

LABORATORIO DI INTRODUZIONE ALLA DATA SCIENCE

Cosa serve:

R-project → per scaricare l'ultima versione potete usare questo link (<https://www.r-project.org/>). Cliccate su "download R" e scegliete il mirror più vicino a voi (per l'Italia c'è l'università di Padova e quella di Milano)

RStudio → andate su questo link (<https://www.rstudio.com/>) e scaricate la versione free di RStudio Desktop

NOTA: per gli utenti Windows è utile scaricare ed installare anche l'ultima versione di Rtools che trovate a questo link (<https://cran.r-project.org/bin/windows/Rtools/>)

Programma

MODULO 1 (Dott. Stefano Moro e Dott. Riccardo Giubilei)

Prima parte – Introduzione a R ed RStudio

1. R, RStudio - Introduzione
2. Primi passi – Funzioni, oggetti e operatori logici
3. Basi di statistica: indici di posizione, variabili e indici di variabilità
4. Data Visualization con tidyverse: ggplot2 e ggpubr
5. Data wrangling con tidyverse: dplyr e tidyr
6. Programmare in R: cicli, funzioni e accenni di programmazione funzionale

Seconda Parte - Analisi di dati con R

1. Test base (T-test, Test del Chi-quadro, Test di Shapiro-Wilk)
2. ANOVA, regressione semplice e multipla
3. GLM – Modelli lineari generalizzati

MODULO 2 (Prof. Pierfrancesco Alaimo e Prof. Marco Mingione)

Terza Parte – High dimensional regression and variable selection

1. Lasso e Ridge
2. Alberi di classificazione
3. Introduzione alle reti neurali

Quarta Parte – Costruzione di interfaccia web per l'analisi statistica (R Shiny)

1. Il pacchetto Shiny
2. Struttura di una ShinyApp e l'ambiente *reactive*
3. Costruzione di una ShinyApp: caso studio e *deploy*

References

- *Mastering Shiny*, Hadley Wickham, <https://mastering-shiny.org/>
- *An Introduction to Statistical Learning*, Hastie e Tibshirani, <https://www.statlearning.com/>
- *R for Data Science*, Wickham e Grolemund, <https://r4ds.had.co.nz/index.html>