

Analisi dei dati con R

Docente: Maria Chiara Debernardi

Lingua del corso

Italiano

Descrizione del corso e obiettivi

Il corso intende fornire un'introduzione all'utilizzo di R, presentando le caratteristiche e i principali comandi che R mette a disposizione per l'analisi dei dati utilizzando l'ambiente di sviluppo RStudio.

Nella seconda parte del corso saranno fornite alcune nozioni di programmazione con R e di utilizzo di package dedicati al data mining.

Il corso si pone un duplice obiettivo: da un lato presentare la struttura del software e la filosofia per l'analisi dei dati in esso implementata; dall'altro lato mostrarne le potenzialità attraverso la presentazione di alcune applicazioni in ambito economico e aziendale.

Al termine del corso i partecipanti saranno in grado di:

- Manipolare la maggioranza degli oggetti disponibili in R
- Effettuare analisi descrittive usando tabelle e grafici in R
- Creare semplici funzioni in linguaggio R
- Utilizzare le funzionalità dei principali package del Tidy-verse

Avviso: Il corso presenta il linguaggio R e le sue caratteristiche principali e non può essere considerato un sostituto di un corso di statistica o di econometria, quindi non saranno trattati i dettagli delle metodologie statistiche ed econometriche.

Destinatari

Il corso è aperto a tutti gli studenti Bocconi. In particolare si rivolge:

- A coloro che, a qualsiasi titolo, sono o saranno coinvolti in progetti che prevedono l'estrazione di informazioni da un set di dati di interesse per la redazione del progetto finale di triennio o della tesi
- A tutti coloro coinvolti nella preparazione di report contenenti analisi quantitative di dati nel contesto economico-sociale

Prerequisiti

È fondamentale che i partecipanti abbiano frequentato e superato positivamente il corso di informatica previsto dal proprio piano di studi o possiedano competenze equivalenti.

È inoltre auspicabile una buona conoscenza dei fondamenti di statistica descrittiva e inferenziale, corrispondenti al primo esame di Statistica del proprio piano di studi.

Durata

12 ore

Calendario

Lezione	Data	Ora	Aula
1	mar 20/11/2018	18.00 - 19.30	Info 6
2	mer 21/11/2018	18.00 - 19.30	Info 6
3	lun 26/11/2018	18.00 - 19.30	Info AS05
4	mer 28/11/2018	18.00 - 19.30	Info 6
5	lun 03/12/2018	18.00 - 19.30	Info 6
6	mer 05/12/2018	18.00 - 19.30	Info 6

Programma delle lezioni

Lezione	Argomenti	Riferimenti bibliografici
1	Introduzione a R <ul style="list-style-type: none"> - Introduzione a R - I primi passi nell'ambiente di RStudio - I package CRAN - Gli script in R 	Capp. 1, 4, 6
<i>Esercizi</i>		
2	Lavorare con i dati <ul style="list-style-type: none"> - I tipi di dati e di strutture complesse - Creazione di variabili e di data frame - Gestione dei formati, conversioni tra formati diversi - Il tempo in R 	Capp. 10, 14, 15, 16
<i>Esercizi</i>		

Lezione	Argomenti	Riferimenti bibliografici
3	Analizzare i dati <ul style="list-style-type: none"> - Operazioni sui vettori/matrici - Statistiche di sintesi uni e bi-variate - Importazione / Esportazione di dati - Il package tidyR 	Capp. 11, 12, 13, 20
<i>Esercizi</i>		
4	Analisi avanzata dei dati <ul style="list-style-type: none"> - Il package dplyr - Modelli lineari (cenni) - Rappresentare i dati attraverso i grafici 	Capp. 3, 5, 18, 23, 28
<i>Esercizi</i>		
5	Programmare in R <ul style="list-style-type: none"> - Il linguaggio R: le strutture di programmazione - Creare funzioni in R 	Capp. 19, 21
<i>Esercizi</i>		
6	Data mining in R <ul style="list-style-type: none"> - Come fare data mining con R - Alcuni problemi di classificazione - Il package broom 	Capp. 22, 25
<i>Test di autoverifica finale</i>		

Bibliografia consigliata

Testo fortemente orientato all'illustrazione del Tidy-verse (capitoli indicati nei riferimenti bibliografici):

Hadley Wickham, Garrett Golemund, *R for Data Science, 1st Edition*, O'Reilly Media, 2016

Anche consultabile online al link: <http://r4ds.had.co.nz/>

A supporto e integrazione del materiale fornito in aula:

Vito M. R. Muggeo, Giancarlo Ferrara, *Il linguaggio R: concetti introduttivi ed esempi, 2ª edizione*, 2005

Scaricabile dal link: <https://cran.r-project.org/doc/contrib/nozioniR.pdf>

Software di riferimento

Ultima release disponibile di R (versione 3.5.0 o superiore); si veda il sito

<https://www.r-project.org>

Ultima release disponibile di RStudio (versione 1.1.456 o superiore); si veda il sito

<https://www.rstudio.com>

Posti disponibili

110