

# BioSta guida al

MASSIMO GUERRIERO\* - MARCO SANDRINI\*\*

---

*Ecco una breve, ma accurata guida per chi, occupandosi di biostatistica ed epidemiologia, deve scegliere tra i numerosi software e packages statistici per l'analisi dei dati e il calcolo scientifico attualmente disponibili*



# istica SOFTWARE



In questo saggio vogliamo tracciare una sintetica panoramica delle caratteristiche dei principali e più diffusi packages statistici per personal computer dedicati all'analisi dei dati di natura medica ed epidemiologica. Particolare enfasi viene data al software di pubblico dominio reperibile su Internet. Nella sezione conclusiva vengono esposti i principali criteri che devono a nostro avviso guidare nella scelta di un programma di analisi statistica, unitamente ad alcuni riflessioni sulle conseguenze che possono derivare da un "abuso" di questi strumenti informatici.

Con l'avvento dei mainframe, ma più ancora con la comparsa dei primi personal computer (PC), avvenuta intorno alla seconda metà degli anni '80, si è assistito alla progressiva comparsa sul mercato del software di una serie di prodotti per il trattamento, l'analisi, la modellizzazione e la rappresentazione grafica dei dati mediante tecniche statistiche. Il numero di funzionalità disponibili in questi "packages" è andato crescendo negli anni di pari passo con le sempre maggiori capacità computazionali dell'hardware immesso sul mercato e con lo sviluppo di sistemi operativi e linguaggi di programmazione potenti, stabili e in grado di sfruttare sempre meglio queste potenzialità.

A partire dai primi anni '90, la possibilità di condividere su scala planetaria in modo semplice, veloce e a basso costo informazioni, idee, esperienze e software attraverso la grande rete informatica Internet ha reso possibile la nascita di gruppi spontanei di esperti (solitamente appartenenti al mondo accademico) impegnati a distanza su progetti comuni per lo sviluppo di software scientifici di vario tipo. Lungo queste strade si è perve-

nuti allo scenario attuale che vede a disposizione di chi si occupa di biostatistica ed epidemiologia, o più in generale di chi si occupa di calcolo scientifico, una vastissima gamma di programmi per coltore, utili per le più disparate analisi: programmi più o meno specializzati, più o meno versatili, più o meno semplici da utilizzare, dotati di una più o meno ricca documentazione, più o meno esenti da errori ed instabilità di funzionamento, più o meno costosi, taluni di essi distribuiti addirittura gratuitamente (e fra questi ve ne sono, come vedremo, di ottimo livello davvero).

Da un lato la disponibilità di questi strumenti di calcolo ha di fatto reso possibile lo sviluppo di alcune importantissime tecniche statistiche di tipo "computer intensive", come ad esempio i metodi esatti e il bootstrap, e, più in generale, ha consentito l'applicazione dei diversi metodi statistici a grandi moli di dati. Ma più ancora, l'accesso alle varie tecniche di analisi dei dati è diventato ora possibile ad un pubblico sempre più ampio, anche non tecnico, non specializzato in campo statistico o addirittura senza nemmeno una base di tipo matematico/statistico.

Accanto a questi effetti palesemente positivi, l'avvento di pacchetti statistici sempre più "user friendly" ha contribuito d'altro canto a diffondere, specie fra i non addetti ai lavori, l'idea secondo cui condurre uno studio, una ricerca scientifica metodologicamente corretta si riduce a saper utilizzare (spesso in modo nemmeno troppo approfondito) qualcuno di questi strumenti informatici. Tale tendenza rischia di rendere difficoltoso il dialogo, di per sé fecondo, fra lo statistico e il ricercatore impegnato in una qualche disciplina scientifica e si oppone al raggiungimento dello scopo per il quale la Statistica nacque e si sviluppò: fornire uno schema concettuale rigoroso per la metodologia dell'indagine scientifica, dalla fase iniziale di definizione degli obiettivi della ricerca, sino all'utilizzazione finale dei risultati (i).



Il presente saggio nel secondo paragrafo mostrerà, senza alcuna pretesa di esaustività, alcuni fra i più diffusi software commerciali di tipo statistico, mentre nel terzo paragrafo verranno presentati quelli di pubblico dominio. Nel quarto paragrafo si discute su alcuni aspetti inerenti la scelta e l'uso di questi strumenti.

### SOFTWARE COMMERCIALE BMDP (BioMeDical computer programs - P series)

BMDP è uno dei pacchetti più presenti sul mercato dal maggior numero di anni. Il suo sviluppo iniziò negli anni '60 su mainframe. Si tratta di un insieme di programmi (ora anche per PC) per l'analisi statistica dei dati, che vanno da semplici statistiche descrittive e tabelle di frequenza, all'analisi della varianza, alla regressione, fino all'analisi della sopravvivenza, analisi di serie storiche e analisi delle corrispondenze. Molti di questi programmi mettono a disposizione opzioni che sono difficilmente rintracciabili in altri pacchetti statistici (vedi ad es. (ii) per quanto concerne le caratteristiche di BMDP nell'analisi di sopravvivenza). Il linguaggio dei comandi di BMDP è relativamente facile da imparare e da usare, anche se non intuitivo. I manuali sono ricchi di esempi che agevolano la comprensione dell'uso di questo package.

### SAS

Si tratta probabilmente del software che gode di maggior reputazione in ambito statistico ma anche medico, epidemiologico e farmaceutico. Nato negli anni '70, questo prodotto è andato diventando un vero e proprio standard di riferimento per il calcolo statistico e conta ora migliaia di installazioni in tutto il



mondo. La Food and Drug Administration (FDA) ha da qualche anno individuato in SAS lo standard per la ricezione, l'archiviazione e l'elaborazione statistica di dati originati da sperimentazioni cliniche e da poco tempo anche il Ministero della Sanità italiana richiede SAS per la preparazione del Final Report nella fase di registrazione di un prodotto farmaceutico (iii). Numerose importanti aziende farmaceutiche lo utilizzano sia per l'analisi di dati legati alla sperimentazione clinica che, più in generale, per il trattamento delle informazioni.

Come SPSS, anche SAS è strutturato in moduli. Ve ne sono per tutte le esigenze ed elencarli tutti sarebbe qui impossibile. Il modulo SAS/STAT è sicuramente quello di maggior interesse nel contesto del nostro tema e include potenti procedure per l'analisi di varianza (ANOVA, MANOVA, per misure ripetute, etc.), regressione (lineare, non lineare, GLM, etc.), dati categoriali (misure e test di associazione, regressione logistica, modelli loglineari, etc.), per l'analisi multivariata (analisi di correlazione canonica, analisi discriminante, modelli di equazioni strutturali, etc.), l'analisi di sopravvivenza, l'analisi dei cluster, analisi non parametrica e molto altro.

Complessivamente SAS risulta essere più difficile da apprendere rispetto ad altri pacchetti di pari potenza. La sintassi del linguaggio non è propriamente intuitiva e la possibilità di un suo uso guidato attraverso menù a tendine implica l'acquisto di un altro ulteriore modulo (SAS/INSIGHT). La manualistica è a dir poco sterminata e talora disorienta l'utente, specie se alle prime armi.

Il prodotto è in ultima analisi di fascia alta, non solo come pre-

stazioni ma anche come costo. È in generale consigliato a coloro che hanno una consolidata esperienza in campo statistico e una certa dimestichezza con i linguaggi di programmazione.

### S-Plus

Questo software, ricco di funzionalità di vario tipo, è per molti versi simile al prodotto di pubblico dominio R, di cui parleremo più avanti. S-Plus ed R sono accomunati dal fatto di avere la medesima origine, di disporre entrambi di un linguaggio di programmazione estremamente potente e flessibile (e con una sintassi di fatto coincidente) e di mettere a disposizione dell'utente una vastissima gamma di analisi statistiche, anche per grandi moli di dati. Va comunque riconosciuto che S-Plus, grazie alla sua intuitiva e ben or-



ganizzata interfaccia grafica permette di essere subito utilizzato anche dall'utente meno esperto e non impone (per un uso di base) quel non indifferente investimento di tempo ed energie altrimenti richiesto da R. Per maggiori dettagli rimandiamo a quanto detto riguardo R nella sezione del software di pubblico dominio.

### SPSS

Si tratta di un software di ampia diffusione (per la nostra esperienza, il più diffuso) fra chi si occupa di biostatistica ed epidemiologia, ma anche in tutte le scienze sociali (di fatto l'acronimo sta per Statistical Package for the Social Sciences). Le ragioni per cui questo prodotto gode di un così grande numero di utenti vanno ricercate principalmente nella sua semplicità d'uso, nella sua vastissima disponibilità di strumenti di analisi e non ultimo nel fatto che si tratta di uno dei prodotti presenti sul mercato dal maggior numero di anni, sempre sostenuto da un'abile politica di marketing.

SPSS è organizzato in moduli. Il modulo base ha funzionalità piuttosto limitate che comprendono statistiche descrittive, bivariante (t-test, ANOVA, test non parametrici), regressione lineare, cluster

analysis e non molto altro ancora. Ottimo l'editor dei dati, caratterizzato da un formato di foglio elettronico semplice ed efficace per l'immissione e la visualizzazione rapida dei file di dati. I grafici ad alta risoluzione e qualità sono parte integrante del modulo base. Oltre ai menù e alle finestre di dialogo, in SPSS è disponibile un linguaggio a comandi che consente di creare ed eseguire sessioni di lavoro; ad alcune funzioni del sistema è possibile accedere solo mediante la sintassi dei comandi. SPSS è complessivamente di uso molto semplice ed intuitivo, forse uno dei prodotti che richiedono il minor investimento di tempo per il suo apprendimento.

Esiste un modulo aggiuntivo per qualsiasi esigenza. Regressione logistica e modelli probit con il modulo SPSS Regression Models; GLM, modelli loglineari, ANOVA per misure ripetute e analisi della sopravvivenza con SPSS Advanced Models; alberi di classificazione e regressione con Answer Tree; test esatti con SPSS Exact Tests; trattamento dei dati mancanti con Spss Missing Value Analysis. E molto, molto altro ancora.

Accanto ad una così grande offerta di prodotti, i costi sono piuttosto alti e collocano questo software in una fascia di prezzo medio-alta.

### Stata

Questo software, a nostro modesto avviso, rappresenta per molti versi una scelta "ottima". Utilizzo in linea di massima assai semplice (la recentissima versione 8 per la prima volta include anche analisi guidate da menù a tendina), costo contenuto, elevatissimo numero di funzionalità (specie nel campo della biostatistica e della epidemiologia (iv)), capacità di tratta-



re grandi moli di dati con velocità ed affidabilità, disponibilità di un'ottima manualistica e di un periodico (lo *Stata Journal*) che propone interessanti aggiornamenti ed approfondimenti, sintassi dei comandi semplice ed elegante, linguaggio di programmazione potente (anche se piuttosto farraginoso), vasta gamma di possibilità di rappresentazione grafica dei dati (specie nella versione 8) rendono questo prodotto appetibile sotto moltissimi punti di vista. *Stata* è uno strumento ormai diffuso su scala planetaria con migliaia di installazioni ed è destinato sia allo studente che a chi si occupa di ricerca scientifica.

Sulla base della nostra esperienza (ormai più che decennale) nel campo del calcolo scientifico per la medicina e l'economia, ci sentiamo di poterlo segnalare come un prodotto di grande interesse.

### SOFTWARE DI PUBBLICO DOMINIO

Una ricerca, anche sommaria, del software statistico/matematico di pubblico dominio disponibile sulla Rete per l'analisi di dati di natura biomedica rivela la presenza di una quantità davvero considerevole di materiale (v).

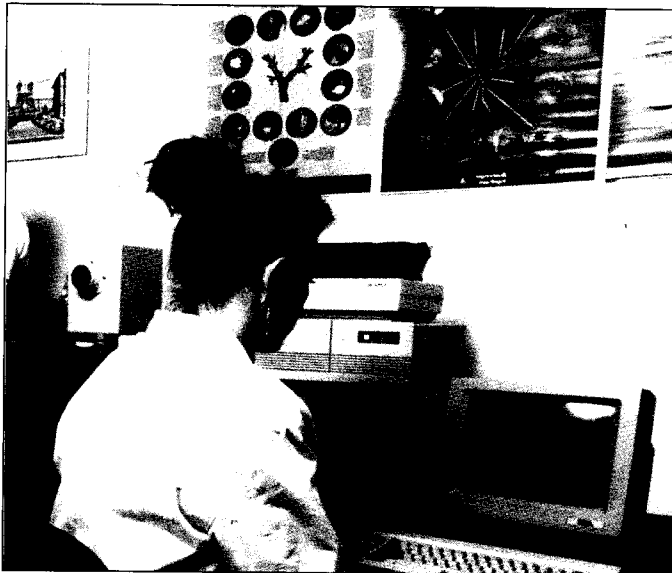
In linea di massima detto software può essere distinto in due grandi categorie: i package di uso generale, in grado di effettuare un'ampia gamma di operazioni di pre-processamento e analisi statistica dei dati, e package più specifici che si occupano di circoscritte aree della statistica, quali ad esempio il calcolo della dimensione campionaria, la meta-analisi o la visualizzazione di dati multidimensionali.

#### Epi Info

Epi Info è un programma di uso

semplice ed intuitivo che permette di creare e trattare un questionario o una form in pochi minuti, di personalizzare il processo di data entry e di analizzare poi i dati raccolti. Esso viene quotidianamente utilizzato da numerose équipes di ricercatori e professionisti impegnati nella sanità pubblica di tutto il mondo per effettuare inchieste epidemiologiche, per la gestione di dati sanitari, di registri di malattie, per la sorveglianza epidemiologica e molto altro ancora.

Oltre alla possibilità di creare tabelle, grafici e analisi statistiche di una certa complessità (stima di modelli logistici, curve di sopravvivenza, etc.), questo prodotto offre piena compatibilità con database ODBC ed SQL; offre inoltre la possibilità agli utenti di sviluppare moduli aggiuntivi personalizzati e di interfacciarsi con dati GIS in



formato ArcView per la creazione di mappe spaziali di sanità pubblica. Buona la documentazione. Senza dubbio alcuno, possiamo affermare che la conoscenza di Epi Info dovrebbe far parte del bagaglio di strumenti di indagine di ogni epidemiologo. Disponibile sul sito <http://www.epiinfo.it/>

#### INSTAT

Nato circa vent'anni or sono, Instat è un package statistico generale di utilizzo semplice ed intuitivo per piattaforma Windows, sviluppato e distribuito dallo Statistical Services Centre dell'Università di Reading, UK. Destinato in origine all'insegnamento delle idee chiave su cui si fonda la scienza statistica, questo prodotto è andato di fatto diventando nel tempo uno strumento utile per la ricerca in tutte quelle discipline che richiedono delle analisi statistiche dei dati. Instat contiene numerose funzionalità di manipolazione e trattamento dei dati, intervalli di confidenza e test di ipotesi (parametrici e non parametrici) per il confronto di medie, proporzioni e conteggi, stima di modelli di regressione lineare e modelli loglineari, varie possibilità di rappresentazione grafica (istogrammi, grafici di dispersione, boxplot, stem-and-leaf plot).

Disponibile sul sito: <http://www.rdg.ac.uk/ssc/instat/instat.html>

#### PEPI (computer Programs for EPidemiologists)

Attualmente PEPI si compone di 43 programmi che permettono di accedere ad una gamma molto vasta di analisi per dati epidemiologici (e non solo). Si va dalla regressione logistica, curve ROC, curve di sopravvivenza, ai numerosi test parametrici e non parametrici, test esatti, calcolo della numerosità campionaria e molto altro ancora. L'uso è semplice e immediato. Il solo manuale esistente è il libro di Abramson e Gahlinger (vi). Questa collezione di programmi è un ottimo strumento didattico, ma anche di ricerca per chi, con nessuna spesa, voglia effettuare tutta una serie di analisi, anche avanzate, sui propri dati. Caldamente consigliato.

Si veda il sito: <http://sagebrushpress.com/pepibook.html>.

#### R

Questo prodotto merita qui una speciale menzione perché rappresenta uno strumento di una complessità e di una potenza difficilmente raggiunti da altri prodotti, anche commerciali. R è, al pari di S-Plus, un vero e proprio linguaggio di programmazione e al contempo un ambiente per l'analisi statistica dei dati e la loro visualizzazione. R ed S-Plus hanno una origine comune: il linguaggio S sviluppato presso i Bell Laboratories intorno alla metà degli anni '70 da Becher e Chambers (vii). Cercare di elencare tutte le funzionalità disponibili in R è praticamente impossibile. Molte di queste sono già disponibili al momento dell'installazione del software sul proprio PC, moltissime altre possono esse-



Nemico dei tuoi occhi.

Amico dei tuoi occhi.



Forse non lo sai, ma le cause di allergia possono essere migliaia: alcune rarissime, altre relativamente diffuse. Il polline, ad esempio, è una delle cause di allergia più frequenti. Come reazione, l'occhio libera l'istamina, sostanza che provoca dilatazione dei capillari e un maggior afflusso di sangue nei tessuti oculari. Ecco perché in presenza di una causa di allergia gli occhi si gonfiano, si arrossano e danno un fastidioso senso di prurito. In questi casi, Imidazyl Antistaminico può aiutarti. La sua formula infatti è a base di una sostanza antistaminica specifica per neutralizzare la reazione allergica. Imidazyl Antistaminico è disponibile anche nella versione monodose senza conservanti. Ecco perché i tuoi occhi trovano in Imidazyl un amico.

collirio  
**imidazyl Antistaminico**

Una risposta alle irritazioni allergiche dei tuoi occhi.

RECORDATI OTC

Per ulteriori informazioni visitate il sito [www.familycare.it](http://www.familycare.it)

È un medicinale. Ha controindicazioni. Leggere attentamente il foglio illustrativo. Non usare in caso di glaucoma e sotto i 12 anni. Evitare l'uso prolungato. Depositato presso il Ministero della Salute in data 27/11/2002

re aggiunte caricando i cosiddetti "contributed packages", cioè degli insiemi di nuovi comandi, sviluppati da utenti di R. Fra le numerosissime tecniche statistiche disponibili: modelli lineari e lineari generalizzati (GLM), modelli di regressione non lineare, regressione non parametrica (GAM), test di ipotesi di tipo parametrico e non parametrico, tecniche bootstrap, analisi della varianza, numerose tecniche di clustering, analisi discriminante, tecniche di partizionamento ricorsivo (CART), analisi della sopravvivenza, analisi delle serie temporali e molto, molto altro ancora. Per quanto concerne le potenzialità grafiche troviamo praticamente tutti i tipi di rappresentazioni grafiche a due e tre dimensioni (istogrammi, grafici di dispersione, boxplot, grafici di comparazione dei percentili, superfici, curve di livello, etc) e la tecnica avanzata di rappresentazione dei dati multidimensionali denominata Trellis (viii).

R non è in generale di utilizzo facile ed intuitivo, come la maggior parte degli strumenti dotati di grande potenza e flessibilità. Mal si adatta dunque all'uso da parte di un'utente priva di una formazione matematico/statistica e di una conoscenza delle nozioni di base di almeno un linguaggio di programmazione. Esso rappresenta però un investimento con grandi ritorni: superato lo scoglio iniziale dell'apprendimento, l'utente diventa padrone di un ambiente di calcolo in cui praticamente qualsiasi trattamento, analisi e rappresentazione grafica dei dati è possibile. A ciò si aggiunge l'inevitabile vantaggio di un prodotto completamente gratuito, straordinariamente documentato, collaudato da un elevatissimo numero di membri della comunità statistica internazionale e tenuto costantemente aggiornato.

Questo software è disponibile sul il sito: <http://www.r-project.org/>.

#### Vista

"See what your data seem to say", questo è il motto con il quale viene presentato al pubblico questo interessantissimo strumento sviluppato da Forrest W. Young, professore emerito in psicologia quantitativa e collaboratore presso il L.L. Thurstone Psychometric Laboratory dell'Università della North Carolina. Come per altri prodotti di questo tipo (vedi ad es. Xgobi (ix)), l'idea di fondo è quella di fornire uno strumento di esplorazione di dati multidimensionali basato su grafici altamente dinamici ed interattivi, capaci di condurre l'utente in un vero e proprio "viaggio esplorativo" all'interno delle osservazioni campionarie, viaggio attraverso cui poter acquisire conoscenza sui dati e sui legami esistenti fra le variabili in gioco e per poter formulare congetture ed ipotesi (da verificare a quel punto attraverso le tecniche classiche della statistica).

Fra gli strumenti disponibili in questo package troviamo grafici di dispersione e matrici di grafici di dispersione, istogrammi, boxplot, grafici basati sulle coordinate parallele, biplot, percentili e grafici di comparazione dei percentili, guided tour plots, spin plots, spread plots (gruppi di finestre grafiche interagenti dinamicamente ed interattivamente) e



molto altro ancora. A questi si aggiungono strumenti per l'analisi statistica uni e multivariata, quali l'analisi della varianza, l'analisi delle componenti principali, il multidimensional scaling e l'analisi delle corrispondenze.

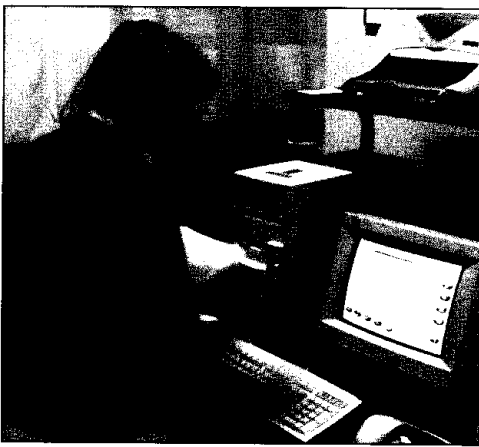
È infine disponibile una buona documentazione che comprende un manuale e un ottimo corso sull'analisi dei dati con Vista.

Il prodotto può essere scaricato dal sito: <http://forrest.psych.unc.edu/research/>.

### CONCLUSIONI

I criteri che devono guidare nella scelta di un software di analisi dei dati sono in generale molteplici. Il costo è il primo elemento da valutare, se si dispone di budget limitati. L'assoluta indisponibilità di fondi non può ovviamente che guidare verso il software di pubblico dominio. Per chi ha la possibilità di accedere anche ai packages più costosi, vale una regola fondamentale: più costoso non significa necessariamente migliore. In taluni casi un pacchetto reperibile gratuitamente su Internet può risultare ancora la scelta ottimale.

Altro fattore determinante è l'uso che si intende fare del software. Se le proprie esigenze sono modeste e si reputa che rimarranno tali anche in futuro, allora ci si può orientare verso strumenti dalle limitate capacità computazionali, che di solito offrono in cambio una maggiore facilità di apprendimento ed un costo inferiore o addirittura nullo. Congiuntamente a questo elemento, ne vanno considerati altri due: il proprio livello di conoscenza della metodologia statistica e la propria propensione verso l'uso di strumenti informatici di una certa complessità (che richiedono un più consistente investimento iniziale di tempo per lo studio del loro funzionamento). È in linea di massima sconsigliabile avventurarsi in package molto ricchi di funzionalità quando non si è disposti a dedicare tempo ed energie per imparare ad usarli con padronanza e/o per approfondire la conoscenza della teoria statistica di base. In questi casi è più ragionevole optare per strumenti semplici e con un "raggio d'azione" più circoscritto. D'altronde arrivare ad utilizzare con dimestichezza un software statistico potente significa allargare i propri orizzonti e significa disporre



di un'arma importante per poter affrontare e risolvere una gamma molto ampia di problemi di natura statistica. Da non sottovalutare poi il fatto che la scelta di uno strumento largamente diffuso fra i membri della comunità scientifica di appartenenza significa maggiore possibilità di dialogo e di scambio.

A livello più tecnico, è importante che (a) il pacchetto non sia chiuso su sé stesso, ma che permetta di importare ed esportare dati

agevolmente e in modo affidabile con il maggior numero di pacchetti statistici e con i più diffusi database management systems; (b) permetta una facile visualizzazione ed editing dei dati caricati; (c) che permetta di creare grafici e tabelle esportabili nei più comuni word processor; (d) che i grafici prodotti siano facilmente ed estesamente personalizzabili.

Come si è detto nell'introduzione, la disponibilità di pacchetti di sempre più facile utilizzo, con analisi guidate attraverso menù a tendina e finestre di dialogo, non ha portato con sé solo effetti positivi, ma ha contribuito a nostro avviso al diffondersi, anche fra medici e biologi, della tendenza a ritenere che l'apporto fornito dalla Statistica nell'ambito di una ricerca scientifica si esaurisca al momento dell'elaborazione dei dati con il calcolatore. Le cinque fasi in cui si articola uno studio statistico (definizione degli obiettivi, rilevazione dei dati, elaborazione metodologica, presentazione ed interpretazione dei risultati, utilizzazione dei risultati) coprono in realtà l'intero arco di sviluppo di una ricerca empirica. Se il contributo viene a mancare in una o più di queste tappe, il rigore scientifico di tutto lo studio rischia di crollare e l'applicazione di più o meno sofisticate metodologie di analisi con più o meno sofisticati strumenti hardware e software si riduce ad un vano dispendio di tempo ed energie.

\*Dip. E.S.I. Sez. Statistica, Facoltà di Economia, Università di Verona

\*\*C.I.D.E. - Centro Interdipartimentale di Documentazione Economica, Università di Verona

### BIBLIOGRAFIA

- i D. Piccolo, Statistica (seconda ed.), Il Mulino, Bologna, 1998.
- ii F.E. Harrell Jr. and R. Goldstein, A Survey of Microcomputer Survival Analysis Software: The Need for an Integrated Framework, *The American Statistician*, 51, 4, 1997.
- iii R. Fasani, G. Reggiardo - Medi Service, Genova, Analisi di dati clinici con SAS System, Convegno per gli utenti SAS del mondo scientifico e accademico, Milano, 30 maggio 2000, Aula Magna Università degli Studi, sede di Bicocca.
- iv M. Hills and B.L. De Stavola, A Short Introduction to Stata for Biostatistics, Timberlake Consultants Press, 2002.
- v Un sito che rappresenta un buon punto di partenza per muovere i primi passi in questo vasto mare di informazioni è: <http://members.aol.com/johnp71/javasta2.html>. Un ricchissimo e quasi esaustivo elenco di software per la statistica è reperibile all'indirizzo: [http://www.stata.com/links/stat\\_software.html](http://www.stata.com/links/stat_software.html).
- vi J.H. Abramson and P.M. Gahlinger, Computer Programs for Epidemiologic Analyses: PEPI v. 4.0, Sagebrush Press.
- vii R.A. Becker and J.M. Chambers, 'S': A Language and System for Data Analysis, Bell Laboratories, September, 1978. <http://cm.bell-labs.com/cm/ms/departments/sia/doc/comp.html>
- viii R.A. Becker, W.S. Cleveland, M-J Shyu and S.P. Kaluzny A Tour of Trellis Graphics. <http://cm.bell-labs.com/cm/ms/departments/sia/doc/trellis.tour.col.ps>
- ix D.F. Swayne, D. Cook and A. Buja, XGobi: Interactive Dynamic Data Visualization in the X Window System, *Journal of Computational and Graphical Statistics* 1998, 7 (1). <http://www.research.att.com/areas/stat/xgobi/>.