

Erasmus Vassallo

**Laboratorio di
Statistica Economica
con R**



Copyright © MMIX
ARACNE editrice S.r.l.

www.aracneeditrice.it
info@aracneeditrice.it

via Raffaele Garofalo, 133 A/B
00173 Roma
(06) 93781065

ISBN 978-88-548-2875-9

*I diritti di traduzione, di memorizzazione elettronica,
di riproduzione e di adattamento anche parziale,
con qualsiasi mezzo, sono riservati per tutti i Paesi.*

*Non sono assolutamente consentite le fotocopie
senza il permesso scritto dell'Editore.*

I edizione: novembre 2009

*A mia moglie Anna e
a mio figlio Gabriele*

*“Spesso ci si attacca ai numeri
come gli ubriachi si attaccano
ai lampioni, non per farsi illu-
minare ma per farsi sostenere”.*

George Bernard Shaw

Indice

<i>Prefazione</i>	11
<i>1. Nozioni di R</i>	17
<i>2. Aggregati di contabilità nazionale</i>	29
<i>3. Numeri indice</i>	41
<i>4. Serie storiche in un approccio classico</i>	55
<i>5. Specificazione di modelli lineari</i>	93
<i>6. Analisi delle serie storiche in un contesto stocastico</i>	129
<i>7. Povertà e concentrazione del reddito</i>	167
<i>8. Produzione e produttività</i>	179
<i>9. Efficienza statistica</i>	193
<i>10. Carte di controllo</i>	215
<i>11. Indici di bilancio</i>	233
<i>12. Shift-share, tavola input-output ed altri metodi</i>	243

Prefazione

Questo libro costituisce materiale di supporto ed integrazione alle lezioni di LABORATORIO DI STATISTICA ECONOMICA svolte nell'ambito del Corso di Laurea in STATISTICA ED INFORMATICA PER LA GESTIONE E L'ANALISI DEI DATI (SIGAD) della Facoltà di Economia dell'Università degli Studi di Palermo.

La ragione di questo progetto editoriale è nell'assenza di un testo dedicato da offrire agli studenti. Ciò, naturalmente, non esclude che il volume possa risultare utile anche in contesti didattici più generali quale supporto per un percorso applicato di alcuni classici temi economici ben aldilà dello specifico corso per il quale è stato progettato.

Nell'offerta formativa dell'Ateneo di Palermo, il LABORATORIO DI STATISTICA ECONOMICA è stato pensato già nel 2002 dal Prof. Gaetano Cusimano in una chiave originale e con una specificità di forma e contenuto non presente in corsi simili di altre sedi universitarie. Il sottoscritto, che ha avuto l'onore di subentrargli nelle lezioni, ha tentato di mantenere inalterato lo spirito originario pur non trascurando i necessari aggiornamenti.

In particolare, nel piano formativo della Facoltà di Palermo, il LABORATORIO è posizionato al terzo ed ultimo anno e segue la STATISTICA ECONOMICA posta al secondo anno; di essa, il LABORATORIO riprende argomenti e temi affrontandoli con applicazioni e simulazioni tramite software statistico.

Questo volume è aderente al corso di riferimento e, pertanto, privilegia un approccio empirico e pragmatico volto alla presentazione, analisi e soluzione del problema economico, omettendo volutamente quelle considerazioni teoriche e metodologiche che il lettore dovrebbe

conoscere dallo studio propedeutico della STATISTICA e della STATISTICA ECONOMICA.

Con questo difficile obiettivo, nella definizione dei contenuti del testo si sono enfatizzate le caratteristiche tipiche di una disciplina di confine qual è per appunto la STATISTICA ECONOMICA, in stretto contatto con l'economia, l'econometria e, naturalmente, la statistica; a tal proposito, è interessante ricordare la declaratoria prevista dal D.M. del 4-10-2000: "Il settore si incentra sulla misurazione ed analisi statistica dei fenomeni economici latamente intese: dalla misura di grandezze e di sistemi di grandezze economiche (contabilità nazionale), all'analisi della dinamica e alle previsioni economiche, alla stima e verifica di modelli di comportamenti economici, alla valutazione di politiche. Elaborazione di sistemi e modelli di riferimento, progettazione e gestione di sistemi di dati e indicatori economici, sviluppo e impiego di appropriati metodi statistici per lo studio empirico-quantitativo del comportamento economico, in chiave sezionale, spaziale e temporale costituiscono elementi fondanti del settore, ai vari livelli (dal micro al macroeconomico). Specifica attenzione viene riservata anche alle analisi di mercato, alla gestione e alle decisioni aziendali, con particolare riguardo al controllo statistico e alla valutazione della qualità dei prodotti e dei servizi".

Già nel 1943 Libero Lenti definisce la STATISTICA ECONOMICA quale analisi dei fenomeni economici con il metodo quantitativo. Benedetto Barbieri nel 1958 la considera un ramo della statistica applicata che tratta della rilevazione ed analisi quantitativa dei fenomeni economici. Antonino Giannone nel 1973 la presenta come un complesso di tecniche speciali di natura logico-matematica impiegate per lo studio dei fenomeni economici collettivi al fine di scoprire leggi od uniformità che li governano. Renato Guarini nel 1982 considera la STATISTICA ECONOMICA come osservazione e misurazione dei fenomeni economici, classificazione, ricerca e verifica di leggi e relazioni esistenti tra i fenomeni stessi. Giuseppe Alvaro nel 1999 si sofferma sulla capacità della disciplina di individuare, definire, rilevare, misurare ed analizzare le manifestazioni dei fenomeni economici nella loro continuità temporale ed articolazione territoriale.

Tuttavia, la definizione che maggiormente piace ricordare è quella di Gaetano Cusimano con cui il sottoscritto ha avuto il privilegio di

lavorare per circa 15 anni. Egli scriveva nel 1984 che la STATISTICA ECONOMICA è lo studio dei fenomeni economici collettivi mediante l'impiego di convenienti metodi e modelli statistici costruiti per soddisfare le esigenze specifiche di ciascun problema: è descrizione dei fenomeni economici, ricerca delle leggi che ne regolano il comportamento e delle relazioni che intercorrono fra i fenomeni economici stessi.

La STATISTICA ECONOMICA, dunque, analizza fenomeni economici collettivi sia in un approccio microeconomico (basato sul comportamento di singole unità economiche considerate nel loro insieme, per esempio la produzione delle imprese, la nati-mortalità imprenditoriale, il comportamento dei lavoratori disoccupati, etc.), sia in un approccio macroeconomico (si studiano i fenomeni economici aggregati che riguardano un sistema economico nel suo complesso, per esempio tassi di cambio, inflazione, crescita dell'economia, etc.).

Ecco, allora, che se risulta ammissibile (e certamente lo è) specificare in tal modo i contenuti della STATISTICA ECONOMICA, il LABORATORIO, e quindi il testo corrispondente, risulta naturalmente definito nei suoi contenuti con un marcato accento pratico e con un utilizzo esteso dell'elaboratore. Con tale scopo, il volume fa un ampio ricorso a **R** che oggi, tra i tanti software disponibili, è quello che più rapidamente si sta diffondendo nella comunità degli statistici.

Il volume è diviso in 13 capitoli la cui sequenza segue un criterio generale da argomenti meno strutturati verso argomenti più complessi o che richiedono un utilizzo più intensivo del software. Il contenuto di questi 13 capitoli ed i temi trattati sono presentati di seguito in dettaglio.

Il capitolo 1 si sofferma su alcune nozioni di base di **R** introducendo il software al lettore con minori competenze. In particolare, sono descritte le funzioni di lettura dei dati ed i comandi più significativi per la manipolazione degli oggetti.

Il capitolo 2 elenca i conti principali della Contabilità Nazionale e svolge alcuni esempi di calcolo con riferimento agli aggregati considerati di maggiore rilievo.

Il capitolo 3 si sofferma sui Numeri Indice con specifica attenzione a quelli di prezzo e su alcune tecniche di deflazione. Con **R** vengono calcolate le serie a prezzi costanti e verificate le differenze tra valuta-

zioni a prezzi correnti e prezzi costanti, nonché le modalità e le conseguenze dei cambi di base.

Il capitolo 4 descrive l'analisi delle serie storiche in un contesto classico con un approccio generalmente ma non esclusivamente macroeconomico. Viene quindi descritto come scomporre la serie di dati nelle principali componenti di trend, ciclo e stagionalità tramite diverse tecniche. Gli esempi evidenziano come **R** possa essere adoperato per smussare i dati tramite medie mobili di ordine dispari ovvero di ordine pari nel caso di lag stagionale pari; ma il software è utile anche per calcolare più complesse funzioni di smussamento esponenziale per serie con o senza periodo stagionale. Inoltre, un primo riferimento alle stime di modelli lineari permette di individuare trend e ciclo con opportuna specificazione modellistica, così come la componente stagionale è ottenibile tramite variabili dummy opportunamente definite. Il maggiore interesse verso la componente di trend offre lo spunto per introdurre le specificazioni tipiche dei modelli autoregressivi e dei modelli a ritardi distribuiti. Tuttavia, l'introduzione dei modelli lineari e delle loro stime all'interno del capitolo necessita alcune sottolineature rinviate alla sezione successiva.

Il capitolo 5 tratta dunque gli aspetti statistici, econometrici ed interpretativi della specificazione lineare con particolare riferimento ai limiti ed alle cautele che si rendono necessarie. Si affronta la mal-specificazione dei modelli con omissione di variabili rilevanti ed inclusione di variabili irrilevanti, la presenza di multicollinearità piuttosto che autocorrelazione dei residui, ovvero l'assenza di stabilità strutturale dei parametri o di omoschedasticità, per citare solo alcuni dei contenuti trattati nel capitolo per mezzo degli esempi proposti. Tuttavia, una migliore specificazione del modello, utile soprattutto a fini previsivi, potrebbe richiedere una formulazione più complessa che ipotizzi un andamento stocastico della serie.

Il capitolo 6, in particolare, segue l'approccio moderno assumendo che una serie storica venga generata da un processo stocastico e possa essere rappresentata da un modello che includa una componente autoregressiva e/o una componente in media mobile. A tal proposito, con **R** sviluppiamo esempi che stimano questi modelli e ne analizzano le capacità descrittive e previsive.

Il capitolo 7 è dedicato alle misure di povertà e di distribuzione del reddito. Tramite il software si calcolano i più noti indici di diffusione e di intensità della povertà; inoltre, alcuni esercizi propongono il calcolo del rapporto e dell'indice di concentrazione di Gini, Theil, Sen ed altri ancora.

Il capitolo 8 si sofferma sui concetti di produzione e di produttività con specifico riferimento alla funzione di produzione Cobb-Douglas di cui vengono analizzate alcune stime.

Il capitolo 9 è dedicato alla misura dell'efficienza tramite DRIFTED OLS (DOLS) e DATA ENVELOPMENT ANALYSIS (DEA). La complessità dell'algoritmo richiede in **R**, unico caso nel testo, una specifica libreria esterna (LPSOLVE) finalizzata alla risoluzione dei problemi di programmazione lineare. Le stime di efficienza sono quindi ottenute secondo il modello DEA tradizionale INPUT-ORIENTED ed OUTPUT-ORIENTED sia con rendimenti di scala costanti che variabili.

Il capitolo 10 tratta le principali carte di controllo Shewhart per attributi e per variabili con limiti 3-SIGMA utili all'analisi dell'evoluzione di un processo produttivo alla ricerca di eventuali anomalie.

Il capitolo 11, a partire dallo Stato Patrimoniale e dal Conto Economico di un'impresa, calcola i principali indici di bilancio di interesse per tracciare un quadro sintetico circa le condizioni di salute di un'azienda.

Il capitolo 12 presenta alcune tecniche utili per analizzare dati localizzati in un territorio. La SHIFT-SHARE consente di scomporre le variazioni osservate di un aggregato in un effetto nazionale, uno di mix industriale ed un terzo dovuto a fattori locali. Ma se il territorio ha un ruolo, gli indici di autocorrelazione spaziale evidenziano le influenze tra aree vicine. Inoltre, anche le classiche tecniche di raggruppamento così come la sintesi di variabili in fattori possono apparire particolarmente utili in un contesto di analisi territoriale. Infine, la TAVOLA INPUT-OUTPUT consente di analizzare le caratteristiche della struttura economica dell'area in esame, studiare gli impatti di una variazione dell'equilibrio tra domanda ed offerta e, per questa via, confrontare territori diversi.

Il capitolo 13, in ultimo, mostra come collegare **R** a funzioni esterne scritte in linguaggio C/C++ poichè in questo linguaggio sono spes-

so disponibili routine non ancora direttamente implementate nel software statistico.

Non è comunque necessaria una lettura ordinata dei capitoli, anche se per ovvie ragioni espositive le sezioni successive del testo non replicano le descrizioni delle funzioni di **R** presentate nelle parti precedenti del volume.

Ai 13 capitoli è inoltre aggiunta una sezione on-line con indirizzo

<http://www.erasmo.vassallo.unipa.it/laboratorio>

In essa il lettore troverà approfondimenti, capitoli aggiuntivi, dati, errata corrige e quant'altro opportuno per una migliore fruizione del testo.

E' evidente che i temi e le applicazioni trattati nel volume forniscono un quadro generale, semplificato e non esaustivo dei principali argomenti della STATISTICA ECONOMICA, ma è anche vero che questi coprono gran parte delle esigenze formative di base previste dalla disciplina. Non sappiamo se il risultato ottenuto sia o meno soddisfacente; sarà il giudizio del lettore a stabilire l'utilità di questo testo. A chi scrive non resta che ringraziare quanti vorranno segnalare errori, sviste e refusi presenti nel volume.

Prima di chiudere questa Prefazione, si deve rivolgere un saluto speciale al Prof. Gaetano Cusimano, maestro ed amico, che ci ha lasciato improvvisamente il 27 ottobre 2009 all'età di 76 anni mentre questo testo stava per essere consegnato alle stampe. E' doveroso ricordarne la straordinaria figura di uomo e di docente la cui passione verso la didattica e la ricerca era seconda soltanto all'amore nei confronti della sua famiglia.

Palermo, 29 ottobre 2009

Erasmus Vassallo

1. Nozioni di R

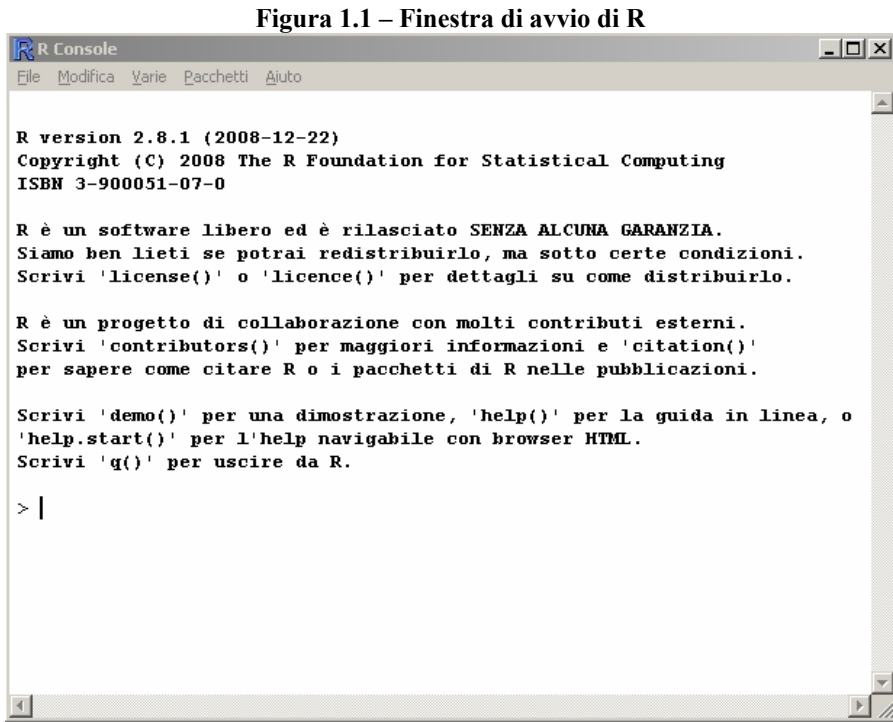
1.1 **R** è un ambiente di sviluppo statistico open-source con licenza GNU-GPL basato su un linguaggio di programmazione orientato agli oggetti ed ormai ben noto non soltanto agli addetti ai lavori. Nasce nel 1996 da un progetto di Ross Ihaka and Robert Gentleman presso l'Università di Auckland come derivazione del linguaggio statistico S, tant'è che all'inizio gran parte delle routine giravano sull'una e sull'altra piattaforma di calcolo senza la necessità di compiere modifiche al codice; non è un caso che John M. Chambers, uno dei fondatori del linguaggio S, sia stato tra i programmatori più attivi nel sostegno di **R**.

Tuttavia, le versioni più recenti di **R** si sono maggiormente differenziate da S enfatizzando le specificità di un linguaggio che ha come suo principale punto di forza la totale gratuità. E' questa la vera chiave del successo di **R** che, oggi, può vantare su una vasta e sempre crescente libreria di routine statistiche aggiunte continuamente da programmatori occasionali, esperti o semplici utenti. Non è dunque un caso che l'edizione del 7 Gennaio 2009 del prestigioso New York Times ha dedicato un articolo proprio ad **R** ed alla sua crescente diffusione negli ambienti statistici.

Per queste ragioni, il testo dedica specifica attenzione a **R** con riferimento alla release 2.8.1 anche se i comandi utilizzati funzionano con altre versioni del software: il lettore troverà le informazioni necessarie ed i file di lavoro all'indirizzo web già segnalato in Prefazione

<http://www.erasmo.vassallo.unipa.it/laboratorio>

La Figura 1.1 riporta la finestra di avvio di **R** con il cursore in riga di comando



La finestra di **R** è nota al tipico utilizzatore di questo testo poiché il Corso di Laurea in Statistica di Palermo offre già nel primo anno alcuni corsi orientati all'apprendimento del software, mentre il lettore con minori competenze potrà utilmente consultare i riferimenti di cui al già citato indirizzo web.

Qui, pertanto, riportiamo soltanto alcuni comandi utili a sintetizzare le funzionalità di base di **R** come stimolo per gli approfondimenti che il lettore riterrà necessari.

1.2 Iniziando dalle assegnazioni di valore, il comando

x <- 4