## About the Authors

Kent Weaver is a senior fellow in the governmental studies program at the Brookings Institution. T. J. Pempel is Glenn B. and Cleone Orr Hawkins Professor of Political Science at the University of Wisconsin-Madison. Nancy N. Davidson is acquisitions editor at the Brookings Institution. Sidney Verba is Carl H. Pforzheimer University Professor and director of the University Library at Harvard University. Keith G. Banting is director of the School of Policy Studies at Queen's University in Kingston, Ontario. Order of authorship was decided by a weighted index of time spent preparing the article and number of humorous one-liners used.

# Windows Statistics Packages

**Carl Grafton and Anne Permaloff,** *Auburn University at Montgomery*

The first article in this *PS* series appeared a decade ago and covered the first generation of microcomputer statistics packages.[1] They were useful tools that performed many of the tasks that previously could only be managed on mainframe or minicomputers. Their memory requirements were 48K or 64K, silicon pocket change by today's standards.

Our second and third *PS* statistics package articles appeared in 1988.[2] They covered programs that represented enormous improvements over their predecessors. They were faster, offered a far greater variety of data manipulation and statistical techniques, and several provided graphics output that was unavailable in earlier products. These second generation programs also required more powerful hardware. Entirely new machines were needed; memory requirements were an order of magnitude greater (instead of 64K, 640K); and a hard disk was a near necessity.

All of the second-generation statistics packages were text-oriented DOS programs (except for a few that ran on the Macintosh), although some moved into the graphics mode when displaying plots. The second-generation programs were later improved with two or sometimes three new versions. These changes, such as the

addition of menus to SYSTAT and improved data-handling capacity for STATGRAPHICS, were often welcome, but most of these programs retained their basic personalities.

In the last few months a third generation of statistics packages has been introduced. The new programs require Microsoft Windows 3.0 or 3.1. Once more, memory requirements have increased by an order of magnitude so that 4-8 megabytes are required, and new platforms are necessary. Many machines that would run a second-generation program without difficulty simply cannot be used with Windows. Windows demands more RAM, more hard disk space, a faster hard disk, faster clock speed, and graphics of higher resolution. The cost of such equipment has plummeted, but given especially severe budget constraints that plague higher education nearly everywhere, even these low prices may be too high for institutional or individual purchase.

## The Distinguishing Features of Windows Statistics Packages

A program is not necessarily better because it has been ported to Windows. In addition to greater

hardware demands and consequent costs, a Windows program may be slower than its DOS counterpart, although we have observed two startling exceptions to this normally valid rule.[3] A Windows program may have on-screen help capabilities that are not as good as its DOS counterpart, and its menu structure may not be any easier to use.

Fundamentally, what distinguishes the better Windows statistics packages from DOS programs is greater ease of use. It is not clear, however, that the convenience of the Windows programs is due to Windows per se as much as it is the result of general improvements in statistics programs. These changes have made statistics packages progressively more convenient as they have shifted from mainframe platforms through several DOS or Macintosh generations to the present.

Windows and even more hardware intensive systems are squeezing DOS software out of the marketplace. Therefore to avoid buying a soon-to-be-discontinued product, prospective purchasers of a DOS program may want to inquire whether the company involved intends to continue to improve its DOS merchandise. However, buying into a DOS dead end may not be a mistake if the price is

right and performance meets the user's needs. The purchaser of a DOS program should also ask whether it will be possible to update from the DOS version to Windows without having to pay the full cost of a new Windows program. This is an especially important consideration with expensive site licenses. As of this writing, one major manufacturer does not permit such a transition.

All of the programs listed below are full-featured statistics packages except for Origin, which is listed separately. Origin is a statistical graphics program designed to produce better graphics than a standard statistics package, but it is not intended for general statistical analysis.

Although some of these programs can theoretically run with 2 megabytes of RAM, anyone using Windows should have at least twice that much memory so a disk cache can be configured. The prospective user of SPSS, SYSTAT, or SAS should have 8 megabytes of RAM.

## The Programs

With one exception, the prices listed below are for purchases of single copies only. Site license costs are too complex to list here.

**MINITAB for Windows,** Release 9. MINITAB, Inc., 3081 Enterprise Dr., State College, PA 16801-3008. Tel. (814) 238-3280. Fax (814) 238-4383. $895 list price, $495 academic discount price. $100 upgrade. Plus $10 S&H for all.

**SAS for Windows.** SAS Institute, SAS Circle, Box 8000, Cary, NC 27512-8000. Tel. (919) 677-8000. Fax (919) 677-8123. SAS is rented, not purchased. Prices depend on site license terms and the number of extra cost modules required.

**SPSS for Windows,** Release 5. SPSS, Inc., 444 North Michigan Ave., Chicago, IL 60611-3962. Tel. (800) 543-2185. Fax (312) 329-3668. $695 list price, $495 academic discount price. Only Value Plus members can take advantage of an upgrade price from the DOS to Windows version.

**SPSS for Windows, Professional Statistics, Advanced Statistics, Tables, Trends, Categories, Lisrel, MapInfo.** Address and telephone number above. All except Advanced, Lisrel, and MapInfo, $395 per module list price, $295 academic discount price. Advanced and Lisrel are $495 list and $395 academic discount. MapInfo $995 list price and $895 academic discount. All modules require base system listed above. All modules except MapInfo may be purchased together at discounts.

**SYSTAT for Windows.** SYSTAT, Inc., 1800 Sherman Ave., Evanston, IL 60201. Tel. (708) 864-5670. Fax (708) 492-3567. $895 list price. $627 academic discount price. $250 upgrade from DOS version.

**UNISTAT for Windows,** Version 1.13. Adhoc Reading Systems, Inc., 28 Brunswick Woods Drive, East Brunswick, NJ 08816. Tel. (908) 254-7300. Fax (908) 254-7310. $610. This program is copy protected.

**WinSTAR,** Version 1.51. Anderson-Bell Corporation, P.O. Box 5160, Dept. U73A, Arvada, CO 80005. Tel. (800) 745-9751. Fax (303) 940-0595. $449 list price. $337 academic discount price. $59 upgrade from the DOS version. A fully functional student version (64 variables by 500 observations) ranges from $23 to $50 depending on the quantity ordered.

**WinSTAT,** Version 2.0. Kalmia Company, Inc., 71 Dudley Street, Cambridge, MA 02140. Tel. (617) 864-5567. $395 list price. $355.50 academic discount price.

**Origin,** Version 2.94. MicroCal, Inc., 22 Industrial Drive East, Northampton, MA 01060. Tel. (800) 969-7720 and (413) 586-0149. $495 list price. $396 academic discount price.

## Common Analytical Techniques

All of the statistical packages described here offer basic database management functions, transforms using logical statements (except for WinSTAT), descriptive statistics, crosstabulation, ANOVA, correlation, and multiple regression. They all plot histograms, box and whisker, and scatter plots. These programs offer enough statistical power for most tasks.

Prospective buyers who require advanced tools should use this article only to narrow their search. Then, descriptive materials from individual companies should be obtained and telephone calls made to insure that the product meets the purchaser's needs. Space limitations prohibit us from discussing or even listing all the tools offered by the more sophisticated programs included here. For example, among the three leading programs two dozen advanced ANOVA statistics are represented.

## Statistical Techniques

SYSTAT provides the greatest number and variety of statistical techniques if it is compared to the SPSS or SAS base systems (see Table 1). SPSS is superior to SYSTAT in this regard only if several of its modules are included, but SYSTAT remains competitive even under this comparison especially if graphical techniques are included. SAS is difficult to compare to SPSS and SYSTAT because many of its modules are not accessible via standard Windows menus. If only those parts of SAS that have been fully encompassed in the Windows environment are included in a comparison, SAS falls well behind SYSTAT and the SPSS base system.

The prospective SPSS purchaser is also offered several modules, but all are part of a fully operational Windows program and except for the unusual LISREL package all are fully integrated into the base system. The SPSS modules include:

• Trends. A comprehensive time series package. SYSTAT matches this module in most features. MINITAB and WinSTAT do not offer nearly as rich a time series feature set. The explanation of time series analysis in the WinSTAT manual is far better than MINITAB's.

• Professional. Discriminant analysis. Factor analysis. Cluster analysis. Distance and proximity measures. Reliability analysis. Multidimensional scaling. Weighted least squares regression. Two stage least squares. SYSTAT provides all of these features except the last. UNISTAT, MINITAB, and WinSTAT contain several (see Table 1).

• Advanced. Logistic analysis. Factorial analysis of variance. Multivariate analysis of variance. Repeated measures analysis of variance. Hierarchical loglinear models. Nonlinear regression. Probit and Logit. Follow-

**TABLE 1**
**Statistical Techniques**

|  | MINI | SPSS | SYSTAT | UNI | WinSTAR | WinSTAT |
|---|---|---|---|---|---|---|
| *Advanced Crosstabulation* | | | | | | |
| Cohen's Kappa | No | Yes | Yes | Yes | No | Yes |
| Contingency | Yes | Yes | Yes | Yes | No | Yes |
| Eta coefficient | No | Yes | No | No | No | No |
| Mantel-Haenszel chi-sq | No | Yes | Yes | Yes | No | No |
| *ANOVA* | | | | | | |
| Number advanced 1-way | 4 | 21 | 13 | 12 | 5 | 7 |
| More than 1-way | 2 | 5 | 5 | 3 | 4 | 2 |
| MANOVA | Yes | Adv | Yes | No | No | No |
| *Nonparametric* | | | | | | |
| Cochran's Q | No | Yes | No | Yes | No | No |
| Cohen's Kappa | No | Yes | Yes | Yes | No | No |
| Friedman | Yes | Yes | Yes | Yes | No | Yes |
| Jonckheere's trend | No | No | No | Yes | No | No |
| Kendall coef. of concord. | No | No | Yes | No | No | No |
| Kendall's W | No | Yes | No | No | No | No |
| Kolmogorov-Smirnov | No | Yes | Yes | Yes | Yes | Yes |
| Kruskal-Wallis | Yes | Yes | Yes | Yes | Yes | Yes |
| Lilliefurs test | No | Yes | Yes | Yes | No | No |
| Mann-Whitney U | Yes | Yes | Yes | Yes | Yes | Yes |
| McNemar | No | Yes | Yes | Yes | No | No |
| Mood | Yes | No | No | No | No | No |
| Moses extreme react. | No | Yes | No | Yes | No | No |
| Walch | Yes | No | No | Yes | No | No |
| Wald-Wolfowitz runs | No | Yes | Yes | No | Yes | No |
| Wilcoxon Signed Rank | Yes | Yes | Yes | Yes | Yes | Yes |
| *Various* | | | | | | |
| Cluster analysis | No | Pro | Yes | Yes | No | Yes |
| Discriminant analysis | Yes | Pro | Yes | Yes | No | Yes |
| Distance and proximity | No | Pro | Yes | No | No | No |
| Factor analysis | Yes | Pro | Yes | Yes | No | Yes |
| Factorial ANOVA | No | Adv | Yes | No | No | No |
| Hierarchical loglinear | No | Adv | Yes | No | No | No |
| Lisrel | No | Lis | No | No | No | No |
| Logistic analysis | No | Adv | Yes | No | No | No |
| M-Estimators | No | Yes | No | No | No | No |
| Multidimensional scaling | No | Pro | Yes | Yes | No | No |
| Nonlinear regression | No | Adv | Yes | Yes | No | No |
| Probit and Logit | No | Adv | Yes | No | No | No |
| Reliability analysis | No | Pro | Yes | No | No | No |
| Time series | Yes | Trend | Yes | No | No | Yes |
| Trimmed mean | Yes | Yes | No | No | No | No |
| Weighted least sq. reg. | Yes | Pro | Yes | No | No | No |

SAS features are excluded from these and other tables (except the last) because its modules (many of which are not full Windows programs) would require the addition of too many footnotes.

up life tables. Kaplan-Meier survival analysis. Cox regression. SYSTAT offers all except the last three, and the others provide a few (see Table 1).

• Categories. Conjoint analysis and optimal scaling. These features are unique to SPSS.

• Tables. Allows presentation quality multi-dimensional or nested tables similar in appearance to Lotus Improv displays. This module is unique to SPSS.

• Lisrel. Structural equation modeling. This module is unique to SPSS.

• MapInfo. Extremely powerful geographical information system. SYSTAT contains geographical mapping capabilities adequate for most political science needs as does SAS as an extra module, but MapInfo is far more powerful. None of the other programs contain mapping features.

UNISTAT offers the next largest variety of statistical techniques, followed by MINITAB, WinSTAT, and WinSTAR. Despite WinSTAT's low ranking, it provides a number of nicely implemented techniques that are unexpected in a relatively low end product. WinSTAT includes cluster

analysis, discriminant analysis, factor analysis, nonlinear regression, and time series analysis.

**Database Capabilities**

SPSS and SAS are limited in their data handling capacities only by disk space (see Table 2). SYSTAT permits 255 variables and as many observations as disk space permits. WinSTAT allows a 65,000 × 65,000 matrix, UNISTAT a 32,600 × 32,600 matrix, WinSTAR 512 variables and 32,000 observations, and Origin 100 variables and 32,000 observations. WinSTAR's capacities are the smallest with 256 variables and 1,600 observations. MINITAB permits a maximum of 1,000 variables and 4.0 million data points; the maximum number of observations depends on the number of variables used and the constraint on total number of data points.

SPSS provides the best data entry and editing capabilities that we have seen in a statistics package. We would be inclined to enter data directly into SPSS rather than the spreadsheet we usually employ for this purpose, partly because cursor movements are smooth and intuitive, but also because SPSS will save data easily in many different formats. These include Excel, WKS, WK1, WK3, SYLK, dBASE II, III, and IV, tab delimited and fixed ASCII, as well as SPSS/PC+ and SPSS Portable. If one needs to perform analyses beyond the program's capacities, the data are easily transferred elsewhere.

When SPSS is first loaded, the spreadsheet-like data input screen appears. One may then immediately begin naming variables and inputting data or reading a file. Graphics, statistics, and other operations are listed at the top of the screen in the normal manner for Windows programs. The MINITAB, SYSTAT, and Origin data entry and editing screens resemble SPSS's, but editing is not quite as handy, and none saves data in as many formats as SPSS.

Data entry in WinSTAR is performed on a well-implemented database management style screen rather than one that resembles a spreadsheet. This is a reasonable design because WinSTAR is especially oriented toward the processing of inter-

*December 1993*

791

view data, and it can even be used by a telephone interviewer to record responses. On the other hand, when numerical data such as population figures are entered one variable at a time for all observations, it is only possible to see one observation at a time, that is, the one into which data are being entered. This is a disadvantage if the user wants to visually scan data, hunting for obvious typographical errors that often stand out when viewed in a column.

Data importation from a spreadsheet into WinSTAR requires specification of a worksheet cell range even if the entire worksheet is being imported. The error message that results if the cell range is omitted is not informative.

WinSTAT uses a spreadsheet format for data entry, but it is very poorly designed and useful only for limited data editing. The cursor will not move downward or to the right when the <ENTER> key is hit. When the cursor is moved with arrow keys to a cell that contains data, changes cannot be made without a mouse point and click operation. The unnecessary hand motions involved in these procedures would be extremely annoying and time consuming if a great deal of editing were being done.

WinSTAT also responds inappropriately when numbers of 10 digits or more are entered. When such a number is keyed in, it appears on the screen without mishap, but when the <ENTER> key is depressed, an entirely different number replaces the original. Thus 8,000,000,001 became −5,899,934,591.

UNISTAT's data entry and editing screen uses a spreadsheet format, and it offers all of the control features of SPSS. However, UNISTAT's spreadsheet is so slow on a 20 Mhz computer as to be useless for data entry and only marginally functional for data editing. Even at excruciatingly slow typing speeds, data are lost due to UNISTAT's turtle-like performance. At least a 33 Mhz machine would be required to permit relatively accurate data entry in UNISTAT. These problems are mitigated somewhat by the effortless importation of a wide variety of file formats listed in Table 2.

SAS's primitive data entry screen

### TABLE 2
### Data Storage and Manipulation

| | MINI | SPSS | SYSTAT | UNI | WinSTAR | WinSTAT |
|---|---|---|---|---|---|---|
| *Database capacity:* | | | | | | |
| Max no. of variables | 1,000 | Disk | 255 | 32,600 | 512 | 65,000 |
| Max no. of obs. | RAM | Disk | Disk | 32,600 | 32,000 | 65,000 |
| Max no. of data pts. | 4.0 M | Disk | Disk | 1.06 B | 16.4 M | 4.225 B |
| *Import:* | | | | | | |
| Lotus WKS & WK1 | Yes | Yes | Yes | Yes | Yes | No |
| Lotus WK3 | Yes | Yes | No | Yes | Yes | No |
| Excel | No | Yes | Yes | Yes | No | No |
| DIF | No | No | Yes | Yes | No | Yes |
| SPSS from other operating systems | No | Yes | No | No | No | No |
| SYSTAT from other operating systems | No | No | Yes | No | No | No |
| SYLK | No | Yes | No | Yes | No | No |
| dBASE II | Yes | Yes | Yes | Yes | Yes | No |
| dBASE III | Yes | Yes | Yes | Yes | Yes | Yes |
| dBASE IV | Yes | Yes | Yes | Yes | Yes | No |
| Tab-delimited ASCII | Yes | Yes | Yes | Yes | Yes | Yes |
| Fixed ASCII | Yes | Yes | Yes | Yes | Yes | No |
| Space-delimited ASCII | Yes | Yes | Yes | Yes | Yes | Yes |
| Symphony | Yes | No | Yes | No | No | No |
| Oracle | No | Yes | No | No | No | No |
| SQL Server | No | Yes | No | No | No | No |
| Import parts of a file | No | Yes | Yes | No | Yes | No |
| Merge files | Yes | Yes | Yes | Yes | Yes | No |
| Transpose cases & variables | Yes | Yes | Yes | Yes | No | No |
| *Data entry & editing:* | | | | | | |
| Cursor control with Enter | Yes | Yes | Yes | Yes | Yes | No |
| Insert, delete rows or cols | Part | Yes | Yes | Yes | Yes | Del |
| Sort | Yes | Yes | Yes | Yes | Yes | Yes |
| Value labels | No | Yes | Yes | No | Yes | No |
| Long variable labels | Yes | Yes | No | Yes | Yes | Yes |
| Search | No | Yes | Yes | Yes | Yes | No |
| Full set of math transforms | Yes | Yes | Yes | Yes | Yes | No |
| Full set of cond. transforms | Yes | Yes | Yes | No | Yes | No |

prohibits convenient cursor movement, and data-edit commands must be accessed through a series of awkward menu choices. As with UNISTAT, SAS's data entry and editing limitations are made more bearable by its ability to import many micro, mini, and mainframe file formats.

### Transforms

Arithmetic, logical, and statistical transforms are easily performed in SPSS, SYSTAT, MINITAB, and SAS via well designed menu choices and fill-in-the-blank forms.

WinSTAR's transform capabilities do not contain all the mathematical functions offered by SYSTAT, SPSS, and SAS but they could be described as full featured nonetheless.

Origin's and UNISTAT's transforms are limited to mathematical expressions with no conditional state-

ments permitted. WinSTAT is almost entirely lacking in transform capabilities.

### Graphics

SYSTAT, SPSS, and UNISTAT (in that order) offer the greatest variety of plotting options, although SPSS and UNISTAT change places if only the SPSS base system is considered (see Table 3). Surprisingly, Origin is a poor fourth in terms of plot varieties, although it contains unique formatting and other graphics features.

SYSTAT is the clear leader in the graphics arena. It can produce many different presentation quality three-dimensional plots with a variety of surface smoothing techniques. It can also generate many specialized statistical plots for principal components analysis, discriminate analysis, multidimensional scaling, and factor

**TABLE 3**
**Graphics**

| | MINI | SPSS | SYSTAT | UNI | WinSTAR | WinSTAT |
|---|---|---|---|---|---|---|
| Presentation quality | Yes | Yes | Yes | Yes | No | No |
| Export graphics | No | No | Yes | Yes | No | No |
| Direct graph edit | Yes | Limit | Yes | No | No | No |
| 3-D GRID | No | No | Many | Yes | No | No |
| Andrews' Fourier plot | No | No | Yes | No | No | No |
| Bubble | No | No | Yes | No | No | No |
| Casement | No | Yes | Yes | No | No | No |
| Confidence ellipse & centroid | No | No | Yes | No | No | No |
| Contour | No | Yes | Yes | No | No | No |
| Curve fit | No | No | Yes | Yes | No | No |
| Dendrogram (cluster anal.) | No | Pro | Yes | Yes | No | No |
| Discrim anal. territor. map | No | Pro | Yes | No | No | No |
| Dit, fuzzy, strips | No | Yes | Yes | No | No | No |
| Error bars | No | Yes | Yes | No | No | No |
| Factor scores plot | No | Pro | Yes | Yes | No | No |
| Geographical maps | No | MI | Yes | No | No | No |
| Icon plot | No | No | Yes | No | No | No |
| Jittered scatterplot | Yes | No | Yes | No | No | No |
| Link plot points to data | No | No | Yes | No | No | No |
| Multidimensional scaling | No | Pro | Yes | No | No | No |
| Functions | No | No | Yes | Yes | No | No |
| Principal components | No | Pro | Yes | Yes | No | No |
| Quantile or Q | No | No | Yes | No | No | No |
| Right y axis | Yes | No | Yes | Yes | No | No |
| Sunflower | No | No | Yes | No | No | No |
| Three variable plot (3-D) | No | Yes | Yes | Yes | No | No |
| Variety of curve smoothing | No | Yes | Yes | Yes | No | No |
| Vector plot (factor loadings) | No | Yes | Yes | No | No | No |

analysis. SYSTAT can plot maps, and its options even for routine techniques such as bar charts and scatter charts go far beyond all the other programs except Origin.

SYSTAT allows titles or comments of varying sizes and print styles to be placed anywhere on a plot, an extremely handy feature. Plotted data points can be highlighted. The comparable data in the SYSTAT spreadsheet are then automatically highlighted as well. This is especially useful for locating outliers in a large data set.

After producing a series of three-dimensional plots, SYSTAT suddenly began generating blank sheets from a LaserJet IIP, and then crashed. SYSTAT technical staff diagnosed the difficulty as the program's tendency to hang onto memory resources after a plot has been completed. They recommended not running other programs while doing SYSTAT plots even with 8 megabytes of RAM (used by the authors), and exiting the graphical component if blank sheets begin emerging from the printer. We followed both suggestions and experienced no further problems.

SYSTAT's graphics are amazingly fast in the Windows version. A simple two-dimensional plot requires only a few seconds (compared to SPSS's 15 to 20 seconds), and even a three-dimensional SYSTAT surface plot takes only two or three minutes. DOS SYSTAT requires 15 to 20 minutes to perform comparable plots.

It is important that a package be able to export a plot. It may be desirable to send a plot to a drawing program so that it can be modified by adding clipart or comments or to a word processing program where the plot can be integrated with text. There are two major graphic formats used by the programs discussed here —bitmap and metafile. The metafile format is the better of the two. The advantage of the metafile format is that graphics can be expanded or reduced with no loss of resolution. An expanded bitmap graphic has a grainy appearance, and information is lost in a reduced bitmap, making text unattractive or unreadable. SYSTAT exports in both graphics formats making its plots readable by old and new programs.

UNISTAT plots can be saved in either metafile or bitmap formats and is comparable to SYSTAT in plotting speed. UNISTAT offers three-dimensional plotting in several forms including topographical or surface, a right or double Y axis, five curve-fitting techniques, function plotting, a probability plot (if data lie on a straight line it is normally distributed), and several multivariate analysis plots. It also offers five curve-smoothing techniques. Axes, tick marks, grids, plot symbols, titles, labels, and all other parts of its plots can be customized, but it is not possible to add comments on the plot.

SPSS omits a three-dimensional surface plot, although it includes a three-dimensional scatter plot. Curve-smoothing capabilities are less powerful than SYSTAT, UNISTAT, and Origin. One or two line footnotes can be added, a feature that should be standard on all plotting software. Another potentially useful feature has been botched in its implementation by SPSS. Comments can be added anywhere on a plot, but they can be only 20 characters long. Furthermore, a comment cannot be dragged to another location with the mouse. Its location can only be changed by entering new X–Y coordinates, an operational constraint one would only expect to see in a DOS program.

An SPSS chart can only be saved in its own format. It is unclear what that format is. We were not successful in importing an SPSS chart into graphics programs including one capable of reading more than 15 formats, and there is no discussion of graphics exporting in the SPSS manual.

SAS produces a basic collection of attractive plots with clearly designed, easy to use screens. The same can be said of MINITAB.

WinSTAR and WinSTAT provide only bare-bones, two-variable plotting. Even though both programs can use a laser printer, their plots are so unattractively designed they might as well be produced on a dot matrix printer. Neither program allows graphics exporting except via Windows' severely limited clipboard utility.

As a dedicated graphics program, Origin should offer a greater variety

of plots than full-fledged statistics packages, but it does not. It performs few advanced statistical plotting techniques, and its newly added three-dimensional capability produces plots that do not come close to SYSTAT in variety or clarity of output. Although it has many excellent features, its lack of plotting variety will eliminate it from consideration for many readers. Virtually any dedicated business oriented graphics program such as Freelance Graphics would be more useful.

For the small number of plotting types it will produce, Origin's formatting capabilities are far superior to the other products represented here. It can handle more than three combinations of axes (3 X axes and 3 Y axes) with different scales. (Even three each is impossibly confusing.) Origin can create axis breaks, and text can be placed anywhere on the plot without limit. Straight or curved arrows, straight lines, rectangles, and circles can also be added. Virtually every individual feature of an Origin plot down to individual data points may be customized. Origin's text formatting is superior to some word-processors. It can also place multiple plots on one page including a magnified image of part of a plot.

Origin's time series plot automatically creates major and minor axis tick labels of seconds, minutes, hours, days, weeks, months, quarters, or years. Major tick marks can be in years and minor ones in quarters or months or any other combination for which there is sufficient space.

Origin's curve fitting capabilities are also the best among these programs. Its built-in curve-fitting functions include: linear regression; polynomial regression; four kinds of exponential decay; Gaussian and

multiple Gaussian; Lorentzian and multiple Lorentzian; and sigmodial. User-defined, curve-fitting functions may also be used. Origin has three curve-smoothing algorithms. It can also plot user-inputted functions and perform first or second order derivatives and integration.

Origin comes with a scripting language that can be related to screen objects in a manner familiar to users of hypertext, Visual BASIC, or other similar environments. The Origin language can be used to build highly attractive hypertext-like demos, tailor-made analytical graphics tools, and the like.

## Accuracy

Six accuracy tests with two data sets were performed on all six statistics programs. The first used the following basic data set:

| X | BIG | LITTLE | HUGE | TINY |
|---|------|--------|----------|--------|
| 1 | 99991 | .99991 | 10000000 | .00001 |
| 2 | 99992 | .99992 | 20000000 | .00002 |
| 3 | 99993 | .99993 | 30000000 | .00003 |
| 4 | 99994 | .99994 | 40000000 | .00004 |
| 5 | 99995 | .99995 | 50000000 | .00005 |
| 6 | 99996 | .99996 | 60000000 | .00006 |
| 7 | 99997 | .99997 | 70000000 | .00007 |
| 8 | 99998 | .99998 | 80000000 | .00008 |
| 9 | 99999 | .99999 | 90000000 | .00009 |

Reported in Table 4 are the results of the six tests. With five tests the maximum number of digits that a program could handle and correctly perform calculations on are reported. For example, SPSS could correctly plot HUGE vs. TINY when HUGE had 13 digits (1 through 9 followed by 12 zeros) and TINY 12 digits (a decimal point and 11 zeros followed by 1 through 9). At the other extreme, WinSTAT could barely produce a correct plot with only two digits for BIG vs. LITTLE. The BIG

vs. LITTLE plot was also difficult for SPSS, UNISTAT, and WinSTAT but represented no challenge for SYSTAT.

The mean and standard deviation were calculated for all variables in the data set except X as was a correlation matrix. In each case, X was excluded because it is only a single digit. Finally, a regression analysis of X on BIG and LITTLE was performed. Ideally, this analysis sould produce a warning message concerning the singularity of the model.

The final test was a calculation of the variance and standard deviation of the three numbers 8000001, 8000002, and 8000003. Zeros were added to each number until the programs produced incorrect results, refused to accept the numbers, or crashed.

The importance of the results to readers will depend on their individual needs, although UNISTAT and WinSTAT fall well behind the others.

## Documentation

Except for the LISREL manual, documentation that accompanies SPSS for Windows is the best we have seen for any statisics package. Each statistical technique is explained theoretically and with clear examples. The SPSS manuals also provide detailed instructions for operating the program itself, but since its use is so intuitive, the user will rarely need to use the manuals for this purpose except for the LISREL module, which offers a technique with which many users may be unfamiliar. The LISREL manual assumes that the user possesses a high level of expertise. It is almost completely lacking in the clear explanations found in all other SPSS manuals.

## TABLE 4
## Accuracy

| Max digits handled | MINI | SAS | SPSS | SYSTAT | UNI | WinSTAR | WinSTAT |
|---|---|---|---|---|---|---|---|
| Plot HUGE v. TINY | 18, 7 | 8, 8 | 13, 12 | 12, 11 | 7, 7 | 13, 12 | 9, 6 |
| Plot BIG v. LITTLE | 5, 4 | 8, 8 | 3, 3 | 12, 11 | 4, 2 | 13, 12 | 2, 2 |
| Basic statistics | 5, 4, 18, 4 | 8, 8, 8, 8 | 13, 2, 13, 2 | 6, 9, 12, 9 | 7, 7, 7, 7 | 8, 6, 13, 12 | 6, 6, 6, 6 |
| Correlation | 7, 6, 18, 7 | 8, 8, 8, 8 | 13, 10, 13, 12 | 12, 10, 12, 11 | 7, 6, 7, 7 | 8, 6, 13, 12 | 6, 6, 6, 6 |
| Var. & stnd. dev 8XXX1 | 7 | 8 | 13 | 13 | 7 | 13 | 6 |
| Reg. X on BIG & LITTLE | Warning | No warning | No warning | Warning | Warning | Warning | No Warning |

The SYSTAT manual is also superb, but it is not quite as helpful for someone unfamiliar with a particular technique. The MINITAB, UNISTAT, WinSTAR, and Win-STAT manuals are clearly written but spare. A user unfamiliar with a particular statistical technique will be obliged to consult a statistics text. WinSTAR's printed documentation is supplemented by an extraordinarily complete on-line help system. The Origin manual provides clear and complete explanations of the program's operations.

The set of manuals that supports the fully installed SAS system measure over a foot high when stacked. With few exceptions they are dismally written and poorly organized. This is documentation written by and for SAS mavens, not occasional users.

### Installation and Miscellaneous Irritants

UNISTAT is copy protected. It must be installed on a hard disk from original floppies, and it must be deinstalled before it can be put somewhere else. Furthermore, our attempt to deinstall the program to remove it from the hard disk produced only error messages. A hard disk crash or virus would leave the user without the program for the time it would take to ship new disks.

UNISTAT does not come close to taking full advantage of the Windows environment. It is essentially a DOS program in Windows clothing. It uses a few Windows features while suffering from its worst disadvantages, most notably sluggish performance. Its worst speed problems are not in calculations (it is one of the fastest programs in this regard), but in data entry where the user sees far too much of the infamous Windows hour glass.

SAS also appears to come to Windows grudgingly. Two extra cost modules (ASSIST and INSIGHT) must be installed before SAS resembles other Windows programs, and even then SAS's use of the Windows environment is often nonstandard, incomplete, and awkwardly rendered. To construct a SAS system that barely matches SYSTAT, the base SPSS program, or MINITAB, one must

install seven modules that require 65.9 megabytes of disk space. Making matters worse, SAS's poor screen designs, incomplete use of Windows conventions, and inadequate help screens require frequent forays into the stack of SAS manuals.

All programs except WinSTAT were easy to install to a standard hard disk or a Bernoulli disk.[4] When we attempted to install WinSTAT to a Bernoulli drive with a 90 megabyte capacity, WinSTAT informed us that it could not be used on a floppy disk and ended the procedure. We solved the problem by installing it on our fixed disk and copying it to the Bernoulli.

WinSTAT crashed under conditions that the other programs could handle. In the data input and editing screen, if one fails to press the Enter key after the last number is typed and then selects a statistics option, the program crashes to the Windows menu, and unsaved data are lost. The other programs either treat the last number as missing or make it obvious that the number has not been entered. Other WinSTAT crashes seemed to be caused by data sizes exceeding the program's constraints.

## Recommendations

We do not recommend UNISTAT. Its copy protection is a strong negative; data entry and editing are barely functional; and it is highly overpriced.

If the reader wants a Windows statistics package and is obliged to use SAS, he or she should demand that the ASSIST and INSIGHT modules be installed, although the typical user will need many other modules (with their manuals) as well.

WinSTAR and WinSTAT are very limited and overpriced. Several DOS statistics programs costing less are at least as functional. However, Win-STAR's $59 upgrade from its DOS product probably represents fair value.

Origin offers too few plotting techniques to recommend as a graphical analysis supplement to a standard statistics package. Its greatest strength lies in its language that allows the creation of demos and specialized tools.

MINITAB is a solid product in terms of performance and the variety of tools it provides. The Windows version is an enormous improvement over the previous DOS program, and the satisfied MINITAB user should probably take advantage of the reasonable $100 upgrade price. However, even at its academic discount price it is not as sound a value for the new purchaser as SYSTAT.

SPSS and SYSTAT for Windows are joys to use. The SPSS manuals should be considered the standard for the industry, but the SYSTAT manual is excellent as well. At $627 SYSTAT represents by far the best value among the programs represented here. If we were asked to make a purchasing decision with no funding constraints, we would choose the complete SPSS system. Site license prices and specific features needed by the analyst may often be the determining factors in the choice between them. Both are excellent products.

Despite the high quality of the SPSS and SYSTAT Windows renditions, many satisfied users of the DOS versions who cannot afford to shift to them (hardware as well as software costs may be involved) need not feel aggrieved that they cannot make the transition at this time. The Windows versions are fine products, but they represent only marginal improvements for most current DOS users.

---

## Notes

1. Carl Grafton and Anne Permaloff, "Statistical Packages for Personal Computers," *PS: Political Science & Politics* 16 (Spring 1983), 182-88.

2. Anne Permaloff and Carl Grafton, "Top of the Line: SPSS-PC, SAS, and SYSTAT," *PS: Political Science & Politics* 21 (Summer 1988), 657-66. Carl Grafton and Anne Permaloff, "Microcomputer Statistical Packages," *PS: Political Science & Politics* 21 (Winter 1988), 71-82.

3. Strictly speaking, Windows programs are also DOS programs, but it is common to distinguish between applications that operate directly with DOS and those that reside in Windows, which in turn interacts with DOS.

4. SAS had already been installed by the university SAS representative who encountered no difficulties in the long SAS installation.