basic data analysis and modeling.

MODALITÀ DI SVOLGIMENTO DELL’INSEGNAMENTO

Lectures and practical activities and data analysis in R.
PREREQUISITI RICHIESTI
Basics of linear algebra and statistics.

FREQUENZA LEZIONI
Mandatory.

CONTENUTI DEL CORSO
Use of the statistical software in R regarding:


Probability. Random number generation and data modeling according to different probability distributions: uniform, binomial, Poisson, Gaussian.


TESTI DI RIFERIMENTO

ALTRO MATERIALE DIDATTICO
See https://www.dmi.unict.it/ortis/StatsLab/

PROGRAMMAZIONE DEL CORSO

<table>
<thead>
<tr>
<th>Argomenti</th>
<th>Riferimenti testi</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Introduction to R, Basic Commands in R, Indexing Data, Matrices and Lists, Loading Data</td>
<td>Lecture Notes</td>
</tr>
<tr>
<td></td>
<td>Graphs, Data Types and Structures, Conditional Statements and Loops, Graphs and Data Visualization</td>
</tr>
<tr>
<td>---</td>
<td>------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>3</td>
<td>Mean, Median, Variance, standard deviation, quantiles, percentiles, interquartile distance, boxplot, outlier detection</td>
</tr>
<tr>
<td>4</td>
<td>Functions in R, data filtering</td>
</tr>
<tr>
<td>5</td>
<td>Bivariate analysis, statistical inference, contingency table, joint probability, marginal probability, chi-squared test, t-test, linear regression.</td>
</tr>
</tbody>
</table>

**VERIFICA DELL'APPRENDIMENTO**

**MODALITÀ DI VERIFICA DELL'APPRENDIMENTO**

Practical activity and data analysis with R

**ESEMPI DI DOMANDE E/O ESERCIZI FREQUENTI**

Perform a univariate analysis considering the attribute X

Report the correlation matrix among the attributes considering 2 digits precision

Perform a Linear Regression analysis of the relationship between the two features X and Y with the variable Z. Report below the output of the summary() function applied on the linear regression model obtained using lm(). Then, comment the results.

Is the dataset balanced with respect to the attribute X?

Visualize the scatter plot considering the variables X and Y. Report below the code used to create the plot.