

CORSO DI ANALISI DATI CON R – BREVE DESCRIZIONE DEL CORSO
DOCENTE – DR. BRUNO BELLISARIO, PHD

L'aumentata necessità di strumenti di analisi dati in ambito biologico ed ecologico va di pari passo con la crescente disponibilità di dati (spesso di differente origine) legata allo sviluppo di nuove e più performanti tecnologie di indagine. In tale contesto, la conoscenza di adeguati strumenti di analisi dati e/o linguaggi di programmazione diventa essenziale. Nello specifico, l'ambiente di lavoro R rappresenta uno degli strumenti maggiormente utilizzato in numerose discipline scientifiche, permettendo la gestione e l'analisi dei dati, nonché la visualizzazione di grafici. È un software libero (distribuito con la licenza GNU GPL) e multiplatforma (Unix, Linux, Mac OS X, Windows). Il presente corso ha lo scopo di introdurre le principali metodologie di analisi dati in ambito biologico ed ecologico sfruttando il potenziale offerto dal linguaggio di programmazione. Il corso è strutturato in due fasi, base ed avanzato, ciascuna delle quali suddivisa in quattro moduli da due ore ciascuno, unendo ad ogni argomento teorico una esercitazione numerica finalizzata alla familiarizzazione con l'ambiente di lavoro R.

DATA ANALYSIS WITH R – A BRIEF INTRODUCTION
INSTRUCTOR – DR. BRUNO BELLISARIO, PHD

The development of new and more performing technologies goes hand in hand with the increasing availability of data (often of different origins), which requires the development and use of up-to-date analytical and statistical tools. In this context, the knowledge of adequate programming languages becomes essential. Specifically, the R-language and working environment represents one of the most used and effective tools in many scientific disciplines, allowing the management and analysis of data. R is a free, multiplatform language distributed under the GNU GPL license (Unix, Linux, Mac OS X, Windows). This course aims to introduce the main methods of data analysis in the biological and ecological fields by exploiting the potential offered by the programming language. The course is structured in two phases, basic and advanced, each of which divided into four modules of two hours each, combining each theoretical topics with numerical exercises aimed at familiarizing yourself with the R work environment.

STATISTICA DI BASE CON R (IL PROGRAMMA DI MASSIMA POTREBBE SUBIRE VARIAZIONI)

(4 MODULI DA 2 ORE CIASCUNO)

Data: 22/06/2023 – 23/06/2023 - 26/06/2023 - 27/06/2023

Ora: 9-11

Introduzione a R

Modalità di download e procedure di installazione del software R
Pannelli e barre degli strumenti
Acquisizione di un file dati esterno al sistema
Uso dell'help e ricerca di comandi
Come scrivere ed eseguire uno script
Cicli e istruzioni condizionali per la definizione di routine
Cenni sui packages sviluppati in ambiente R

Statistica descrittiva

Cenni sulla trasformazione di variabili
La standardizzazione
Misure di tendenza centrale
Misure di variabilità
Statistiche per gruppi
Standard scores
Correlazioni
Metodologie di rappresentazione dei dati

Inferenza statistica#1

Verifica della normalità di una distribuzione
Statistica parametrica
Statistica non parametrica
Stima di un modello di regressione lineare
Regressione lineare multipla
Test di ipotesi per modelli di regressione
Verifica del modello
Selezione del modello

Inferenza statistica#2

Analisi tra gruppi
One-way, two-way ANOVA
Tecniche di classificazione
Clustering

STATISTICA AVANZATA CON R
(IL PROGRAMMA DI MASSIMA POTREBBE SUBIRE VARIAZIONI)

(4 MODULI DA 2 ORE CIASCUNO)

Data: 28/06/2023 – 29/06/2023 - 30/06/2023 - 03/07/2023

Ora: 9-11

Analisi multivariata

Tecniche di ordinamento

Analisi delle componenti principali (PCA)

Non metric multidimensional scaling (nMDS)

Analisi delle corrispondenze canoniche (CCA)

Redundancy analysis (RDA)

Regressione logistica

Modelli lineari generalizzati (GLMM)

Modelli additivi generalizzati (GAM)

Verifica del modello

Selezione del modello

Modelli probabilistici ed inferenza statistica

Introduzione al machine learning

Definizioni

k-nearest neighbors (knn)

Support vector machines

Alberi decisionali

Boosted Regression Trees