

R: un ambiente *opensource* per l'analisi statistica dei dati

Versione 0.9 -30 novembre 2004

Vito Ricci
vito_ricci@yahoo.com

E' garantito il permesso di copiare, distribuire e/o modificare questo documento seguendo i termini della Licenza per Documentazione Libera GNU, Versione 1.1 o ogni versione successiva pubblicata dalla Free Software Foundation. La Licenza per Documentazione Libera GNU è consultabile su Internet:
originale in inglese: <http://www.fsf.org/licenses/licenses.html#FDL>
traduzione in italiano: <http://www.softwarelibero.it/gnudoc/fdl.it.html>

La creazione e distribuzione di copie fedeli di questo articolo è concessa a patto che la nota di copyright e questo permesso stesso vengano distribuiti con ogni copia. Copie modificate di questo articolo possono essere copiate e distribuite alle stesse condizioni delle copie fedeli, a patto che il lavoro risultante venga distribuito con la medesima concessione.

Indice

Abstract

1 Introduzione

2 Che cos'è R

3 Le funzionalità dell'ambiente R

4 Breve disamina di alcune risorse disponibili su R

5. R e la statistica ambientale

6 Conclusioni

Bibliografia

Abstract

L'articolo vuole essere una presentazione non esaustiva del linguaggio ed ambiente per l'analisi statistica R senza entrare in dettagli troppo tecnici. Dopo una panoramica sullo stesso vengono sommariamente fatti riferimenti alle funzionalità e alle potenzialità dell'ambiente. Una rassegna delle risorse attualmente disponibili conclude l'intervento.

This paper would be a short introduction and overview about the language and environment for statistical analysis R, without entering in specific details too much computational. I give a look about this opensource software pointing out its main features, its functionalities, its pros and cons describing some libraries and the kind of analysis they support. I supply a summary, with a short description, about many resources concerning R that can be found in the Web: the most are in English language, but there are also some in the Italian language.

The aim of this work is to contribute in increasing of the use of the R environment in Italy among statistical researchers trying to "advertise" this software and its opensource philosophy.

1. Introduzione

Esiste in circolazione una gamma assai vasta di software specializzati nell'analisi statistica dei dati, basta ricordare SAS, SPSS, STATA, STATGRAPHICS PLUS, SHAZAM, S PLUS, MINITAB, GAUSS, etc., solo per citarne alcuni che sono in commercio. Sono prodotti che costituiscono senz'altro un fondamentale ed insostituibile ausilio per il lavoro dello statistico, tuttavia molti di questi programmi sono anche alquanto costosi ed è consentito l'uso su licenza da parte del produttore.

Da alcuni anni a questa parte, soprattutto in ambito universitario (ma non solo) si sta sempre più diffondendo un nuovo package che merita di sicuro una debita trattazione e l'interessamento da parte degli statistici e di coloro che fanno analisi dei dati e che costituisce anche una valida alternativa ai software sopraccitati. Ci stiamo riferendo al software **R**. Probabilmente qualcuno forse lo conosce già e lo adopera, altri ne avranno solo sentito parlare, per altri ancora è una novità assoluta. **R** è un ambiente statistico scaricabile gratuitamente da Internet sul sito di **The R Project for Statistical Computing** il cui indirizzo Internet è <http://www.r-project.org>¹. Esso è il frutto del lavoro collettivo svolto da un gruppo, sempre più folto, di ricercatori in campo statistico ed informatico a livello mondiale. **R** più che un software statistico può essere definito come un ambiente, costituito da una varietà di strumenti, orientato alla gestione, all'analisi dei dati e alla produzione di grafici, basato sul linguaggio S creato da AT&T Bell Laboratories, ambiente dal quale è nato un altro software commerciale più noto, S-Plus, prodotto prima da MathSoft e ora da Insightful. **R**, a differenza di quest'ultimo è un *GNU-Software*², ossia è disponibile

¹ Attualmente del software R è disponibile la versione 2.0.1 rilasciata il 15/11/04

² Per un approfondimento del progetto GNU si consulti il sito Internet <http://www.gnu.org>

gratuitamente sotto i vincoli della GPL (General Public License)³: chiunque può accedere al codice sorgente e modificarlo, migliorarlo, integrarlo e mettere a disposizione della comunità scientifica il proprio lavoro nella logica e nella filosofia del software *opensource*.

La versione iniziale di **R** venne sviluppata nel 1996 per il sistema operativo MacOS da Ross Ihaka e Robert Gentleman del dipartimento di Statistica dell'Università di Auckland in Nuova Zelanda⁴. Successivamente altri ricercatori iniziarono ad aggiungersi e a fornire il loro contributo alla scrittura del codice sorgente e al miglioramento dell'applicazione, sviluppata in seguito anche per l'ambiente UNIX. Nel 1997 è nato l'**R Development Core Team**, gruppo formato da statistici di tutto il mondo che si occupano dello sviluppo e della distribuzione del programma e, ancora più recentemente, nel 2003 è stata costituita dagli stessi membri dell'**R Development Core Team** la **R Foundation for Statistical Computing**. Si tratta di una organizzazione non profit che lavora per il pubblico interesse allo scopo di promuovere lo sviluppo (attuale e futuro) e la diffusione del software, di fornire un punto di riferimento per persone, istituzioni ed imprese che desiderano sostenere il progetto e interagire con la community che ruota intorno ad **R**, di gestire e tutelare il copyright di **R** e della relativa documentazione.

Quest'anno dal 20 al 22 maggio si è svolta a Vienna la prima conferenza dedicata agli utilizzatori di **R**⁵.

2. Che cos'è R?

R può essere definito come un sistema di analisi statistica, che è contemporaneamente un linguaggio ed un software. Lo stesso **R Development Core Team**⁶, parlando di **R**, preferisce definirlo come un ambiente integrato e coerente, piuttosto che come la somma di un insieme di strumenti poco flessibili, oppure come un semplice software o package al pari di prodotti simili, stando implicitamente a significare la presenza di un maggior "*valore aggiunto*" nel "prodotto" **R**.

Le sue caratteristiche principali possono essere così riassunte:

- a) semplicità nella gestione e manipolazione dei dati;
- b) disponibilità di una suite di strumenti per calcoli su vettori, matrici ed altre operazioni complesse;

³ Sulla GPL si veda: <http://www.gnu.org/copyleft/gpl.html> e <http://www.softwarelibero.it/gnudoc/gpl.it.txt>

⁴ Ross Ihaka and Robert Gentleman, R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3):299-314, 1996

⁵ Per informazioni sulla conferenza si veda l'URL: <http://www.ci.tuwien.ac.at/Conferences/useR-2004/>

- c) accesso ad un vasto insieme di strumenti integrati per l'analisi statistica;
- d) produzione di numerose potenzialità grafiche particolarmente flessibili;
- e) 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.

E' distribuito gratuitamente sotto i vincoli della GPL (General Public License) ed disponibile per diverse architetture hardware e sistemi operativi: Unix, Linux, Windows, MacOS. E' quello che tecnicamente si definisce un software *multiplatforma*. Sul sito <http://www.r-project.org> è possibile scaricare, oltre che il programma base, anche una serie di moduli aggiuntivi e un'ampia manualistica (in lingua inglese) sull'ambiente che va dall'installazione del software al suo utilizzo nell'analisi dei dati.

Sul sito citato sono disponibili i file binari di **R** (contenenti il codice sorgente scritto parte nel linguaggio C e parte nel linguaggio Fortran) alla cui realizzazione e al miglioramento partecipano una pluralità di soggetti che operano nella parti più disparate del mondo e che riescono a collaborare e cooperare grazie ad Internet, utilizzando come strumenti di comunicazione i forum, le liste di discussione, la posta elettronica.

R si presenta come uno strumento completo e ben si adatta a molteplici esigenze: è il software maggiormente utilizzato a livello mondiale dai ricercatori in campo statistico. Esaminiamo ora i suoi punti di forza ed anche alcune sue debolezze e/o mancanze.

- 1) **R** è un software *opensource*, come già detto, e in virtù di ciò:
 - a) si ha la possibilità di accedere al codice sorgente e di modificarlo
 - b) ha un costo zero per l'utente finale e presenta una facilità nell'accesso a particolari categorie di utenti che non possono permettersi costosi software per analisi statistiche come gli studenti oppure gli utilizzatori e i ricercatori dei Paesi in via di sviluppo
 - c) dispone di una vasta manualistica (in lingua inglese) consultabile e scaricabile da Internet (si veda oltre)
 - d) è possibile accedere tramite Internet ad una vasta gamma di librerie per analisi statistiche molte dettagliate create e messe a disposizione di tutti da parte di ricercatori di tutto il pianeta
 - e) è possibile contare sul supporto e sull'assistenza dell'**R Development Core Team** e di tutti gli utenti di **R** (una vera community a livello mondiale) grazie al sito Internet e alle liste di discussione tramite le quali ci si

⁶ Cfr. la pagina web del sito The R Project for Statistical Computing, "What is R?" all'indirizzo: <http://www.r-project.org/about.html> [consultata in data 19/01/04]

può avvalere dell'aiuto di esperti in statistica e informatica di tutto il pianeta

- 2) **R** è un ambiente estremamente versatile:
- a) con possibilità di creare strumenti personali di analisi statistica ad hoc necessari per le proprie ricerche
 - b) con disponibilità di strumenti di analisi statistica (già pronti) dai più elementari ai più sofisticati
 - c) con notevoli possibilità grafiche rispetto ad altri software statistici (probabilmente **R** è il software che consente di ottenere il più elevato numero di rappresentazioni grafiche che si possono definire da sole oppure utilizzando diversi pacchetti basati su Java, OpenGL, etc.)
 - d) con possibilità di programmare **R** creando funzioni e programmi definiti dall'utente⁷
 - e) dotato di funzione di help in linea per ciascun comando facilmente richiamabile dal programma
 - f) è un software in rapidissimo sviluppo
 - g) possiede tutti i vantaggi dei software multiplatforma
 - h) ha la possibilità di integrazione, con l'uso di apposito software gratuito e scaricabile da Internet, con interfaccia grafica (GUI)
 - i) può essere integrato facilmente, con l'uso di apposito software gratuito e scaricabile da Internet, con alcuni dei più diffusi prodotti di office automation, come ad esempio Microsoft Excel in ambiente Windows
 - j) può essere interfacciato, con l'uso di apposito software gratuito e scaricabile da Internet, con i principali programmi di gestione di database come MySQL, Microsoft Access, Oracle, PostgreSQL per l'importazione ed esportazione di dati⁸
 - k) consente, tramite un'apposita libreria scaricabile gratuitamente da Internet, di importare i file dei dati da software statistici quali: Minitab, S, SAS, SPSS, Stata
 - l) è abbastanza semplice nell'impiego a livello didattico nell'insegnamento della statistica
 - m) è un linguaggio di programmazione orientato ad oggetti (object oriented)⁹ e consente la manipolazione di matrici e vettori

⁷ Si veda: R Development Core Team, Writing R Extensions, versione 2.0.1 del 15/11/04 all'indirizzo Internet: <http://cran.r-project.org/doc/manuals/R-exts.pdf>

⁸ Si veda: R Development Core Team, R Data Import/Export, versione 2.0.1 del 15/11/04 all'indirizzo Internet: <http://cran.r-project.org/doc/manuals/R-data.pdf>

⁹ Si veda: R Development Core Team, The R language definition, 2.0.1 del 15/11/04 all'indirizzo Internet: <http://cran.r-project.org/doc/manuals/R-lang.pdf>

Esiste un'importante differenza a livello di filosofia tra **R** e la maggior parte degli altri packages statistici. Con gli altri software un'analisi statistica porta ad una grande quantità di output di informazioni e dati, mentre con **R** l'analisi statistica è condotta come una serie di passi, con risultati intermedi memorizzati in oggetti. Ad ogni passo dell'analisi gli output sono minimi e l'utente ha la possibilità di visualizzarli e manipolarli richiamando, quando è necessario, gli oggetti nei quali i risultati dell'analisi sono contenuti¹⁰.

Tra i difetti e le lacune di **R** qualche autore¹¹ ha riscontrato che:

- a) è un ambiente basato su un'interfaccia utente a carattere (CUI), anche se è possibile implementare un'interfaccia GUI (Graphical User Interface), le funzioni e le istruzioni vengono immesse da una linea di comando (command line interface), che, tuttavia, è molto potente e semplice da usare
- b) possono mancare alcune funzioni statistiche
- c) la sintassi del linguaggio a volte risulta alquanto inconsistente

Tuttavia a tali affermazioni possiamo chiaramente controbattere che:

- 1) esistono delle interfacce GUI per **R** come (R Commander) che sono assolutamente equivalenti a quelle dei software commerciali¹²
- 2) alla mancanza di alcune funzione statistiche si può sopperire scrivendo il codice e creando da sé questi strumenti, anche se ciò richiede maggiore padronanza di strumenti informatici avanzati che non tutti gli statistici possiedono
- 3) circa la presunta inconsistenza della sintassi bisogna ricordare che nel 1998 John Chambers ha ricevuto un importante riconoscimento (The Association for Computing Machinery award for Software Systems) per la definizione del linguaggio S (il progenitore di **R**) precipuamente per la sua superiorità nella struttura e nella consistenza della sintassi.

¹⁰ Cfr. Jeff Racine, Rob J. Hyndman, Using R to Teach Econometrics, Working Paper 10/2001, Department of Econometrics & Business Statistics, Monash University, Melbourne, novembre 2001

¹¹ Jeff Racine, Rob J. Hyndman, Using R to Teach Econometrics, op. cit., pag. 3

Come può evincersi, anche da un punto di vista puramente quantitativo, gli aspetti positivi del software **R** prevalgono nettamente su quelli negativi e ne hanno decretato il successo a livello mondiale all'interno della comunità di ricercatori in ambito statistico.

3. Le funzionalità dell'ambiente **R**

L'ambiente **R** è basato sul concetto di "*package*" tradotto di solito in italiano con pacchetto. Un package è un insieme di strumenti che svolgono determinate funzioni, ma può anche contenere solo dati oppure solo documentazione. Attualmente è disponibile una vasta gamma di packages (scaricabili, per il sistema operativo Windows, dall'URL <http://cran.at.r-project.org/bin/windows/contrib>) utilizzabili per la risoluzione di specifici problemi o per analisi statistiche molto particolareggiate. Il cuore di **R** è rappresentato dal modulo base (che offre gli strumenti fondamentali per l'analisi statistica) e attorno a questo modulo "ruotano" una serie di altre librerie addizionali, alcune delle quali sono già comprese nel programma **R** al momento in cui lo si installa, mentre altre librerie ancora, in relazione alle esigenze e necessità, possono essere aggiunte e installate dall'utente dopo averle scaricate dal sito suddetto. **R** è un ambiente computazionale molto dinamico ed in continua evoluzione e frequentemente nuovi ulteriori packages sono realizzati e messi a disposizione dei ricercatori.

Esaminiamo sinteticamente i principali packages di **R** puntualizzando alcune delle funzionalità che questi consentono di operare. Quella che segue è la situazione valida sino alla versione 1.8.1.

- 1) Package "*base*": come già detto questo modulo è il motore dell'ambiente **R**, viene caricato automaticamente e fornisce all'utente gli strumenti per le più importanti e diffuse analisi statistiche come: ANOVA, regressione lineare, statistica descrittiva ed inferenziale, analisi esplorativa dei dati, grafici elementari, modelli lineari generalizzati, generazione di campioni delle più comuni variabili casuali, operazioni su matrici e vettori.
- 2) Package "*ctest*": consente di effettuare tutti i principali test statistici per la verifica delle ipotesi (test t, test F, test di normalità, test non parametrici, test per l'omoscedasticità delle varianze, test Chi-quadro, etc.)
- 3) Packages "*ts*" e "*tseries*": sono due packages dedicati all'analisi delle serie temporali
- 4) Package "*spatial*": analisi dei dati di serie spaziali

¹² Per ulteriori notizie su R Commander si consulti l'URL: <http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/>

- 5) Packages "*grid*", "*lattice*", "*rgl*" e "*scatterplot3d*": permettono di realizzare e manipolare grafici avanzati e tridimensionali
- 6) Packages "*mva*", "*amap*", "*multidim*" e "*multiv*": permettono di eseguire le principali analisi su dati multidimensionali (analisi delle componenti principali, analisi fattoriale, correlazione canonica, scaling multidimensionale, clustering gerarchico)
- 7) Package "*cluster*": è il modulo specializzato nell'analisi dei gruppi (cluster analysis)
- 8) Package "*nls*": effettua l'analisi della regressione con modelli non lineari applicando il metodo dei minimi quadrati (Nonlinear Least Squares)
- 9) Package "*matrix*": strumenti per l'uso avanzato di matrici e vettori con metodi numerici per l'algebra lineare
- 10) Packages "*survival*" e "*survrec*": pacchetti specializzati nell'analisi della sopravvivenza
- 11) Package "*nlme*": modelli lineari e non lineari con effetti misti (Linear and Nonlinear mixed effects models)
- 12) Package "*foreign*": contiene tutta una serie di funzioni utili per importare file da software statistici quali Minitab, S, SAS, SPSS, Stata.

Con il rilascio della versione 1.9.0.¹³ avvenuta il 12 aprile 2004 ci sono state importanti modifiche nella struttura organizzativa dei packages all'interno di **R**. In particolare il package che nella precedente release era denominato "*base*" è stato diviso in quattro pacchetti denominati: "*base*", "*graphics*", "*stats*", "*utils*". Tutti e quattro sono caricati in una installazione standard, ma tale divisione alleggerisce il lavoro svolto dal sistema, consentendo, anche grazie ad altre modifiche apportate, migliori prestazioni. Inoltre si è riusciti ad aumentare la velocità di start up di una sessione di **R**: il tempo impiegato con l'ultima versione è circa i due terzi di del tempo della versione 1.8.1. Il package "*base*" continua a contenere i tools strettamente di base del software, mentre i packages "*ctest*", "*eda*", "*modreg*", "*mva*", "*nls*", "*stepfun*" and "*ts*" sono confluiti nel pacchetto "*stats*". Anche tutti gli archivi (datasets) contenenti serie storiche sono stati spostati in quest'ultimo package. "*graphics*" contiene tutti gli strumenti di grafica di base, invece "*utils*" contiene delle utility del sistema tra cui i comandi per help, l'installazione e la rimozione di nuovi packages, il lancio di sessioni dimostrative delle funzionalità del programma (demo), etc..

Con la release 2.0.0¹⁴ disponibile dal 4 ottobre 2004 si sono avute ulteriori modifiche e aggiornamenti. Tra le principali novità

¹³ Si veda l'URL: <http://cran.r-project.org/bin/windows/base/NEWS.rw1090>

¹⁴ Si veda l'URL: <http://microarrays.unife.it/CRAN/bin/windows/base/NEWS.rw2000>

- 1) è stato implementato un nuovo sistema di caricamento del programma e dei packages detto '*lazy loading*': al posto di mantenere in memoria gli oggetti di **R**, questi sono memorizzati in un database su disco rigido e caricati solo al primo uso. Ciò consente un più veloce start up con una riduzione del 40% rispetto alla versione 1.9.x e un uso più efficiente della memoria dell'elaboratore;
- 2) tutti i datasets che prima si trovavano nei packages "*base*" e "*stats*" sono stati inseriti in un nuovo package denominato "*datasets*";
- 3) il package "*graphics*" è stato diviso in due pacchetti "*grDevices*" (dispositivi grafici condivisi tra i grafici di base e quelli del package "*grid*") e "*graphics*" (grafici di base);

Riteniamo i packages sopraccitati i più significativi che abbracciano le più importanti e diffuse tipologie di analisi statistiche dei dati, sul sito del **The Comprehensive R Archive Network - CRAN** (all'URL sopraccitata) è possibile scaricare più di duecentocinquanta packages che spaziano nei più disparati campi della statistica applicata.

In conclusione di paragrafo si vuole segnalare un ambiente molto specialistico basato su **R**, formato da una collezione di packages e funzioni: si tratta di "*Rmetrics*"¹⁵ che può essere un utile strumento per l'insegnamento e le applicazioni legate alle analisi finanziarie e alla finanza computazionale.

4. Breve disamina di alcune risorse disponibili su R

E' disponibile su Internet una vasta gamma di materiale utile per avvicinarsi all'ambiente **R** e aiutare l'utente, anche novizio, nell'apprendimento dell'uso di questo software statistico. Cercheremo di fornire alcune indicazioni e una breve descrizione per le risorse più interessanti e fondamentali, indicando gli indirizzi da cui poterle scaricare. Si è fatta una distinzione tra la manualistica ufficiale predisposta dall'**R Development Core Team** e altri contributi disponibili scritti da ricercatori e studiosi, soffermandoci più a lungo su quelli in lingua italiana.

a) Manualistica ufficiale

La pagina del sito del **CRAN** da cui si può accedere ad una vasta scelta della manualistica ufficiale su **R** è:

<http://cran.r-project.org/manuals.html>. Da questa URL è possibile accedere all'ultima versione dei manuali. Al momento la versione corrente è la 2.0.1 rilasciata in data 15/11/04 e i manuali (in lingua inglese) sono disponibili in formato PDF leggibili con il software gratuito Acrobat Reader, ormai presente su tutti i personal computer,

¹⁵ Si veda l'URL <http://www.itp.phys.ethz.ch/econophysics/R/>

oppure scaricabile dal sito della Adobe: <http://www.adobe.com/products/acrobat/readstep2.html>. In bibliografia per ciascun manuale è indicato l'URL precisa da cui effettuare il download.

Un primo approccio con **R** lo si può avere con la lettura di "*An Introduction to R*" che fornisce le indicazioni di base e i concetti introduttivi sull'ambiente e su come usare **R** per realizzare alcune analisi statistiche e i grafici. Dopo aver preso una certa dimestichezza con l'ambiente di lavoro si può passare a "*The R language definition*", testo di livello più avanzato che tratta dell'uso di **R** come un vero e proprio linguaggio di programmazione orientata ad oggetti. Altri manuali trattano di problematiche specifiche come, ad esempio, l'importazione e l'esportazione di dati da e verso altri programmi che gestiscono database (si veda "*R Data Import/Export*"), le indicazioni per poter creare personali librerie ad hoc in ambiente **R** utili a risolvere le proprie specifiche problematiche (si veda "*Writing R Extentions*"), oppure sulle modalità di installazione di **R** nei vari sistemi operativi e sulle diverse piattaforme di computers (si veda "*R Installation and Administration*"). È disponibile, inoltre, "*The R Reference Index*" il manuale completo di riferimento del software **R** che, per le più diffuse librerie, fornisce una descrizione dei singoli comandi e delle istruzioni ed il loro funzionamento. Inoltre ogni libreria aggiuntiva dispone di una specifica manualistica con dettagli sulle proprie funzionalità. Si ricorda che dalla linea di comando del programma è sempre possibile richiamare l'aiuto on line e dalla barra dei menù si possono consultare i principali manuali di cui si è detto sopra (sia in formato PDF che in formato HTML) nonché una pagina delle FAQ (Frequently Asked Questions) su **R**, consultabile anche su Internet all'URL: <http://cran.r-project.org/faqs.html>.

b) Risorse in lingua italiana

L'ambiente **R** sta prendendo sempre più piede anche nel nostro paese, soprattutto come software utilizzato nell'ambito dei corsi universitari di statistica. Esistono su Internet manuali, dispense e appunti realizzati e predisposti da docenti per i propri corsi e, in alcuni casi, è possibile trovare anche delle esercitazioni pratiche con **R**.

Una prima risorsa è quella scritta da Claudio Agostinelli, "*Introduzione ad R*" (150 pagine, scaricabili dall'indirizzo Internet: <http://www.dst.unive.it/~claudio/R/index.html#manuale>) che costituisce una buona base di partenza per cominciare lo studio del linguaggio **R** trattando i concetti fondamentali e soffermandosi sulla regressione lineare. Assieme al manuale è possibile scaricare dei files contenenti i dati per svolgere le esemplificazioni pratiche trattate nello stesso. Di livello leggermente più avanzato è la dispensa di Angelo M.

Mineo, "*Una guida all'utilizzo dell'ambiente statistico R*", (67 pagine, scaricabili dall'indirizzo Internet: <http://cran.r-project.org/doc/contrib/Mineo-dispensaR.pdf>), il quale, oltre a trattare i fondamenti di **R**, dà qualche cenno sull'uso del linguaggio di programmazione. Tale risorsa manca, tuttavia, di esempi pratici e di riferimenti concreti alle applicazioni statistiche. Seppure estremamente sintetica, può essere utile la dispensa di Vito M. R. Muggeo, "*Il linguaggio R: concetti introduttivi ed esempi*" (33 pagine, scaricabili dall'indirizzo Internet: <http://cran.r-project.org/doc/contrib/nozioniR.pdf>) per avere giusto una panoramica sull'ambiente e un interessante e breve accenno all'impiego dei modelli lineari generalizzati con **R**.

"*Introduzione ad R*" di Roberto Baggiani (100 pagine scaricabili dall'URL: <http://digilander.libero.it/robicox/manuali/pdf/mainr.pdf>) è un testo ben strutturato e completo che spazia dagli aspetti meramente introduttivi del software, ad alcuni brevi cenni sulla programmazione e sullo scripting, sino alla trattazione analitica delle principali tecniche di analisi statistica univariata (grafici, regressione, tests, ANOVA). La presenza di parecchio codice in **R**, con relativa esemplificazione, corona questa risorsa preziosa per l'utente che si avvicina al package statistico.

"*La statistica applicata attraverso il programma R*" di Francesca Parpinel (26 pagine, scaricabili dall'indirizzo Internet: http://venus.unive.it/statcomp/r/man_Parpinel.pdf) è un testo alquanto pratico e tratta di alcune applicazioni del linguaggio **R** in diversi ambiti statistici (serie storiche, intervalli di confidenza, analisi delle dipendenza). Altrettanto pratico, dopo aver fornito al lettore alcuni contenuti di livello elementare sull'ambiente **R**, è pure "*Note sul linguaggio e ambiente statistico R*" realizzato da Luca Scrucca (32 pagine, scaricabili dall'indirizzo Internet: <http://www.stat.unipg.it/~luca/R-note.pdf>) che si occupa di statistica univariata, rappresentazioni grafiche, modelli lineari e ANOVA con l'uso di **R**.

Per chi voglia passare subito all'azione usando direttamente il software **R**, senza indugiare troppo sugli aspetti teorici, si consigliano le seguenti dispense con delle esercitazioni di laboratorio d'informatica:

- Alessio Pollice, "*Esercitazioni con R*", materiale utilizzato nel corso di Statistica IV presso la facoltà di Economia dell'Università di Bari, si sofferma soprattutto sull'impiego di **R** per l'analisi di dati multivariati (si può scaricare dall'URL: <http://www.dip-statistica.uniba.it/html/docenti/pollice/materiale.htm>)

- Lea Petrella, Silvia Poletti, "*Laboratorio di Statistica I, metodi matematici e statistici*", utilizzato per l'omonimo corso presso l'Università Roma 3, si occupa dell'applicazione di **R** a problemi di calcolo delle probabilità e di statistica matematica (si può scaricare dal seguente indirizzo Internet: http://www.mat.uniroma3.it/didatticacds/corsi/didattica_interattiva/aa_01_02/st1/st1.html).

- A. Brazzale, M. Chiogna, C. Gaetan e N. Sartori, "*Laboratorio di R, Materiale didattico per i laboratori del corso di Modelli Statistici I*", (80 pagine scaricabili da: <http://www.isib.cnr.it/~brazzale/ModStatI/>): è un testo di carattere applicativo (si tratta di 12 lezioni utilizzate per un laboratorio di R), con ricche esemplificazioni e con files dei dati scaricabili per riprodurre gli esempi riportati negli appunti. Gli argomenti trattati sono relativi alla regressione lineare semplice e multipla, l'analisi della varianza e della covarianza.

Di un certo interesse, anche se non organizzati in forma di manuale, sono gli appunti per un corso introduttivo all'uso dell'ambiente statistico **R** preparati da Stefano Iacus e consultabili al seguente indirizzo Internet: <http://www.economia.unimi.it/iacus/corsoR/>. Si tratta di tre lezioni, in forma estremamente schematica ed efficace, sui fondamenti di **R**, sugli strumenti statistici di base e su alcune tecniche avanzate. È un materiale decisamente pratico e l'utente, attraverso la funzione copia e incolla, può sperimentare direttamente i comandi e le istruzioni riportati nelle lezioni nell'ambiente **R**. Può essere un utile strumento da usare come guida dopo aver preso una certa familiarità a livello teorico con il software **R**.

Per chi vuole affrontare l'analisi delle serie temporali usando le funzionalità messe a disposizione di **R** si consiglia la lettura di "*Analisi delle serie storiche con R*" (64 pagine) realizzato da chi scrive e consultabile al seguente indirizzo web: <http://cran.r-project.org/doc/contrib/Ricci-ts-italian.pdf>. Il manuale cerca di fornire una trattazione delle principali tematiche dell'analisi delle serie storiche (decomposizione, test di specificazione, grafici, processi stocastici) attraverso delle esemplificazioni pratiche.

c) Risorse in lingua inglese

In lingua inglese esiste una più vasta letteratura sull'ambiente **R**. Accenneremo ad alcune risorse disponibili gratuitamente sul Web rinviando alla bibliografia riportata a conclusione dell'articolo per i riferimenti a testi cartacei.

"*R for beginners*" di Emmanuel Paradis (31 pagine, scaricabili dall'URL: http://cran.r-project.org/doc/contrib/rdebuts_en.pdf), come lascia supporre il titolo, è una dispensa abbastanza elementare diretta ad utenti poco esperti che vogliono avvicinarsi al software **R**, è a carattere meramente introduttivo. Sicuramente di livello più avanzato (per i quali si presuppone una certa padronanza e dimestichezza con **R**, oltre che, in alcuni casi, una discreta conoscenza di alcune specifiche metodologie statistiche) ed anche con ricche esemplificazioni dell'uso del programma in analisi statistiche sono i tre seguenti manuali:

- John Maindonald, "*Using R for Data Analysis and Graphics, An Introduction*" (112 pagine, scaricabili dall'URL:

<http://cran.r-project.org/doc/contrib/usingR.pdf>) il quale, dopo un breve riferimento alle caratteristiche di **R**, va direttamente alla trattazione di analisi della regressione, modelli lineari generalizzati, analisi della varianza, grafici. Alla fine di ciascun capitolo sono proposti degli esercizi. Assieme al materiale didattico si possono scaricare anche i files contenenti tutti i codici esemplificativi del programma usati nel testo e i files con i dati impiegati nelle esercitazioni.

- John Verzani, "*SimpleR: Using R for Introductory Statistics*", (114 pagine, scaricabili dall'URL: <http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf>): è un testo eminentemente applicativo che presuppone la conoscenza dell'ambiente **R**. Si occupa delle principali tecniche di analisi della statistica descrittiva e di quella inferenziale con esempi concreti (i cui dati sono scaricabili e disponibili in apposito file) riproducibili dall'utente in fase di esercitazione pratica.

- Julian Faraway, "*Practical Regression and Anova using R*", (213 pagine, scaricabili dall'URL: <http://www.stat.lsa.umich.edu/~faraway/book/pra.pdf>). E' un manuale di livello molto avanzato, senza alcuna introduzione al software (salvo un brevissimo accenno in appendice), che si rivolge ad utenti esperti con una buona competenza in materia di statistica inferenziale e multivariata. Come suggerisce il titolo della dispensa, l'autore si sofferma dettagliatamente sull'uso di **R** nella soluzione dei problemi relativi alla regressione e ai modelli lineari (stima, inferenza, diagnostica, trasformazioni di variabili e cambiamento di scala, scelta delle variabili, componenti principali e collinearità, analisi della varianza e della covarianza, metodo dei minimi quadrati generalizzati) e unitamente alla dispensa è possibile scaricare un apposito package, messo a punto dallo stesso autore, indispensabile per la comprensione del contenuto del manuale e che consente all'utente di riprodurre concretamente su proprio PC gli esempi riportati nel testo.

- P.M.E. Altham, "*Introduction to Generalized Linear Modelling in R*" (pagine 54 scaricabili da: <http://www.statslab.cam.ac.uk/~pat/>) è un testo che si occupa di argomenti statistici decisamente avanzati che possono essere trattati e gestiti nell'ambiente **R**, come l'analisi della varianza twoway e le regressioni logistica, binomiale e di Poisson. Il testo è corredato di files di dati relativi alle applicazioni pratiche richiamate nello stesso.

- Oleg Nenadic, Walter Zucchini, "*Statistical Analysis with R, a quick start*" (47 pagine scaricabili da: http://www.statoek.wiso.uni-goettingen.de/mitarbeiter/ogi/pub/r_workshop.pdf) dopo una breve introduzione all'ambiente **R** tratta di argomenti come i modelli lineari, l'analisi delle serie storiche, la grafica avanzata.

Altre due risorse si occupano dell'utilizzo di **R** in particolari ambiti statistici come l'analisi di dati categoriali: "*An*

Introduction to Categorical Data Analysis Using R di Brett Presnell (38 pagine, scaricabili dall'URL: <http://web.stat.ufl.edu/~presnell/Teaching/sta4504-2000sp/R/R-CDA.pdf>) con file di esempio delle analisi statistiche riportate nella dispensa; oppure nel campo degli esperimenti e dei questionari in psicologia: "*Notes on the use of R for psychology experiments and questionnaires*" di Jonathan Baron e Yuelin Li (39 pagine, scaricabili dall'URL: <http://cran.r-project.org/doc/contrib/Baron-rpsych.pdf>).

Forniamo qui di seguito tre indirizzi Internet che possono risultare utili in relazione alla ricerca e al reperimento di ulteriori risorse sull'ambiente **R**:

- a) <http://www.vanderbilt.edu/quantmetheval/r.htm>: questa pagina contiene una serie di link ad utili risorse e materiale disponibili in rete in lingua inglese
- b) <http://cran.r-project.org/other-docs.html>: da questo indirizzo è possibile scaricare altro materiale didattico per l'apprendimento e lo studio di **R** in lingua inglese
- c) <http://www.ku.edu/~pauljohn/R/statsRus.html>: risposte brevi ad una pluralità di quesiti e problemi relativi ad **R**

5. R e la statistica ambientale

In questo paragrafo si vogliono presentare brevemente alcuni packages di **R** di particolare interesse per la statistica ambientale. Si può consultare la seguente URL: <http://agec221.agecon.uiuc.edu/csiss/Rgeo/> relativa ai progetti di analisi di dati spaziali con **R** per avere maggiori e più dettagliate informazioni.

Per quanto riguarda le applicazioni della geostatistica si possono considerare:

geoR: fornisce una serie di funzioni per l'analisi di dati geostatistici¹⁶

geoRglm: modelli spaziali lineari generalizzati¹⁷

gstat: modelli geostatistici univariati e multivariati, previsioni e simulazioni¹⁸

GRASS: interfaccia tra GRASS 5.0 geographical information system ed il software R

¹⁶ Si veda l'articolo: Paulo J. Ribeiro, Jr. and Peter J. Diggle. *geoR*: A package for geostatistical analysis. *R News*, 1(2):15-18, June 2001 consultabile all'URL: <http://agec221.agecon.uiuc.edu/csiss/Rgeo/rnews1.2.15-18.pdf>

¹⁷ Si veda l'articolo: Ole F. Christensen and Paulo J. Ribeiro. *georglm*: A package for generalised linear spatial models. *R News*, 2(2):26-28, June 2002. consultabile all'URL: <http://agec221.agecon.uiuc.edu/csiss/Rgeo/rnews2.2.26-28.pdf>

¹⁸ Si veda il sito: <http://www.gstat.org/>

spatial: funzioni per kriging e point pattern analysis

spatstat: spatial Point Pattern analysis, stima del modello e simulazione¹⁹

splancs: Spatial and Space-Time Point Pattern Analysis²⁰

sgeostat: un object-oriented framework per i modelli geostatistici

vardiag: un package per la diagnostica dei variogrammi

Relativamente alla problematica dei valori estremi è molto valido il package *evd* nel quale sono disponibili una serie di funzioni relative alle seguenti distribuzioni :

Distribuzione del minimo e del massimo

Distribuzione di Frechet

Generalized Extreme Value Distribution

Generalized Pareto Distribution

Distribuzione di Gumbel

Distribuzione di statistiche ordinali

con stima di massima verosimiglianza dei parametri;

Per l'utilizzo del metodo CART (*Classification and Regression Tree*) e degli alberi di regressione generalizzati nello studio dei fenomeni ambientali possono essere utili i packages:

tree: Classification and regression trees

rpart: Recursive partitioning and regression trees

¹⁹ Si veda il sito: <http://www.maths.uwa.edu.au/~adrian/spatstat.html>

²⁰ Si veda il sito: <http://www.maths.lancs.ac.uk/~rowlings/Splancs/>

Si riportano di seguito alcuni comandi del software **R** che fanno riferimento ad alcuni argomenti trattati dal prof. Lamberto Soliani nelle dispense per il corso di *"Statistica univariata e bivariata parametrica e non-parametrica nelle discipline ambientali e biologiche"*. Se non è indicato espressamente i comandi fanno parte del package *stats*. Per richiamare l'help on line digitare: ? nomecomando.

Tabelle di contingenza

table(): tabella di contingenza

Rappresentazioni grafiche di distribuzioni univariate

plot(): diagramma di dispersione

boxplot(): box-and-whisker plot

hist(): istogramma

Le misure di tendenza centrale

mean(): media aritmetica

median(): mediana

quantile(): quantili

summary(): vari indicatori di una distribuzione

fivenum(): Tukey Five-Number Summaries

Misure di dispersione o variabilità

sd(): standard deviation

var(): varianza

Misure di asimmetria e curtosi

kurtosis(): indice di curtosi (package e1071 oppure package fBasics)

skewness(): indice di simmetria (package e1071 oppure package fBasics)

Le seguenti funzioni sono definite dall'utente:

Misure di asimmetria

```
gamma1<-function(x) {  
  m=mean(x)  
  n=length(x)  
  s=sqrt(var(x))  
  m3=sum((x-m)^3)/n  
  g1=m3/(s^3)  
  return(g1)  
}
```

```
skewness2<-function(x){  
  m=mean(x)  
  me=median(x)  
  s=sqrt(var(x))  
  sk=(m-me)/s  
  return(sk)  
}
```

```
bowley<-function(x){  
  q<-as.vector(quantile(x,prob=c(.25,.50,.75)))  
  b=(q[3]+q[1]-2*q[2])/(q[3]-q[2])  
  return(b)  
}
```

```
b3<-function(x){  
  m=mean(x)  
  me=median(x)  
  n=length(x)
```

```
d=sum(abs(x-me))/n
```

```
b=(m-me)/d
```

```
return(b)
```

```
}
```

Misure di curtosi

```
gamma2<- function(x) {
```

```
m<-mean(x)
```

```
s<-var(x)^2
```

```
y<-(x-m)^4
```

```
m4<-sum(y)
```

```
n<-length(x)
```

```
g<-(m4/(s*n))-3
```

```
return(g)}
```

Distribuzioni di probabilità discrete e continue

Beta: The Beta Distribution

Binomial: The Binomial Distribution

Cauchy: The Cauchy Distribution

Chisquare: The (non-central) Chi-Squared Distribution

Exponential: The Exponential Distribution

FDist: The F Distribution

GammaDist: The Gamma Distribution

Geometric: The Geometric Distribution

Hypergeometric: The Hypergeometric Distribution

Logistic: The Logistic Distribution

Lognormal: The Log Normal Distribution

Multinomial: The Multinomial Distribution

NegBinomial: The Negative Binomial Distribution

Normal: The Normal Distribution

Poisson: The Poisson Distribution

SignRank: Distribution of the Wilcoxon Signed Rank Statistic

TDist: The Student t Distribution

Tukey: The Studentized Range Distribution

Uniform: The Uniform Distribution

Weibull: The Weibull Distribution

Wilcoxon: Distribution of the Wilcoxon Rank Sum Statistic

Test statistici parametrici e non parametrici

ansari.test(): Ansari-Bradley Test

bartlett.test(): Bartlett Test for Homogeneity of Variances

binom.test(): Exact Binomial Test

chisq.test(): Pearson's Chi-squared Test for Count Data

cor.test(): Test for Association/Correlation Between Paired Samples

fisher.test(): Fisher's Exact Test for Count Data

friedman.test(): Friedman Rank Sum Test

kruskal.test(): Kruskal-Wallis Rank Sum Test

ks.test(): Kolmogorov-Smirnov Tests

levene.test(): Levene Test (package car)

mantelhaen.test(): Cochran-Mantel-Haenszel Chi-Squared Test for Count Data

mcnemar.test(): McNemar's Chi-squared Test for Count Data

mood.test(): Mood Two-Sample Test of Scale

oneway.test(): Test for Equal Means in a One-Way Layout

pairwise.prop.test(): Pairwise comparisons of proportions

pairwise.t.test(): Pairwise t tests

pairwise.wilcox.test(): Pairwise Wilcoxon rank sum tests

power.anova.test(): Power calculations for balanced one-way analysis of variance tests

power.prop.test(): Power calculations two sample test for of proportions

`power.t.test()`: Power calculations for one and two sample t

`prop.test()`: Test for Equal or Given Proportions

`quade.test()`: Quade Test

`t.test()`: Student's t-Test

`var.test()`: F Test to Compare Two Variances

`wilcox.test()`: Wilcoxon Rank Sum and Signed Rank Tests

Analisi della varianza

`aov()`: Fit an Analysis of Variance Model

`power.anova.test()`: Power calculations for balanced one-way analysis of variance tests

`TukeyHSD()` : Compute Tukey Honest Significant Differences

`contrasts()` : Get and Set Contrast Matrices

Test di normalità

`ad.test()`: Anderson-Darling test for normality (package `nortest`)

`cvm.test()`: Cramer-von Mises test for normality (package `nortest`)

`lillie.test()`: Lilliefors (Kolmogorov-Smirnov) test for normality (package `nortest`)

`pearson.test()`: Pearson chi-square test for normality (package `nortest`)

`sf.test()`: Shapiro-Francia test for normality (package `nortest`)

`shapiro.test()`: Shapiro-Wilk Normality Test

`jarque.bera.test()` : Jarque-Bera Test for Normality (package `tseries`)

`qqnorm()`: normal QQ plot

Regressione lineare semplice e multipla

`lm()`: Fitting Linear Models

`coef()`: Extract Model Coefficients

`fitted()`: Extract Model Fitted Values

`predict()`: Model Predictions

`anova()`: ANOVA tables

lm.influence(): Regression Diagnostics

confint() : Confidence Intervals for Model Parameters

bptest(): Breusch-Pagan Test

dwtest(): Durbin-Watson Test

Correlazione

cor(): coefficiente/matrice di correlazione

cor.test(): Test for Association/Correlation Between Paired Samples

6. Conclusioni

In questo breve articolo introduttivo si è cercato di presentare l'ambiente **R**, utile e indispensabile ausilio informatico per l'analisi statistica dei dati, che ha adottato la filosofia dell'*opensource*. Esso è completamente gratuito e scaricabile da Internet e nel momento attuale è il software maggiormente utilizzato dai ricercatori in campo statistico a livello mondiale. Esso risulta essere uno strumento assai versatile e flessibile con una pluralità di vantaggi, caratteristiche positive e punti di forza che sono stati illustrati nei paragrafi precedenti, così come sono state esaminate le sue principali funzionalità e i campi di applicazione. È stata fatta una recensione, con alcune brevi indicazioni, delle diverse risorse (manuali, appunti, dispense) di pubblico dominio su Internet che possono facilitare l'apprendimento dell'uso di **R** per gli utenti poco esperti o l'approfondimento per gli utenti con maggiore dimestichezza.

Si spera, con la realizzazione di questo lavoro, di poter contribuire alla diffusione dell'ambiente **R** anche tra i ricercatori statistici italiani e di suscitare l'attenzione e la curiosità intorno a questo innovativo programma e alla filosofia che ne è alla base in modo da invogliarne la conoscenza e l'utilizzo.

BIBLIOGRAFIA

Per un quadro completo, dettagliato ed esaustivo consultare la bibliografia del CRAN (The Comprehensive R Archive Network) all'indirizzo Internet: <http://www.r-project.org/doc/bib/R-publications.html>

Paola Bortot, Laura Ventura e Alessandra Salvan, *Inferenza Statistica: Applicazioni con S-PLUS e R*, Cedam, 2002 [a questo indirizzo internet: <http://www.stat.unipd.it/~ventura/libros.htm>, consultato in data 18/02/04, è disponibile l'indice, la premessa, l'errata corrige del libro, nonché i codici *S-PLUS* e *R* di tutti i comandi utilizzati nel volume]

Peter Dalgaard. *Introductory Statistics with R*. Springer, 2002

Ross Ihaka and Robert Gentleman. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3):299-314, 1996.

Stefano Iacus, Guido Masarotto. *Laboratorio di statistica con R*. McGraw-Hill, Milano, 2003; al seguente indirizzo Internet è possibile consultare la prefazione e l'indice del libro:

http://www.ateneonline.it/LibroScien.asp?item_id=1436 [consultato in data 28/01/04]

Diego Kuonen and Valerie Chavez. R - un exemple du succès des modèles libres. *Flash Informatique*, 2:3-7, 2001.

[consultato in data 12/12/03 su Internet al seguente indirizzo: <http://sawwww.epfl.ch/SIC/SA/publications/FI01/fi-2-1/2-1-page3.html>]

Jeff Racine, Rob J. Hyndman, Using R to Teach Econometrics, Working Paper 10/2001, Department of Econometrics & Business Statistics, Monash University, Melbourne, novembre 2001 [consultato in data 15/01/04 su Internet al seguente indirizzo: <http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/2001/wp10-01.pdf>]

Paulo J. Ribeiro, Jr. and Patrick E. Brown. Some words on the R project. *The ISBA Bulletin*, 8(1):12-16, March 2001.

Brian D. Ripley. The R project in statistical computing. *MSOR Connections. The newsletter of the LTSN Maths, Stats & OR Network.*, 1(1):23-25, February 2001. [consultato in data 12/12/03 su Internet al seguente indirizzo: <http://ltsn.mathstore.ac.uk/newsletter/feb2001/pdf/rproject.pdf>]

William N. Venables and Brian D. Ripley. *Modern Applied Statistics with S. Fourth Edition* Springer, 2002

William N. Venables and Brian D. Ripley. *S Programming*. Springer, 2000

Manuali su R

Pagina del sito del The Comprehensive R Archive Network (CRAN) da cui si può accedere ad una vasta scelta della manualistica ufficiale su R in formato PDF:

[consultata in data 29/11/04]

<http://cran.r-project.org/manuals.html>

R Development Core Team, An Introduction to R, versione 2.0.1 del 15/11/04:

<http://cran.r-project.org/doc/manuals/R-intro.pdf>

R Development Core Team, The R language definition, versione 2.0.1 del 15/11/04:

<http://cran.r-project.org/doc/manuals/R-lang.pdf>

R Development Core Team, Writing R Extentions, versione 2.0.1 del 15/11/04:

<http://cran.r-project.org/doc/manuals/R-exts.pdf>

R Development Core Team, R Data Import/Export, versione 2.0.1 del 15/11/04:

<http://cran.r-project.org/doc/manuals/R-data.pdf>

R Development Core Team, R Installation and Administration, versione 2.0.1 del 15/11/04:

<http://cran.r-project.org/doc/manuals/R-admin.pdf>

R Development Core Team, The R Reference Index, versione 2.0.1 del 15/11/04:

<http://cran.r-project.org/doc/manuals/fullrefman.pdf>

Pagina delle FAQ (Frequently Asked Questions) su R:

<http://cran.r-project.org/faqs.html>

Altre risorse su R disponibili in Internet

Italiano

Claudio Agostinelli, Introduzione ad R, versione 0.3, ottobre 2000 (in formato PDF e PostScript)

<http://www.dst.unive.it/~claudio/R/index.html#manuale> [consultato in data 15/01/04]

Roberto Baggiani, Introduzione ad R, versione 6.0, 24 ottobre 2004

<http://digilander.libero.it/robicox/manuali/pdf/mainr.pdf> (in formato PDF)

[consultato in data 26/10/04]

A. Brazzale, M. Chiogna, C. Gaetan e N. Sartori, Laboratorio di R, Materiale didattico per i laboratori del corso di Modelli Statistici I, A.A. 2001/2002, ISIB-CNR, Padova (in formato PDF e PS)

<http://www.isib.cnr.it/~brazzale/ModStat/> [consultato in data 20/05/04]

Stefano Iacus, Corso introduttivo all'uso dell'ambiente statistico R, Dipartimento di Economia Politica e Aziendale, Università di Milano marzo 2003 (in formato HTML)

<http://www.economia.unimi.it/iacus/corsoR/> [consultato in data 18/02/04]

Angelo M. Mineo, Una guida all'utilizzo dell'ambiente statistico R, 2003 (in formato PDF)

<http://cran.r-project.org/doc/contrib/Mineo-dispensaR.pdf> [consultato in data 15/01/04]

Vito M. R. Muggeo, Il linguaggio R: concetti introduttivi ed esempi, versione 1.0, giugno 2002 (in formato PDF)

<http://cran.r-project.org/doc/contrib/nozioniR.pdf> [consultato in data 15/01/04]

Francesca Parpinel, La statistica applicata attraverso il programma R, febbraio 2000 (in formato PDF)

http://venus.unive.it/statcomp/r/man_Parpinel.pdf [consultato in data 15/01/04]

Lea Petrella, Silvia Poletti, Laboratorio di Statistica I, metodi matematici e statistici, Dipartimento di Matematica, Università Roma 3, 2002 (in formato PDF e PostScript)

http://www.mat.uniroma3.it/didatticacds/corsi/didattica_interattiva/aa_01_02/st1/st1.html [consultato in data 15/01/04]

Alessio Pollice, Esercitazioni con R, Dipartimento di Scienze Statistiche, Università di Bari (in formato PDF)

<http://www.dip-statistica.uniba.it/html/docenti/pollice/materiale.htm> [consultato in data 25/10/04]

Vito Ricci, Analisi delle serie storiche con R, novembre 2004 (in formato PDF)

<http://cran.r-project.org/doc/contrib/Ricci-ts-italian.pdf> [consultato in data 30/11/04]

Luca Scrucca, Note sul linguaggio e ambiente statistico R, Dipartimento di Scienze Statistiche, Università degli Studi di Perugia, 18 ottobre 2004 (in formato PDF)

<http://www.stat.unipg.it/~luca/R-note.pdf> [consultato in data 26/10/04]

Inglese

P.M.E. Altham, Introduction to Generalized Linear Modelling in R, Statistical laboratory, University of Cambridge, 1 marzo 2004

In formato PDF: <http://www.statslab.cam.ac.uk/~pat/redwsheets.pdf>

In formato PS: <http://www.statslab.cam.ac.uk/~pat/redwsheets.ps>

File di dati di esempio usati nel manuale: <http://www.statslab.cam.ac.uk/~pat/R.bigdata>

[consultato in data 20/05/04]

Jonathan Baron, Yuelin Li, Notes on the use of R for psychology experiments and questionnaires, agosto 2003 (in formato PDF)

<http://cran.r-project.org/doc/contrib/Baron-rpsych.pdf> [consultato in data 15/01/04]

Julian Faraway, Practical Regression and Anova using R, luglio 2002, (in formato PDF)

<http://www.stat.lsa.umich.edu/~faraway/book/practical.pdf> [consultato in data 15/01/04]

John Maindonald, Using R for Data Analysis and Graphics, An Introduction, 2004 (in formato PDF)

<http://cran.r-project.org/doc/contrib/usingR.pdf> [consultato in data 25/10/04]

Oleg Nenadic, Walter Zucchini, Statistical Analysis with R, a quick start, settembre 2004 (in formato PDF)

http://www.stat.oek.wiso.uni-goettingen.de/mitarbeiter/ogi/pub/r_workshop.pdf [consultato in data 24/11/04]

Emmanuel Paradis, R for beginners, agosto 2002 (in formato PDF)

http://cran.r-project.org/doc/contrib/rdebuts_en.pdf [consultato in data 15/01/04]

Brett Presnell, An Introduction to Categorical Data Analysis Using R, marzo 2000, (in formato PDF)

<http://web.stat.ufl.edu/~presnell/Teaching/sta4504-2000sp/R/R-CDA.pdf> [consultato in data 15/01/04]

Vito Ricci, Time series reference card, novembre 2004, (in formato PDF)

<http://cran.r-project.org/doc/contrib/Ricci-refcard-ts.pdf> [consultato in data 26/11/04]

John Verzani, SimpleR: Using R for Introductory Statistics, versione 0.4, agosto 2002 (in formato PDF)

<http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf> [consultato in data 15/01/04]

<http://www.vanderbilt.edu/quantmetheval/r.htm>

Questa pagina contiene una serie di link ad utili risorse e materiale disponibili in rete in lingua inglese

<http://cran.r-project.org/other-docs.html>

Da questo indirizzo è possibile scaricare altro materiale didattico per l'apprendimento e lo studio di **R** in lingua Inglese

<http://www.ku.edu/~pauljohn/R/statsRus.html>

Risposte brevi ad una pluralità di quesiti e problemi relativi ad **R**