

L'idea di questo libro è nata dall'esigenza di sopperire alla mancanza di una guida di riferimento in lingua italiana all'uso del pacchetto statistico R e si è sviluppata in seguito come un vero e proprio corso di Statistica attraverso l'uso di questo applicativo. L'interesse della comunità statistica nei confronti di R cresce quotidianamente per diverse ragioni: si tratta di un prodotto disponibile gratuitamente, completo e in costante sviluppo curato da statistici ed esperti di computer graphics, molti dei quali di fama internazionale. A questo si aggiunge la caratteristica di essere un vero e proprio linguaggio di programmazione che è possibile interfacciare con routines, scritte in proprio, in differenti linguaggi (C, C++, Fortran). Un ulteriore punto di forza è la sua natura *cross-platform*, ovvero quella di un prodotto che gira sulle più svariate macchine, sia in termini di architettura (Intel, PPC) che di sistemi operativi (da Unix a Linux, da MS-Windows 95 a Windows-XP, da MacOS a MacOSX) tra i quali è possibile scambiare procedure e dati senza incorrere in noiose pratiche di conversione da un sistema all'altro.

Il libro affronta i principali argomenti trattati nei corsi istituzionali di Statistica (descrittiva e inferenziale) e del calcolo delle probabilità con cenni alla simulazione delle principali famiglie di processi stocastici. Lo svolgimento dei temi è trattato in modo elementare (e quindi non esaustivo sul piano teorico) limitando le nozioni a quanto necessario per rendere il lettore in grado di sperimentare subito, attraverso R, i concetti appena introdotti. Ovviamente si tratta solo di una goccia nell'oceano di possibilità offerte da un ambiente statistico quale è R ma è pur sempre, crediamo, un buon punto di partenza. Al contrario di testi simili, in cui l'utente deve sempre districarsi tra il set di funzioni offerte dal pacchetto statistico (che non sempre corrispondono ai concetti e alle tecniche presentate in aula) qui si è sempre cercato di mettere R a servizio della didattica. Durante lo svolgimento del corso verranno quindi costruite delle funzioni in codice R che renderanno il passaggio dalla teoria alla pratica del tutto naturale e, allo stesso tempo, renderanno l'utente sempre più autonomo rispetto al linguaggio di programmazione in questione.

Il volume si apre con un breve excursus sui *fondamentali* di R ovvero: quali sono i tipi di dati e come possono essere trattati, come iniziare e terminare una sessione di lavoro, cosa sono e come si costruiscono delle funzioni e, infine, pochi elementi di algebra. Lo scopo è quello di rendere semplificata la trattazione del resto del materiale presentato. Il secondo e terzo capitolo trattano diversi argomenti di statistica descrittiva. In particolare, nel primo si affrontano tematiche di statistica descrittiva unidimensionale mentre nel terzo ci si occupa dell'analisi congiunta dei fenomeni statistici. Un intero capitolo, il quarto, è dedicato al cal-

colo delle probabilità e alla simulazione. Dagli elementi di calcolo combinatorio si passa alle più importanti famiglie di variabili casuali discrete e continue. Si tratta poi l'argomento della generazione dei numeri (pseudo) casuali. Benché R disponga di ottimi generatori di numeri casuali, in Statistica è spesso necessario simulare variabili casuali *ad hoc*. Inoltre, si affronta il problema della simulazione di alcuni dei processi stocastici più rilevanti in ambito statistico: dalla passeggiata aleatoria alle catene di Markov, dal processo di Poisson ai modelli autoregressivi per finire con cenni sui processi di diffusione. Il capitolo quinto affronta i temi classici della statistica inferenziale ovvero gli intervalli di confidenza e la verifica di ipotesi. Inoltre si considera il problema della verifica di ipotesi per campioni appaiati e si conclude con la presentazione di alcune tecniche per la verifica di ipotesi di adattamento.

Vi sono inoltre due appendici che abbiamo ritenuto utile inserire e che crediamo siano di un qualche interesse anche per gli utenti smaliziati magari abituati a lavorare con altri pacchetti statistici. Nel passare da un ambiente di lavoro all'altro spesso si incorre nella necessità di far digerire all'ambiente di arrivo dati generati con un altro sistema. Nell'Appendice A trattiamo dunque della possibilità e delle modalità con cui si possono importare ed esportare i dati verso gli ambienti più noti: SAS, SPSS, Stata, MS-Excel, Minitab ecc. e come sia possibile interfacciarsi con i database relazionali. La stessa appendice contiene inoltre una risposta, il più possibile esauriente a diverse altre questioni: come si esportano tabelle e grafici in modo che sia possibile inserirli in una relazione? Come si installa un nuovo pacchetto? Inoltre, per chi si occupa di statistica computazionale, abbiamo inserito una breve rassegna sulle procedure di ottimizzazione incluse in R. L'Appendice B invece contiene una breve rassegna della manualistica ufficiale e del materiale bibliografico specifico per R nonché un elenco di risorse di rete dove è possibile trovare dataset e pacchetti utili per la pratica statistica. Questa sezione contiene anche un breve manuale d'uso di tutte le funzioni definite nel testo.

Al libro è anche allegato un CD-ROM in doppio formato MS-Windows e Macintosh. Il CD-ROM contiene una versione di R con un'ampia raccolta di pacchetti aggiuntivi e l'intero archivio dei dataset e delle funzioni utilizzate durante il corso raccolti nel pacchetto `labstatR`. Tale pacchetto aggiornato all'ultima versione di R è reperibile su <http://CRAN.R-project.org>

I comandi R che verranno impiegati nel corso della trattazione vengono trascritti nel seguente modo: comando omettendo la coppia di parentesi “()”. Le definizioni dei concetti statistici verranno evidenziate in **grassetto** mentre in *corsivo* sono riportati termini ritenuti importanti e che eventualmente saranno definiti in quanto segue. Le nuove funzioni R create durante il corso vengono denominate Codice e sono evidenziate nel testo con delle finestre analoghe al Codice 1.

In tutto il testo si assume che l'utente interagisca continuamente con R. Questo in realtà è solo una possibile modalità di utilizzo dell'ambiente R e non è necessariamente la più comoda. A titolo d'esempio si può notare che l'inserimento dei dati avverrà quasi sempre in modalità interattiva. Al contrario, molto spesso i dati sono disponibili su file esterni. I consigli e le modalità di gestione dei file esterni sono riportati nella prima Appendice. Si noti infine che tutti gli esempi di codice R riportati nel testo iniziano con `>` ed eventualmente, se troppo lunghi,

```
KandR <- function(){  
  print("Hello World!\n")  
}
```

**Codice 1** Citazione dal Kernighan and Ritchie [23].

continuano sulla linea successiva con un +. Il lettore non dovrà inserire tali simboli poiché sarà R stesso ad inserirli come spiegato all’inizio del Capitolo 1. Tali simboli sono invece riportati nel testo per rendere con più efficacia l’aspetto di una sessione di lavoro con R.

### *Ringraziamenti*

Non ci resta che ringraziare quanti hanno letto le bozze del libro fornendo utili consigli. Tra questi, in particolar modo, un sentito ringraziamento va a Martin Mächler, Dipartimento di Matematica, Swiss Federal Institute of Technology, Zurigo (ETHZ). Ovviamente non si può non citare l’intero R-Core Team per averci avvicinato, in momenti diversi, allo sviluppo di questo software e per aver creato un così importante strumento di lavoro e ricerca.

Ludovico non se ne avrà a male se il papà gli dedica questo libro non di poesie o intense passioni ma di numeri e calcolatori. In fondo, un computer si sceglie, i genitori “capitano”.

### **Nota alla seconda edizione**

Le parti maggiormente modificate in questa edizione riguardano i capitoli iniziali e le appendici. Infatti in questa seconda edizione si è reso necessario allineare il codice usato nel testo alla versione corrente di R e si è anche voluto rispondere alle richieste di approfondimento rispetto alla descrizione degli oggetti R (capitoli 1 e 2) e della gestione dei pacchetti del software (Appendici). Una differenza evidente tra la precedente edizione e questa attuale è che R è ormai disponibile anche in lingua italiana, quindi quasi tutto l’output prodotto da R e incluso in questa nuova edizione risulta differente rispetto alla prima edizione del libro. La traduzione riguarda solo i messaggi di R e non anche i comandi stessi.

### *Ringraziamenti*

Questa seconda edizione è il frutto dei consigli e richieste di lettori, colleghi e amici raccolte negli ultimi tre anni. Tra questi citiamo, in ordine rigorosamente causale, Sergio Polini, Luciano Matrone, Massimo Borelli, Sergio Invernizzi, Fabio Frascati.

Nel frattempo anche a Lucia sono “capitati” gli stessi genitori.