

SummeR School-DATA MINING AND DATA ANALYSIS WITH R

PROPONENTE

Prof.ssa Luisa Bernardinelli, Prof. Mario Grassi e Dott. Davide Gentilini

OBIETTIVI FORMATIVI

La gestione e l'analisi dei dati rappresenta uno dei fattori più importanti e simultaneamente più critici in molti ambiti lavorativi; spesso ci si avvale di strumenti e software proprietari che sono soventemente costosi o risultano essere in definitiva limitati nella loro applicazione.

Il corso ha l'obiettivo di fornire ai partecipanti le conoscenze necessarie e sufficienti per introdurre ed utilizzare il linguaggio e le potenzialità di R nel proprio lavoro.

R può essere definito come un sistema di analisi statistica e contemporaneamente un linguaggio ed un software. E' uno strumento open source potentissimo e largamente utilizzato, per l'analisi statistica dei dati, inoltre, essendo un vero e proprio linguaggio di programmazione, racchiude in se la potenzialità per creare e sviluppare in autonomia svariate applicazioni utili alla manipolazione, gestione ed analisi di ogni tipo di dato. Le sue caratteristiche principali comprendono infatti la semplicità nella gestione e manipolazione dei dati, la disponibilità di una suite di strumenti per calcoli su vettori, matrici ed altre operazioni complesse, l'accesso ad un vasto insieme di strumenti integrati e funzioni sviluppate da altri e resi disponibili per l'analisi statistica, la produzione di numerose potenzialità grafiche particolarmente flessibili, possibilità di adoperare un vero e proprio linguaggio di programmazione orientato ad oggetti che consente l'uso di strutture condizionali e cicliche, nonché di funzioni create dall'utente.

Offrendo un'elaborata introduzione alla programmazione con R questo corso ha lo scopo di intercettare le esigenze dei partecipanti focalizzandosi in particolar modo su alcuni aspetti fondamentali come la manipolazione e gestione dei dati, la loro analisi attraverso l'identificazione del test statistico più appropriato e la visualizzazione di dati e risultati utilizzando le potenzialità grafiche messe a disposizione da R.

Il corso prevede l'impiego di numerosi "dataset" ed esempi che possano essere familiari alle varie aree di interesse in modo da agevolare i partecipanti nella comprensione ed applicazione delle nozioni acquisite.

NUMERO DI ORE(CFU)/LEZIONI

Il corso avrà una durata di almeno 30 ore (7 CFU), e sarà suddiviso in 5 incontri.

Possono essere frequentati anche singoli incontri.

PERIODO DI SVOLGIMENTO

Il periodo di svolgimento del corso sarà dal 4 al 8 Giugno 2018 (summeR school) durante i mesi di giugno o luglio.

DOCENTI

Prof.ssa Luisa Bernardinelli, Dott. Davide Gentilini, Dott. Antonino Oliveri, Dott.ssa Teresa Fazia

COMITATO SCIENTIFICO

Prof Luisa Bernardinelli, Prof. Mario Grassi, Dr Davide Gentilini, Prof.ssa Elisa Fazzi

COMITATO ORGANIZZATORE

Dipartimento di Scienze del Sistema Nervoso e del Comportamento e Residenza Biomedica – Fondazione Collegio Universitario S. Caterina da Siena, Pavia

SEDE DEL CORSO

Residenza Biomedica –Fondazione Collegio Universitario S. Caterina da Siena, Pavia

POTENZIALI DOTTORATI INTERESSATI

Il corso ha l'intento di essere trasversale e di fornire competenze utili in ogni ambito che abbia a che fare con dati e con la necessità di elaborarli e gestirli. Per tale ragione il corso ha l'obiettivo di essere utile ad ogni tipologia di dottorato, in particolare macroarea di Scienze della vita, Bioingegneria e Bionformatica, Economics, Political Studies, DREAMT. Nell'attività pratica verranno utilizzati esempi e dati nei vari ambiti applicativi di interesse per i diversi Dottorati.

MODALITA' DI VERIFICA DELL'APPRENDIMENTO

Il grado di apprendimento verrà testato al termine di ogni lezione sottoponendo i partecipanti ad un test.

Il test sarà composto da una serie di 10 domande a scelta multipla riguardanti la parte teorica a cui verrà aggiunto un esercizio inerente le argomentazioni trattate.