

Introduzione al software statistico R

Area Informatica
Istat, Via Cesare Balbo, 16
Aula 001

OBIETTIVI FORMATIVI	Il corso è finalizzato a illustrare le principali caratteristiche del software libero R, che nasce come ambiente di programmazione, basato sul linguaggio S, esplicitamente dedicato alle elaborazioni statistiche.
DESTINATARI	Il corso è dedicato alle professionalità statistiche ed informatiche coinvolte nei processi di trattamento, elaborazione e analisi dei dati di indagine.
PREREQUISITI	Buona conoscenza della statistica di base e di elementi di programmazione, esperienza nei processi di trattamento e elaborazione dei dati.
DOCENTI	Docenza interna Istat
METODI	Il corso prevede l'illustrazione teorica dei contenuti ed esercitazioni pratiche sugli argomenti illustrati, anche on-line sulla piattaforma della formazione www.formazione.istat.it . Le lezioni di natura teorica (ogni volta due consecutive) prevedono, dopo una pausa di circa cinque giorni, un ritorno in aula di mezza giornata per svolgere le esercitazioni.
IMPATTO ORGANIZZATIVO ATTESO	L'obiettivo del corso è quello di introdurre i partecipanti a questo nuovo ambiente di lavoro (che si caratterizza per una interfaccia grafica con pochi menù), in modo da capire come utilizzarlo per trattare ed elaborare i dati a fini statistici.

PROGRAMMA

1-a Giornata

9.00 – 17.30



INTRODUZIONE

- ✚ Origini e sviluppo di R
- ✚ Ambiente di lavoro: R-studio
- ✚ Il sito: documentazione e packages aggiuntivi (Task views)

2. OGGETTI BASE

- ✚ a) introduzione agli oggetti base di R: mode, class, typeof
- b) Oggetti vector
 - vector numerici
 - vector di tipo stringa
 - vector di tipo logico
 - alcune semplici operazioni sui vector di tipo numerico, stringa e logico.
 - operazioni tra vector di diversa lunghezza
 - indicizzazione di vector (criterio posizionale e criterio logico)
- c) Oggetti factor
 - relazione con le modalità di variabili categoriche
 - levels
 - alcune funzioni che producono factor. La funzione "cut"
- d) Matrici
 - metodi per definire una matrice
 - dimensioni
 - indicizzazione di (insiemi di) elementi di matrici
 - alcune operazioni semplici su matrici: operazioni component wise; operazioni "globali" (e.g., trasposta, inversa, prodotto righe per colonne, ecc)
 - operazioni tra vector e matrici: regola del "riciclo"

2-a Giornata

9.00 – 17.30



3. OGGETTI COMPLESSI

- ✚ a) Data.frame
 - definizione ed esempi
 - data.frame come insieme rettangolare di dati unità x variabili
 - nomi di righe e colonne di un data.frame
 - selezione di colonne di data.frame
 - selezione di righe di data.frame
 - alcune semplici operazioni su data-frame
 - indicizzazione globale
 - gestione dei missing; differenze con SAS

4. LETTURA E SCRITTURA DI FILE DELIMITED

- ✚ Lettura e scrittura di file
 - funzioni per il percorso dei file
 - funzioni read.table() e write.table(): logica sottostante e descrizione dei parametri
- ✚ b) lettura e scrittura dei dati da Excel
- ✚ c) lettura e scrittura dei dati da SAS

- ✚ d) Analisi dei file
 - acquisizione di singole righe di un file
 - conteggio dei campi
 - conteggio del numero di record
- ✚ e) Lettura di file con criteri di selezione basati sulle variabili
- ✚ f) Lettura e scrittura di oggetti R tramite file
 - funzioni load() e save()
- ✚ g) Lettura di codice R da file
 - funzione source()

5. MANIPOLAZIONE DATI

- ✚ a) Operazioni di composizione
 - funzioni di affiancamento ed impilamento
 - funzioni per il merge
- ✚ b) Estrazione di sottoinsiemi
 - funzione subset ()
 - funzione sample() per sottoinsiemi casuali
- ✚ c) Ricodifica
 - funzione recode() e stringhe di ricodifica
- ✚ d) Ordinamento
 - funzioni sort(), rank(), order()
- ✚ e) Valori mancanti e duplicazioni
 - funzioni is.na(), na.omit(), complete.cases(), unique(), duplicated()
- ✚ f) Editing
 - modifica di dati tramite le funzioni edit() e fix()

3-a Giornata

9.00 – 13.30



Esercitazioni in aula

4-a Giornata

9.00 – 17.30



6. TABULAZIONE

- ✚ a) Funzioni base table(), ftable() e xtabs()
- ✚ b) Principali funzionalità del package aggiuntivo tables, che non è altro che una riproposizione in R della proc SAS tabulate

7. OPERAZIONI RIPETUTE

- ✚ a) Il ciclo FOR
- ✚ b) Istruzioni condizionali: Costrutto IF-ELSE, funzione ifelse()
- ✚ c) Cicli particolari: WHILE, REPEAT ed il loro controllo
- ✚ d) Funzione AGGREGATE: obiettivo, argomenti, risultato, manipolazione del risultato, particolarità, esempi
- ✚ e) Funzione APPLY: obiettivo, argomenti, risultato, manipolazione del risultato particolarità, applicazione a oggetti multidimensionali, esempi

5-a Giornata

9.00 – 17.30



8. ANALISI E VISUALIZZAZIONE DEI DATI

- ✚ a) Funzioni statistiche e modelli
- ✚ I modelli di regressione lineare in R:
 - funzione lm
 - funzioni per estrarre informazioni dal modello: anova, coef, formula, plot, residuals, summary
- ✚ b) Grafici
- ✚ Funzioni disponibili nella distribuzione base di R:
 - di alto livello: plot, coplot, boxplot, qqplot, hist, barplot;
 - di basso livello: points, lines, abline, legend, title

9. UTILIZZO DEL PACKAGE RCMDR: funzioni di base

- ✚ a) Utilizzo di Rcmdr nell'ambito di alcune delle funzionalità esposte nei precedenti moduli
 - Presentazione generale del pacchetto: menù a tendine, funzioni base del pacchetto, finestra degli script

6-a Giornata

9.00 – 13.30



Esercitazioni in aula