Software, Biostatistical

Biostatisticians, and applied researchers using statistics, started to use statistical computer packages (by which I mean pre-written and compiled instructions to the computer for performing some form of statistical analysis) for data analysis during the 1950s. Almost immediately changes occurred in what data were analyzed and in how they were analyzed. Changes in computer hardware have brought changes in the type and quantity of software available. The advent of microcomputers in the late 1970s and early 1980s increased the rate of change and the amount of new software packages. There are currently well over 1000 statistical software packages available on a range of computer hardware platforms.

A database of citations to published reviews of statistical software is available [8]. A good review should tell potential users what the package does, how well it does it, how easy the package is to learn and to use, and how flexible the package is. Also available is information on how certain extendable packages make both vendor-written and user-written extensions available to users; information about the cost, and example contributions, have been presented in the "Editor's notes" of the Statistical Computing Section of *The American Statistician* (see, for example, [8]).

Some Historical Notes

The 1950s saw the first occurrence of statistical software, usually specialized single purpose programs that would run on one type of machine only. Some of these were written by users, but hardware vendors were the first important source (for example, SSP from IBM). The appearance of FORTRAN in the late 1950s saw the first real surge of software and the first occurrence, to my knowledge, of generally useful software not written by a hardware vendor; this was "BIMED", later called BMD, then BMDP, which was started at the University of California at Los Angeles about 1960. By the mid-1960s several other packages had appeared, including PSTAT, SPSS, and SAS in the US and Genstat from England and Australia. All of these packages still exist. A number of other packages also appeared during the 1960s (e.g. OSIRIS, Datatext), but most of these, as far as I know, are no longer available.

These packages were neither well integrated nor comprehensive in coverage by the standards of today. They often used unacceptable algorithms or were prone to coding mistakes which gave wrong, or inaccurate, answers. (For example, Longley [17], using a multicollinear data set, showed problems in a number of packages.) However, prior to the availability of packages such as these, days could be spent, using a mechanical calculator or pencil and paper, to estimate, say, one regression with two covariates on a relatively small data set.

The late 1960s and early 1970s not only saw the appearance of additional software packages, some highly specialized (e.g. just for sample-size calculations) rather than general purpose, but also witnessed the setting up of committees by statistical associations to work on evaluating and designing software: GLIM originated under the auspices of a Royal Statistical Society (RSS) Committee; the American Statistical Association (ASA) set up a Committee on Statistical Program Packages in 1973 to help in evaluating software [3, 4]. The RSS committee, now called the "GLIM Working Party" still exists, as does the software, and the RSS receives a royalty on each version of GLIM sold. The ASA, however, no longer has any such committee.

A Categorization Scheme for Biostatistical Software

The range of software currently available makes any categorization scheme somewhat problematic. The categorization presented here is limited to one dimension: the type of user to whom the vendor expects to sell (and is further limited to software aimed at professionals); a broader categorization scheme can be found in [9]. It is impossible to include all existing packages. I primarily included packages well known to me; within each category the packages are listed alphabetically. Owing to space limitations, only contact information and a brief overview of the package are given. Contact information is for the headquarters of the company; many companies have sales and support offices in other countries.

General Purpose, Useful for Biostatistics

1. Integrated packages, including
   (a) BMDP, purchase via Statistical Solutions, 8 South bank, Crosse's green, Cork, Ireland;

Encyclopedia of Biostatistics, Online © 2005 John Wiley & Sons, Ltd.

This article was published in the *Encyclopedia of Biostatistics* in 2005 by John Wiley & Sons, Ltd.
DOI: 10.1002/0470011815.b2a14033

+353 21 4319629; SPSS Inc., 444 N. Michigan Ave., Chicago, IL 60611, USA; (312) 329-4000; its original design was aimed squarely at biostatistical goals; it is available for several computer platforms (DOS, UNIX, mainframes).

(b) Data Desk, Data Description, Inc., 840 Hanshaw road, 2nd floor, Ithaca, NY 14850; it is available on both Macintosh and Windows platforms.

(c) Genstat, Numerical Algorithms Group, Ltd, Wilkinson House, Jordan Hill Road, Oxford OX2 8DR, UK; (+44) 1865-511245; runs under Windows and several workstation operating systems, including UNIX, VMS and SunOS.

(d) GLIM, Numerical Algorithms Group, Ltd, Wilkinson House, Jordan Hill Road, Oxford OX2 8DR, UK; (+44) 1865-511245; runs under DOS and several workstation operating systems (e.g. UNIX, VMS, SunOS).

(e) JMP, SAS Institute, Inc., SAS Campus Drive, Cary, NC 27513, USA; (919) 677-8000; it is available for both the Macintosh and Windows platforms.

(f) Minitab, Minitab, Inc., 3081 Enterprise Drive, State College, PA 16801, USA; (814) 238-3280; has been widely used in educational environments; it is available for several computer platforms (Macintosh, Windows, UNIX and mainframes).

(g) NCSS, 329 North 1000 East, Kaysville, UT 84037, USA; (801) 546-0445; runs under Windows.

(h) SAS, SAS Institute, Inc., SAS Campus Drive, Cary, NC 27513, USA; (919) 677-8000; runs under several platforms (Windows, UNIX, mainframes).

(i) SPSS, SPSS Inc., 444 N. Michigan Ave., Chicago, IL 60611, USA; (312) 329-4000; originally designed for use by social scientists; runs under several platforms (Macintosh, Windows, UNIX, mainframes).

(j) Stata, Stata Corp., 702 University Drive East, College Station, TX 77840, USA; (800) 782-8272; runs under several platforms (Macintosh, Windows, UNIX).

(k) Statistica, Statsoft, Inc., 2325 East 13th St., Tulsa, OK 74104, USA; (918) 749-1119; runs under Macintosh and Windows operating systems.

(l) Systat, Systat Software. Inc., 501 Suite "C", Point Richmond Tech Center, Canal Blvd., Richmond, CA 94804; SPSS Inc., 444 N. Michigan Ave., Chicago, IL 60611, USA; (312) 329-4000; runs under Macintosh, Windows and UNIX operating systems.

2. Packages based on programming languages; many of these, as well as at least some of the extensible packages mentioned elsewhere, can use subroutine libraries (*see* **Numerical Analysis**):

(a) Gauss, Aptech Systems, Inc., 23804 SE Kent-Kangley Road, Maple Valley, WA 98038, USA, (425) 432-7855; runs under Windows and UNIX; contains numerous statistical routines; there are also several "packages" (sets of Gauss routines) written in Gauss and relevant to biostatistical users.

(b) Matlab, The Mathworks, Inc., 3 Apple Hill Drive, Natick, MA 01760, USA; (508) 647-7000; runs under Windows, UNIX; although most early routines were aimed at engineers, there are now a sizable number of statistical routines.

(c) R, a public domain near-clone of S-Plus; this can be found on Statlib (www address: `http://lib.stat.cmu.edu/`).

(d) SC, Mole Software, 34 Greenville Road, Bloomfield, Belfast BT5 5EP, N. Ireland; (+44) (0) 1232 282654; runs under DOS.

(e) Insightful Corporation, StatSci Division of MathSoft, 1700 Westlake Avenue North, Suite 500, Seattle, WA 98109, USA, (800) 569-0123; based on the AT & T product "S"; S-Plus runs under Windows and UNIX; many new forms of analysis first appear as S (or S-Plus) programs (*see* **S-PLUS and S**).

(f) XLISP-STAT, available for free by anonymous ftp from umnstat.stat.umn.edu; there are versions for the Macintosh, Unix, and Microsoft Windows; there are at least four research groups that have written packages based on XLISP-STAT; an introduction to this package can be found in [22, 23]; introductions to three of the packages can be found in [21], [25], and [26].

Encyclopedia of Biostatistics, Online © 2005 John Wiley & Sons, Ltd.

This article was published in the *Encyclopedia of Biostatistics* in 2005 by John Wiley & Sons, Ltd.
DOI: 10.1002/0470011815.b2a14033

Aimed Specifically at Biostatistical Users

1. General purpose:
   (a) EAST, CyTel Software Corp., 675 Massachusetts Avenue, Cambridge, MA 02139, USA: (617) 661-2011; for design of sequential trials; runs under DOS.
   (b) Epicure, HiroSoft International Corp., 1463 E. Republican Ave., Suite 103, Seattle, WA 98112, USA, (206) 328-5301; runs under DOS and UNIX.
   (c) EpiInfo, originated at the US Centers for Disease Control (CDC) and since then the result of collaboration between the CDC and the World Health Organization; it is available for free on the Internet (`http://www.cdc.gov/epiinfo`); can also be purchased with a printed manual of over 500 pages, from USD, Inc., 2075-A West Park Place, Stone Mountain, GA 30087, USA, (770) 469-4098; runs under DOS.
   (d) Epilog Plus, Epicenter Software, P.O. Box 90073, Pasadena, CA 91109, USA; (626) 304-9487; runs under Windows.
   (e) True Epistat, Epistat Services, 2011 Cap Rock Circle, Richardson, TX 75080, USA; (214) 680-1376; runs under DOS.
2. Special purpose:
   (a) EAST, CyTel Software Corp., 675 Massachusetts Avenue, Cambridge, MA 02139, USA; (617) 661-2011; for design of sequential trials; runs under DOS.
   (b) PEST, The MPS Research Unit, The University of Reading, Earley Gate, Reading RG6 6FN, UK; for design and analysis of sequential trials; runs under DOS.

Special Purpose Software that is Often Relevant to Biostatisticians

1. Randomization software:
   (a) RT, B.F.J. Manly, The Centre for Applications of Statistics and Mathematics, University of Otago, PO Box 56, Dunedin, New Zealand; 64-3-479-7774; randomization procedures for a number of parametric procedures, including anova, linear regression, spatial data, time series; runs on DOS.
   (b) StatXact, LogXact, CyTel Software Corp., 675 Massachusetts Avenue, Cambridge, MA 02139, USA; (617) 661-2011; StatXact includes randomization versions of a large number of nonparametric analyses; LogXact performs exact logistic regression; each runs under Windows.
   (c) Testimate, idv-Datenanalyse und Versuchsplanung, Wessobrunner Strasse 6, D-82131 Gauting/München, Germany; 089/8 50 80 01; includes randomization versions of a large number of nonparametric analyses;
2. Software for estimating sample sizes when designing studies. The following web site has information on more than two dozen such software packages: `http://www.interchg.ubc.ca/cacb/ power`. The following have specific biostatistical orientations:
   (a) EAST, CyTel Software Corp., 675 Massachusetts Avenue, Cambridge, MA 02139, USA: (617) 661-2011; for design of sequential trials; runs under DOS.
   (b) N and NSURV, idv-Datenanalyse und Versuchsplanung, Wessobrunner Strasse 6, D-82131 Gauting/München, Germany; 089/8 50 80 01; runs under DOS.
   (c) PASS, 329 North 1000 East, Kaysville, UT 84037, USA; (801) 546-0445; runs under Windows.
3. Software for correlated data, including longitudinal studies:
   (a) standard software: several of the packages included elsewhere in this list, including BMDP, LIMDEP SAS, S-Plus and Stata, include special routines for this type of analysis.
   (b) software for analyzing surveys; only one package above has adequate routines for dealing with weighted survey data: Stata; there are specialized packages, also:
      (i) SUDAAN, Research Triangle Institute, 3040 Cornwallis Road, P.O. Box 12194, Research Triangle Park, NC 27709, USA; (919) 541-6602; runs under Windows, UNIX and mainframes;
      (ii) WESVAR, Westat, Inc., 1650 Research Blvd., Rockville, MD 20850; (800) westat2, extension 2006.
   (c) software for hierarchical models:
      (i) HLM, Scientific Software International, 7383 N Lincoln Ave., Suite

Encyclopedia of Biostatistics, Online © 2005 John Wiley & Sons, Ltd.

This article was published in the *Encyclopedia of Biostatistics* in 2005 by John Wiley & Sons, Ltd.
DOI: 10.1002/0470011815.b2a14033

100, Lincolnwood, IL 60712, (800) 247-6113; runs under Windows or DOS;

(ii) MLWin, Centre for Multilevel Modeling Project, Institute of Education, University of London, 20 Bedford Way, London WC1H 0AL, UK; +44(0)207 612 6688; runs under Windows.

4. Software from other disciplines: econometric software such as LIMDEP; Econometric Software, Inc., 15 Gloria Place, Plainview, NY 11803, USA; (516) 938-5254; runs under DOS. Many of its routines are of the same type as biostatisticians use and it has some unique features, e.g. the survival analysis routines include left-truncated data and "cure" models.
5. Software for a specific form of analysis:
   (a) Survival; see [10] and [12].
   (b) Spatial; some of the above packages, especially Epilog Plus, Genstat, RT and S-Plus have some routines; there are some very specialized packages but they tend to be oriented to geostatistics and use very different jargon.
   (c) Circular; Oriana, Kovach Computing Services, 85 Nant-y-Felin, Pentraeth, Isle of Anglesey LL75 8UY, Wales, UK; (+44) (0) 1248-450414; specifically oriented to analysis of data in degrees (e.g. angular data such as might be used in a study of spinal injuries) or time (used in health services research); runs under Windows; the only other software I know of are some user-written routines in Stata.
6. Bayesian software; while there are a number of Bayesian software packages, most have never been reviewed anywhere; overviews appear in [5], [6], and [20]. Newer packages have started to appear, including
   (a) BUGS; World Wide Web address: `http://www.mrcbsu.cam.ac.uk/bugs`; versions for Windows and UNIX; "carries out Bayesian inference on complex statistical problems for which there is no exact analytic solution".
   (b) B/D: World Wide Web address: `http://fourier.dur.ac.uk:8000/stats/bd`; runs under Windows; "an interactive programming language which allows complete a priori and diagnostic analyses of Bayesian linear statistical problems".

Some Assessment Criteria

The following issues are of particular importance in assessing any statistical software package, regardless of whether it is specifically aimed at biostatistical users: the quality of the manual, the ease of learning and the ease of use of the package, and the accuracy of its computations.

Although some vendors would have purchasers believe that their package is usable without a manual, there are reasons for users to examine the manual carefully. Information in the manual should include:

1. Information on what is available (though each user must decide whether what is available is what is wanted, and, more importantly, whether it works in the way wanted and whether all the options desired are present).
2. At least one index; if it there is at least one, how good is it?
3. Examples of using the software; are the examples complete? That is to say, do the examples only display how the new commands (menu choices, etc.) work or is everything shown that one would actually need to complete an analysis?
4. Information on other sources of help, including courses, books, web sites, etc.
5. Information on how to interface this package with the operating system and/or with other types of software packages (such as word processing software).
6. Technical information relating to the algorithm used and how the vendor tested the software; there should be citations to the professional literature as well. Note that, as yet, very few vendors actually provide this (for a discussion in the context of a comparative review, see [2]).
7. A list, and explanation, of error messages; these should be clear to someone who does not have a PhD in computer science and should also be given at the same level as the statistical text.

Encyclopedia of Biostatistics, Online © 2005 John Wiley & Sons, Ltd.

This article was published in the *Encyclopedia of Biostatistics* in 2005 by John Wiley & Sons, Ltd.
DOI: 10.1002/0470011815.b2a14033

Manuals can also be used to discover whether the package appears to be aimed at the right type of user; for this, you should examine the manual(s) with the following in mind:

1. What type of language is used in describing and explaining the package? Jargon is rampant and differs dramatically across different disciplines.
2. What level of statistical language is used (e.g. beginner or professional)?
3. What types of graphical output are available and how integrated are the graphics and the statistical routines?
4. What types of checks and diagnostic information are available to help decide whether there are problems with the results of an estimation procedure?
5. How flexible is the software with respect to:

(a) nonstandard problems, e.g. are there choices of algorithms for standard routines such as linear regression?
(b) output; can the user affect the output of the package to ensure that it is in the most usable format for that particular use?

For a discussion of some criteria useful in assessing manuals, see [1], [18], the accompanying discussion of these articles, and the rejoinders by the authors.

A criterion often mentioned is ease of learning of the package. My experience, however, has been that this is really only important for people who will be infrequent users of the software, as these people will essentially be learning the package over again each time they use it. However, for others, the cost of learning is easily overshadowed by ease of use considerations, especially since, for even the hardest-to-learn packages, it rarely takes more than a few hours to learn at least enough to obtain some output.

Ease of use is sometimes, mistakenly, listed with ease of learning as a criterion. It is however both different and much more important. It is also, generally, harder to assess since the determination of whether something is easy to use is heavily dependent on both the level of the user and what the user is trying to do, as well as on the structure of the program. Program structure affects ease of use in many ways; a simple example relates to the difference between typing a command and clicking on a menu item. How this affects a given user depends on whether the menu defaults are what is primarily wanted and how easy it is to choose different options. Of course, at the other extreme, some users want so much of what they choose to do to be dependent on the situation, that no menu-driven program could possibly be considered "easy to use". Furthermore, there are many issues that vendors have never considered and these cannot, of course, be present in a menu. Whether they are available in a command system depends on the amount of thought the vendor put into making the package flexible (a detailed example is provided in [9]). Ease of use can also be aided by the availability of books about the packages, user groups, including e-mail lists and Usenet news groups, vendor newsletters, etc. Integration and ease of recall of various parts of the numerical and graphical output, and integration of the packages to the operating system and to other software (e.g. word processors), are also important here.

Earlier, I mentioned the issue of whether the language used in the manual was appropriate to the statistical expertise of the user. A related issue has to do with the ease with which one can assess one's analytic output. This is affected by numerous factors, including the quality of the error messages, the presence of statistics that can be used to assess assumptions underlying the technique used, and the quality and integration with the statistics of the graphics.

The final criterion to be discussed here is the quality of the numerical algorithms, which affects not only the accuracy of the result, but whether the package provides an answer, and, if it does, the efficiency with which it arrives at the answer.

1. Try the examples in the manual (the vendor should supply all example data sets on the disk with the program). If the examples cannot be reproduced, then immediately contact the vendor. While this appears to be a very simple test that no vendor should ever fail, some packages do fail this test.
2. Check reviews, especially those by reputable statisticians (e.g. reviews in *The American Statistician* or in *Applied Statistics*) (unfortunately, this latter journal is dropping its review section). A good review will supply much more information than just that related to accuracy; in particular, information should be included on the level of user targeted and on the ease of learning and using the package. Furthermore, I believe that

Encyclopedia of Biostatistics, Online © 2005 John Wiley & Sons, Ltd.

This article was published in the *Encyclopedia of Biostatistics* in 2005 by John Wiley & Sons, Ltd.
DOI: 10.1002/0470011815.b2a14033

comparative reviews are much more useful than reviews of individual packages.

3. Look in the literature for test data sets. Many "tests" are so well known that no vendor fails them anymore (this is true, for example, of the Longley [17] data). Furthermore, some tests are not relevant to the work that any particular user does. However, there are valuable benchmarks and tests in the literature (see for example, [7], [14], [15], [19], and [24]) that will help users and vendors test (a) whether the algorithms are appropriate, (b) what happens at the boundaries of either allowed data or standard language, and (c) the quality of the algorithms being used.
4. Build your own library of test data sets that are important in your own work and for which problems have previously been found. Try this library on every new package, and every upgrade received.
5. Examine the "validation" or "certification" section of the documentation, if it exists; unfortunately, most vendors do not yet provide such a section. If such a section exists, look for information regarding the algorithms used and the range and type of issues and of data used to test the software. The documentation should also discuss carefully the issue of how the vendor decided that the test result was acceptable. Also, note whether the vendor says that all tests are re-run after making any change to the software; this "regression testing" is necessary since fixing a bug in software often introduces one or more new bugs and this possibility must be checked.
6. Finally, run the analysis in at least one other software package and carefully compare the results. For this final check, the importance of having algorithm information in the manual is highlighted because for many analyses different algorithms should produce different results. This is especially true in many nonparametric analyses where the treatment of ties greatly affects the results.

Where to Go for More Information

There are no good general sources of information on what software is available. Eventually, there will probably be a source on the Internet which can be added to frequently. The number of software packages, particularly for educational uses, is growing rapidly. Some information, of course, is already available: numerous journals print reviews and a database of citations to these reviews is available [8]; numerous data sets useful for testing exist at statlib (at Carnegie Mellon University; WWW address: `http://lib.stat.cmu.edu/`) and other places on the internet. A "Statistical Software Guide" is produced approximately every two-three years. The most recent appearance of this guide, in print, was [16], but information is currently being gathered for an update report. However, none of this information is either well organized or complete. There is one commercial source of information, SciTech International, which publishes *Software for Science*; however, even their list (almost 2000 products, but including non-analysis packages such as word-processing software) is incomplete; they are especially weak, obviously, regarding shareware and freeware, which is often specialized and is often available on the Internet. Many of the vendors mentioned above have Internet World Wide Web sites; the best source for finding these in general is via a competitor: Stata, at its site (`http://www.stata.com`) maintains links to the sites of other vendors, including several suppliers of free software.

The Future – Maybe

Statistical software has been changing rapidly in recent years. The main changes, as of 1996, relate to (a) the existence of numerous specialized software packages; (b) the movement, slowly, of these specialized routines into general purpose packages; and (c) a heavy emphasis on graphical analysis, especially new types of graphics and new ways of integrating graphics into standard analysis.

While these are valuable, necessary, and will continue, there are two other changes that would be very valuable to the profession. The first relates to a better integration of what we already know about statistical assumptions with our analysis. For example, we know that the two-sample t test is somewhat affected by different variances (the amount depending on the ratio of the groups' sample sizes); many would find it helpful if, along with requested result, the software gave some information about the validity of this, and other, assumptions, for the data used. This might also help guard against the "misuse" of statistical software

Encyclopedia of Biostatistics, Online © 2005 John Wiley & Sons, Ltd.

This article was published in the *Encyclopedia of Biostatistics* in 2005 by John Wiley & Sons, Ltd.
DOI: 10.1002/0470011815.b2a14033

by those who are not well trained in statistics. This issue has been discussed in numerous articles dating back to at least the 1970s [11]. Though noting a number of potential problems, Goodnight clearly favored this type of "offensive validity checking". Haux [13] provides a number of citations on this issue and then gives a detailed example for the Mann–Whitney test and notes that BMDP, SAS and SPSS are each unsatisfactory. Note that the software should not stop the user from doing an analysis. Rather, a user should just want be provided with some additional information without making a number of other requests to the software (e.g. a separate request for equal variances, for symmetry, for heterogeneity, etc.).

A large part of any project relates to data management and data manipulation. Much, and in many projects all, of this is done with the same statistical software used for analysis. However, no current program keeps a reversible history of what the user does to the data and many do not even keep any history (or log or journal) of what was done. The unfortunate result is that often even the analyst cannot reproduce certain results. Thus, another desirable change is the implementation of some form of reversible history of data management and data manipulation so that any particular state of the data could be recreated. Some type of coding scheme should be attached to both this history and to each analysis so that for any given output it would be clear which state of the data was used in its production. The current reliance on *ad hoc*, individual, schemes is inefficient, ineffective, and unnecessary. Version control software, as used in software development, database management and even some word-processing software should be generalizable to statistical software.

References

[1] Berk, K.N. & Francis, I.S. (1978). A review of the manuals for BMDP and SPSS (with comments and rejoinders), *Journal of the American Statistical Association* **73**, 65–70.

[2] Boomsma, A. & Molenaar, I.W. (1993). Four electronic tables for probability distributions, *American Statistician* **48**, 153–162.

[3] Francis, I. & Heiberger, R.M. (1975). The evaluation of statistical program packages – the beginning, in *Proceedings of the Computer Science and Statistics Eighth Annual Symposium on the Interface*, J.W. Frane, ed. Los Angeles, pp. 106–109.

[4] Francis, I., Heiberger, R.M. & Velleman, P.F. (1975). Criteria and considerations in the evaluation of statistical program packages, *American Statistician* **29**, 52–56.

[5] Goel, P.K. (1988). Software for Bayesian analysis: current status and additional needs, in *Bayesian Statistics*, Vol. 3, J.M. Bernardo, M.H. DeGroot, D.V. Lindley & A.F.M. Smith, eds. Oxford University Press, Oxford.

[6] Goel, P.K. (1988). Software for Bayesian analysis: current status and additional needs – II, in *Computing Science and Statistics: Proceedings of the 20th Symposium on the Interface*, E.J. Wegman, D.T. Gantz & J.J. Miller, eds. American Statistical Association, Alexandria.

[7] Goldstein, R. (1987). Linear regression on IBM PC/XT/AT's, in *Computer Science and Statistics: Proceedings of the 19th Symposium on the Interface* R.M. Heiberger & M.T. Martin, eds. American Statistical Association, Alexandria, pp. 219–228.

[8] Goldstein, R. (1994). Editor's notes, *American Statistician* **48**, 254–255.

[9] Goldstein, R. (1997). Computer packages, *Encyclopedia of Statistical Sciences, Update Volume*. Wiley, New York, to appear.

[10] Goldstein, R. Anderson, J., Ash, A., Craig, B., Harrington, D. & Pagano, M. (1989). Survival analysis software on MS/PC-DOS computers, *Journal of Applied Econometrics* **4**, 393–414.

[11] Goodnight, J.H. (1975). Validity checking: how far should we go? in *Proceedings of the Computer Science and Statistics Eighth Annual Symposium on the Interface*, J.W. Frane, ed. Los Angeles, pp. 146–148.

[12] Harrell, F.E., Jr & Goldstein, R. (1997). A survey of microcomputer survival analysis software: the need for an integrated framework, *American Statistician* **51**, 360–373.

[13] Haux, R. (1983). How to detect and prevent errors in computer-supported statistical analysis: an example, *Methods of Information in Medicine* **22**, 87–92.

[14] Heiberger, R.M., Velleman, P.F. & Ypelaar, M.A. (1983). Generating test data with independently controllable features for multivariate general linear forms, *Journal of the American Statistical Association* **78**, 585–595.

[15] Knuth, D.E. (1981). *The Art of Computing*, Vol. 2: *Seminumerical Algorithms*, 2nd Ed. Addison-Wesley, Reading.

[16] Koch, A. & Haag, U. (1995). The statistical software guide '94/95, in the *Statistical Software Newsletter*, *Computational Statistics & Data Analysis* **19**, 237–261.

[17] Longley, J.W. (1967). An appraisal of least-squares programs for the electronic computer from the point of view of the user, *Journal of the American Statistical Association* **62**, 819–841.

[18] Muller, M.E. (1978). A review of the manuals for BMDP and SPSS (with comments and rejoinders), *Journal of the American Statistical Association* **73**, 71–80.

[19] Nash, J.C. (1990). *Compact Numerical Methods for Computers: Linear Algebra and Function Minimization*, 2nd Ed. Adam Hilger, Bristol.

Encyclopedia of Biostatistics, Online © 2005 John Wiley & Sons, Ltd.

This article was published in the *Encyclopedia of Biostatistics* in 2005 by John Wiley & Sons, Ltd.
DOI: 10.1002/0470011815.b2a14033

[20] Press, S.J. (1989). *Bayesian Statistics*. Wiley, New York.
[21] Stine, R. (1997). AXIS: an extensible graphical user interface, in *Statistical Computing Environments for Social Research*, R. Stine & J. Fox, eds. Sage, Thousand Oaks, pp. 175–192.
[22] Tierney, L. (1990). *LISP-STAT: An Object-Oriented Environment for Statistical Computing and Dynamic Graphics*. Wiley, New York.
[23] Tierney, L. (1997). Data analysis using LISP-STAT, in *Statistical Computing Environments for Social Research*, R. Stine & J. Fox, eds. Sage, Thousand Oaks, pp. 66–88.
[24] Velleman, P.F. & Ypelaar, M.A. (1980). Constructing regressions with controlled features: a method of probing regression performance, *Journal of the American Statistical Association* **75**, 839–844.
[25] Weisberg, S. (1997). The R-code: a graphical paradigm for regression analysis, in *Statistical Computing Environments for Social Research*, R. Stine & J. Fox, eds. Sage, Thousand Oaks, pp. 193–206.
[26] Young, F.W. & Bann, C.M. (1997). ViSta: a visual statistics system, in *Statistical Computing Environments for Social Research*, R. Stine & J. Fox, eds. Sage, Thousand Oaks, pp. 207–235.

(*See also* **Software Reliability**)

RICHARD GOLDSTEIN

Encyclopedia of Biostatistics, Online © 2005 John Wiley & Sons, Ltd.

This article was published in the *Encyclopedia of Biostatistics* in 2005 by John Wiley & Sons, Ltd.
DOI: 10.1002/0470011815.b2a14033