

and relatively highly remunerative positions. Subsequently, either tenure itself will be formally abolished, replaced by some variant of rolling term-appointment, or the share of the faculty ensnared in a degenerate tenure system will undergo a long-term contraction toward a lower asymptote of zero.

About the Author

Stephen P. Dresch is professor of economics and dean of the School of Business and Engineering Administration, Michigan Technological University. Previously a research scholar at the International Institute for Applied Systems Analysis (Austria) and chairman of the Institute for Demographic and Economic Studies, he is the author (with K. R. Janson), most recently, of "Giants, Pygmies and the Social Costs of Fundamental Research, or, Price Revisited," *Technological Forecasting and Social Change*, v. 32, no. 4 (1988).

Notes

1. For an incisive analysis of the economic foundations of the system of academic tenure, see Armen A. Alchian. 1958. Private Property and the Relative Cost of Tenure. In P. D. Bradley, *The Public Stake in Union Power*, Charlottesville, VA: The University Press of Virginia, pp. 350-71, reprinted in Armen A. Alchian. 1977. *Economic Forces at Work*, Indianapolis: Liberty Press, pp. 177-202.

2. Of course, denial of tenure is not the equivalent of capital punishment: The termination of a probationary faculty member does not reduce the aggregate supply of faculty services unless the disappointed individual elects as a result to withdraw permanently from the academic market. This would be likely only in a weak academic labor market. Also, the depressive effect of adverse changes in tenure prospects on the supply of faculty services to the individual institution could be compensated by the offer of higher current remuneration; thus, the fact that tenure criteria did not become more stringent elsewhere need not imply a reduction in the supply of faculty to an institution employing this strategy.

Microcomputer Statistical Packages

Carl Grafton

Anne Permaloff

Auburn University at Montgomery

Our first article in this PS series, written five years ago, covered four microcomputer statistics packages (Grafton and Permaloff, 1983). Those programs represented a substantial fraction of the total number on the market at that time. Today there are probably 50 times that many, and virtually all of them offer many times greater speed and data handling capacity, enhanced data transformation capability, and a larger range of statistical techniques.

Here we discuss 15 packages for the IBM PC and compatibles, although several are sold in versions that will run on Apple II and Macintosh equipment. Because it is impossible to cover all the packages being sold, we have selected those whose earlier versions we found useful, others suggested by colleagues, along with "best sellers" and some that we judged from advertising to be potential bargains. Our objectives are to provide the reader with standards by which to evaluate this software, present an overview of the market, and evaluate these particular packages.

The packages examined are in the low to moderate price range. We have squeezed STATGRAPHICS and SYSTAT, the most expensive ones, into the moderate range because, unlike most statistics packages, they can be obtained at a substantial discount from mail order dealers. All except two can run on dual floppy disk systems, although several others perform much better on a hard disk. Most require a color graphics adaptor for all or some graphics displays. Only one is copy protected. Even the ones that are listed as able to run with less than 640K RAM (Random Access Memory) operate more effectively with 640K either because they are able to take advantage of full memory or because they can be used with RAM disk software. RAM disk software allows part of RAM to be set aside so that it resembles an additional disk drive. Since mail

The Profession

order software dealers carry few statistics packages, prices shown are retail from the manufacturer unless otherwise indicated.

The Programs

ABstat, version 5. AndersonBell, Inc., 11479 S. Pine Dr., Parker, CO 80134; (303) 841-9755. \$395. Additional copies half price. Student copies \$22.50 in quantities over 100. Requires at least 256K. A version can use 8087 coprocessor.

DYNA-STAT Professional, version 3.2. Dynamic Microsystems, Inc., 13003 Buccaneer Rd., Silver Spring, MD 20904; (301) 384-2754. \$195. Requires at least 256K.

NCSS, version 5.0. Jerry Hintze, 865 East 400 North, Kaysville, UT 84037; (801) 546-0445. \$99. Student version \$15-\$39 depending on quantity purchased. Requires at least 320K.

STATA, version 1.5 and STATA/graphics. Computing Resource Center, 10801 National Boulevard, Los Angeles, CA 90064; (213) 470-4341. STATA, \$395. STATA/graphics, \$590. Requires at least 256K. Can use 8087 coprocessor.

STATGRAPHICS, version 2.6. Statistical Graphics Corporation, 2115 East Jefferson St., Rockville, MD 20852; (800) 592-0050, (301) 984-5000. \$419 mail order discount. Requires at least 384K. Will run on a dual floppy system, but its frequent disk drive access makes a hard disk nearly a requirement for heavy usage. Can use 8087 coprocessor.

STATISTIX, version 1.1. NH Analytical Software, 1958 Eldridge Avenue, Roseville, MN 55113; (612) 631-2852. \$169. Requires 256K. A version can use 8087 coprocessor. Versions are available for Apple II computers with 64K or 128K at \$99. 20% discount on purchase of five or more copies.

StatPac Gold, version 2.0. Walonick Associates, 6500 Nicollet Avenue S., Minneapolis, MN 55423; (612) 866-9022. \$595. Requires a hard disk.

StatPlan III. The Future Group, 76 Eastern Boulevard, Glastonbury, CT 06033; (203) 633-3507. \$179. Requires at least 256K. A version costing \$189 can use 8087 coprocessor.

STATS-2, version 2.1. StatSoft, 2832 East 10th St., #4, Tulsa, OK 74104; (918)

583-4149. \$149. Requires at least 256K. Enhancement packages for utilizing 512K and/or an 8087 coprocessor are available for \$49 each.

Stat 1. Sugar Mill Software Corp., 1180 Kika Place, Kailua, HI 96734; (800) 367-8047, ext. 261. \$129.95. Student version \$29.95.

SYSTAT, version 3 with the following supplementary modules: DESIGN, LOGIT, PROBIT, TOBIT, TESTAT, and REPORT WRITER. SYSTAT, Inc. 2902 Central St., Evanston, IL 60201; (312) 864-5670. Requires at least 512K. \$439 mail order discount. Most modules \$90 each. Can use an 8087 coprocessor. Also available for the Macintosh, VAX, and many others.

Features and Statistical Functions

Those new to microcomputer statistics package evaluation might be inclined to focus most of their attention on the variety of statistical functions contained in them, but other characteristics are important as well. Once it has been determined that a number of packages offer the required statistical functions, one's choice among them should be determined by such qualities as the size of data sets that can be handled, data input and editing features, data transformation commands, ease of learning (or relearning after long periods of nonuse), graphics capabilities, and price. When both statistical function variety and such features are taken into account, one finds something like a normal distribution between the extremes of packages that are not worth buying in their present form and those that are both powerful and useful.

Four Nonstarters

Four programs are neither listed above nor found on the tables summarizing features and statistical functions because basic flaws ruled them completely out of consideration. They are:

FAKAD, K. I. Macdonald, Nuffield College, Oxford, OX1 1NF, England, 28 pounds. Requires at least 256K. A version can use 8087 coprocessor.

PC-SCA (now PC-UTS). Scientific Com-

puting Associates, 4513 Lincoln Ave., Lisle, IL 60532; (312) 960-1698. \$595. Copy protected using "install-uninstall technique." Requires 640K for practical use and hard disk. Can use 8087 coprocessor.

True Epistat, version 1.0. Epistat Services, 2011 Cap Rock Circle, Richardson, TX 75080-3417; (214) 680-1376. \$249. Requires at least 256K. Will run on a dual floppy system, but a hard disk is better. Can use 8087 coprocessor.

SL-MICRO. Questionnaire Service Company, Box 778, East Lansing, MI 48823; (517) 641-4428. A version is available that runs under CP/M and CP/M-86. \$250. The manual may be purchased separately for \$15. IBM-PC version requires only 128K, CP/M version 48K, and CP/M-86 version 128K.

The main problem with FAKAD was that almost none of its features worked on the first attempt to use them, and some functions never worked. To take the simplest example, the user is told that to turn off the reverse video default one need only key in the command FLIP. This command produced an error message and never worked. Data input was accompanied by many confusing error messages, and there was very little relationship between the manual's sparse instructions and what appeared on screen.

SL-MICRO works, but it would have been out of date even by the standards applied five years ago. SL-MICRO operates entirely by batch runs. Data are entered in one file using a text editor or spreadsheet, and a program using the relatively complex and nonintuitive SL-MICRO syntax is written with a text editor and saved as a second file. Then SL-MICRO is run working off the two files. Mistakes in the batch file require that it be rewritten and resaved. Batch processing can be a useful feature, especially if complex and time consuming runs are to be made. But we cannot recommend any statistics package that does not permit interactive operation and offers only batch processing unless it contains other offsetting features such as unique statistical functions or an especially low price. SL-MICRO offers a solid but not extraordinary array of statistical functions, but at \$250, it is consider-

ably more expensive than several more convenient to use packages.

PC-SCA had one major strike against it before we began to experiment with it. It uses an install-uninstall copy protection scheme. The program itself performed extraordinarily well in accuracy stress tests (described below). Its manual, however, is one of the worst we have ever seen. It is poorly organized and explanations are fragmentary. We were able to get the program to read a simple space-delimited ASCII file only after hours of experimentation, but we never succeeded in saving those data to disk in SCA's own format. Its manufacturer tells us that a step-by-step manual section written especially for the PC user is in development. With such a completely rewritten manual and the removal of copy protection, this program would be worth consideration, but we cannot recommend it in its present condition except perhaps to those who have used it successfully in its mainframe form.

True Epistat was removed from consideration for the reasons discussed below.

Crashes

It is difficult to believe that programs are still being sold that can easily crash, but True Epistat and StatPlan III are crash prone. True Epistat is so bad in this regard that we cannot recommend its purchase. For example, when we asked it to read a nonexistent file, True Epistat crashed rather than simply informing the user that the file did not exist. It also froze and crashed after successfully reading an ASCII file and saving the data in its own format. It crashed again in the course of doing a correlation and again when generating a plot. Other crashes also occurred.

StatPlan III is not nearly as crash prone as True Epistat, but the accuracy stress test caused it to crash when it tried to calculate statistics on or plot large or small numbers. One data set caused it to crash when it tried to calculate the standard deviation for the numbers 800001, 800002, and 800003. It handled five-digit versions of these numbers (with single zeroes removed).

NCSS will crash if its file handling pro-

cedures are not strictly adhered to, and STATS-2 will freeze if the printer is switched off without first informing the software, but these are not sufficiently serious problems to rule these packages out of consideration.

None of the other packages crashed for any reason.

Command Driven versus Menu Driven

There are two basic kinds of statistics packages, command and menu driven. Command-driven programs (SAS and SPSS are classic examples) are operated through combinations of words entered interactively or in batch form. Spelling and syntax must be correct, making such programs considerably more demanding to operate than menu-driven packages that require only that choices be made from lists of procedures. Although this distinction is still largely valid, it is becoming increasingly indistinct. For example, several menu-driven packages use a command format for complex data transformations. On the other hand, some essentially command-driven packages such as STATA and StatPac Gold contain menu options that can be used by novices.

Someone who uses a program infrequently is likely to prefer a menu-driven package. Learning and relearning menu choices is relatively easy if the menus are well designed. Indeed, all of the menu-driven programs represented here can be operated by a first time user with little reference to their manuals. Frequent and advanced users almost invariably prefer command-driven packages since they tend to offer a greater variety of statistical techniques and one is not obliged to wait for a series of menus to clear the screen before data analyses can occur.

The command-driven programs examined here (ABstat, StatPac Gold, SYSTAT, and STATA) have clear and easy-to-use syntaxes. Operating these programs is made even more convenient by clearly written manuals. The prospective purchaser who will be an infrequent user should not be put off by the fact that a program is command driven per se; how-

ever, the quality of the manual becomes more important with a command-driven program.

Among the command-driven programs, we prefer SYSTAT and STATA in terms of ease of use. They offer a variety of built-in functions comparable to the others, but their elegant and logical commands constitute a language that can be used to develop additional functions. SYSTAT incorporates its own version of BASIC as still another mechanism for customizing the package.

Statistical Techniques

Tables 1-3 list the statistical techniques offered by the 11 packages considered usable. FAKAD, PC-SCA, SL-MICRO, and True Epistat are excluded for the reasons discussed above. The reader is cautioned that these techniques are categorized in a variety of ways by software manufacturers and statistics textbook authors. Prospective purchasers should use the data in these tables only to narrow their focus to a few packages. At that point descriptive literature should be obtained from the manufacturers. Then it may be necessary to follow up with telephone calls, because some of the manufacturers' terminology may not be the same as the prospective purchaser's.

Table 1 covers major statistical categories. Many are common to most packages, but the reader is cautioned again that a phrase such as "correlation and multiple regression" may encompass a wide variety of techniques. Tables 2 and 3 focus in greater detail on those techniques likely to be of particular interest to political scientists. These include descriptive statistics, correlation and multiple regression, ANOVA and related techniques, and crosstabs. Even these relatively detailed breakdowns may disguise gaps or obscure features.

Some packages offer LOGIT and PROBIT as modes of analysis; others include them in the data transformation procedures. In the latter case, the transformation restructures the data set and then allows one to use the regression module for the analysis. These techniques are

Table 1. General Statistical Techniques

	ABstat	DYNA- STAT	NCSS	STATA	STAT- GRAPHICS	STATIX	StatPac Gold	StatPlan III	STATS-2	Stat I	SYSTAT
ANCOVA	N	Y	N	Y	Y	Y	N	Y	Y	Y	Y
MANOVA	N	N	N	N	Y	N	N	N	N	Y	Y
Cluster analysis	N	N	Y	N	Y	N	N	N	N	N	Y
Discriminant analysis	N	Y	Y	N	Y	N	N	N	N	N	Y
Factor analysis	N	Y	Y	N	Y	N	Y	N	N	N	Y
Non-parametric tests	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y
Sample size determination	N	N	N	N	Y	N	N	N	N	N	Y
Time series	N	Y	Y	N	Y	N	Y	Y	N	N	Y

All have ANOVA, correlation, multiple regression, crosstabs, descriptive statistics, F-tests, and t-tests.

Key: Y = Yes, N = No.

Table 2. Regression Techniques

	ABstat	DYNA- STAT	NCSS	STATA	STAT- GRAPHICS	STATIX	StatPac Gold	StatPlan III	STATS-2	Stat I	SYSTAT
Nonlinear regression	N	Y	N	N	Y	N	Y	Y	N	N	Y
Nonparametric correlation	Y	Y	Y	N	Y	Y	Y	N	Y	N	Y
Residual analysis	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Ridge regression	N	N	N	N	Y	N	N	N	N	N	N
Robust regression	N	N	Y	N	Y	N	Y	N	N	Y	Y
Stepwise regression	N	Y	Y	N	Y	N	Y	N	Y	Y	Y
LOGIT*	N	N	N	Y	N	N	N	N	N	N	Y
PROBIT*	N	N	N	Y	N	N	Y	Y	N	N	Y

They all have simple and multiple regression and standard correlation and regression statistics.

Key: Y = Yes, N = No, * = Accomplished directly on the data.

Table 3. Frequencies and Crosstabulation

	ABstat	DYNA- STAT	NCSS	STATA	STAT- GRAPHICS	STATIX	StatPac Gold	StatPlan III	STATS-2	Stat I	SYSTAT
Frequencies	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
2-way table	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Multi-variable	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Maximum number	*	—	*	*	*	5	*	*	3	*	*
Created from raw data	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Variable labels	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y
Value labels	Y	N	Y	Y	Y	N	Y	Y	N	Y	Y
Cell display format											
Raw numbers	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Column percentage	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y
Row percentage	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y
Cell percentage	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y
Expected frequencies	N	Y	Y	N	N	N	N	N	N	Y	N
Weighted variables	N	N	Y	Y	Y	Y	N	N	N	Y	Y

All have Chi-square.

Key: Y = Yes, N = No, * = Memory dependent. (All but Stat I can use sort or SELECT IF options to increase table size.)

Table 4. Graphics

	ABstat	DYNA- STAT	NCSS	STATA	STAT- GRAPHICS	STATIX	StatPac Gold	StatPlan III	STATS-2	Stat I	SYSTAT
Box-and-whisker	N	N	Y	Y	Y	N	Y	N	N	N	Y
Notched box- and-whisker	N	N	N	Y	Y	N	N	N	N	N	Y
Casement	N	N	N	Y	Y	N	N	N	N	N	S
Contour	N	N	N	N	Y	N	N	N	N	N	N
Draftsman	N	N	N	Y	Y	N	N	N	N	N	S
Factor (analysis) plot	N	N	Y	N	Y	N	N	N	N	N	Y
Log	N	N	N	N	Y	N	Y	N	N	N	Y
Pie charts	Y	N	N	Y	Y	N	Y	N	N	N	N
Scatter plots	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Standard error bar	N	N	N	Y	Y	N	N	N	N	N	N
Star symbol	N	N	N	N	Y	N	N	N	N	N	N
Three-dimensional bar	N	N	N	N	Y	N	N	N	N	S	N
Three-dimensional scatter	N	N	N	N	Y	N	N	N	N	N	N
Presentation quality?	Y	N	N	Y	Y	N	S	Y	N	N	N

They all have bar charts and histograms.

Key: Y = Yes, N = No, S = Some.

listed both in Table 2 and later in Table 6 which covers data transformations.

SYSTAT and STATGRAPHICS offer the largest variety of statistical techniques useful to political scientists, although several other packages would serve many applications just as well. It is impossible to rank these remaining packages because individual users' requirements vary so widely. It is clear, however, that there is little relationship between cost and the variety of statistical techniques offered.

Graphics

Statistics packages vary widely in the variety and quality of plots that they can produce. Here, even more than with statistical techniques, differences in terminology can be highly confusing, and some graphics tools may not be as familiar to readers as are statistical techniques. Less common graphics tools include:

1. Box-and-whisker and notched box-and-whisker. A box-and-whisker plot consists of a rectangle that encompasses the middle portion (usually 50 percent) of data values together with two lines (whiskers) that extend from the box to the extremes of the data. A notched box-and-whisker

plot includes a notch in the box the width of which indicates a confidence interval for the median. The width of the box itself is proportional to the square root of the number of observations in the data set. The notched box-and-whisker plot is especially useful for multiple data set comparisons.

2. Casement. A series of two-variable scatterplots in groups determined by the values of a third and/or fourth variable.

3. Contour. Similar to a three-dimensional bar chart but printed like a topological map.

4. Draftsman. A series of two-variable scatterplots for all combinations of three or more variables.

5. Factor (analysis) plot. Shows which variables make up which factors.

6. Standard error bar. Each point in the plot is the mean of a data set, and the point is the intersection for lines showing standard error for both variables.

7. Star symbol. A multivariate plotting technique. Each star represents one observation, and a star consists of rays drawn from a center point. Each ray represents one variable.

Statistics packages offer two kinds of plots: character and pixel. Character

Table 5. Accuracy and Speed

	ABstat	DYNA- STAT	NCSS	STATA	STAT- GRAPHICS	STATIX	StatPac Gold	StatPlan III	STATS-2	Stat I	SYSTAT
Accuracy:											
8 x 9 data set (large & small numbers)											
plots	5	1	1	5	5	4	1	1	3	1	5
descriptive statistics	5	3	1	5	5	2	1	2	1	4	5
regression	5	3	1	5	5	2	1	1	2	4	5
Bureau of Standards regression											
regression	5	2	1	4	3	1	3	1	1	1	5
Key: 5 = Accurate, 3 = Incomplete, 1 = Complete breakdown											
Variance and standard deviation, large number (Max. digits read)											
	11	7	6	7	14	9	8	5	5	10	15
Speed (in seconds):											
Reading an 18 x 159 ASCII file											
	47	215	205	45	320	97	—	73	—	500	205
Multiple regression on the 18 x 159 data set											
	513	250	164	90	615**	170	—	177	135**	390	325

*Tests performed on an AT clone with hard disk.

**Accepted only a 14 x 159 data set.

graphics are made up of the basic ASCII character set such as asterisks and lines. These are adequate for analytical purposes. They run quickly on a printer, but they may not be of presentation quality. Pixel graphics have the potential to be of higher resolution. They can be of presentation quality if the programmer has a good design sense and if the software allows the plot to be reproduced on a printer or plotter.

STATGRAPHICS is by far the best program in terms of graphics. It offers the greatest variety of two-dimensional plots. It allows three-dimensional bar, scatter, and contour plots as well. Recent versions of STATGRAPHICS feature printer and plotter drivers that produce presentation quality graphs. STATGRAPHICS also includes a graphics editor that permits the user to modify the wording, location, and size of default titles and to add new titles.

Most statistics packages offer basic plotting routines such as scatter plots and bar

charts together with a few extras such as box-and-whisker. Compared to STATGRAPHICS, STATA/graphics places a poor second with a modest variety of presentation quality graphs. SYSTAT ranks just behind STATA/graphics with comparable variety but no presentation quality output. The other programs are tied for last. Recently, NCSS added an optional presentation quality graphics module costing \$59.00. We have not had a chance to test it, but its specifications suggest that it offers considerably greater variety than STATA/graphics.

Accuracy and Speed

All of the statistics packages reviewed here can produce accurate results with data that are not troublesome. However, when required to perform analyses on relatively large or small numbers, wide variations in performance are revealed.

These variations might render particular programs useless for some applications.

We used three sets of data to test for data handling capability. The first was a table of eight variables and nine observations containing numbers ranging from single-digit to 12-digit integers and 12-digit fractions (Wilkinson). A variety of operations were performed on these data including descriptive statistics, regression, and plots.

The second data set was used by the National Bureau of Standards to evaluate multiple linear regression programs (Wampler, 1978). This data set is as follows:

200,000	1.0	1.00001	0.99999
600,000	1.0	2.00001	1.99999
400,000	1.0	1.99999	2.00001
400,000	1.0	0.99999	1.00001

The third data set consisted of three observations: 8,000,000,001; 8,000,000,002; and 8,000,000,003 (Lehman, 1987). The test consisted of calculating the variance and standard deviation (1.0). If the program could calculate these statistics, a zero was added to each number, and if it could not calculate the statistics, a zero was removed from the numbers and the calculation repeated. This process was continued until the largest number for which the statistics could be correctly calculated was determined. Table 5 records this figure.

SYSTAT, STATGRAPHICS, ABstat, and STATA produced perfect or almost perfect results for all the accuracy tests. SYSTAT could work accurately with the largest numbers (15 digits) in the third data set, and STATGRAPHICS could handle 14 digits. The other packages produced scores substantially inferior to STATGRAPHICS, SYSTAT, ABstat, and STATA.

We tested speed by reading a space delimited ASCII file containing 18 variables and 159 observations and by performing a multiple regression using all variables in the file. Times for these operations are shown in Table 5.

Time required for file reading and regression analysis are not burdensome even for the slower packages, although STATGRAPHICS with its 320-second time

for reading the ASCII file and 615-second time for performing a regression analysis on a 14-variable data set is near the limit of what we would tolerate. These times were recorded on an all-floppy Zenith 150, the equivalent of an original IBM-PC running at the slowest possible speed. An 8087 coprocessor was not used. Newer equipment, especially a PC-AT or equivalent with a hard disk drive reduces these times drastically.

Data Base Capabilities

Virtually all statistics packages allow the user to enter, edit, and store data using a module contained in the statistics package itself. The best of these, STATGRAPHICS, StatPlan III, NCSS, and SYSTAT (in order from first to fourth place), all use a spreadsheet format that allows free cursor movement to enter data by variable or case and also allow for convenient correction, addition, and deletion of data. The STATGRAPHICS and StatPlan III spreadsheets work best, but they are not as responsive to user needs as Lotus 1-2-3. NCSS also uses a spreadsheet format, but it works interactively with the disk on which data are stored. Every time the cursor is moved to a row or column that is off screen, the disk drive runs. On an all-floppy system this results in a very annoying delay unless data are stored on a RAM disk. RAM disk use makes NCSS a fast and effective data management tool, but one must be careful to copy data to be kept permanently from the RAM disk onto a real disk at the end of a session. The SYSTAT data and editing screen is balky but adequate for editing or short data entry sessions.

During data input only STATGRAPHICS, StatPlan III, NCSS, and SYSTAT display data in a neat tabular format while allowing free cursor movement (and editing) around the data base. For example, the relatively expensive STATA displays data observation by observation in a format not conducive to proofing. However, basic editing can be done with all of the packages.

StatPac Gold contains an unusual module that allows the user to design data

Table 6. Data Base Capabilities

	ABstat	DYNA- STAT	NCSS	STATA	STAT- GRAPHICS	STATIX	StatPac Gold	StatPlan III	STATS-2	StatI	SYSTAT
Data Base Capacity:											
Max. number of variables	128	100	250	99	—	120	500	RAM set	NA	RAM set	200
Max. number of observations	16,000	16,000	32,000	24,000	—	32,600	32,767	RAM set	NA	RAM set	32,000
Max. number of data points	1,920,000	16,384	8,000,000	239,000	8,500	3,912,000	16,383,500	50,000	8,000	50,000	6,400,000
Export ASCII:											
Unformatted data files	Y	Y	Y	Y	Y	Y	Y	N	Y	N	Y
Unformatted analysis results	Y	N	Y	Y	Y	N	Y	N	N	N	Y
Formatted data/results	S	N	Y	N	Y*	N	N	N	N	N	N**
Export DIF files	N	N	N	N	Y	N	N	N	N	N	N
Export WKS files	N	N	N	N	Y	N	N	N	N	N	N
Import ASCII:											
Blank delimited	Y	N	Y	Y	Y	Y	N	Y	Y	Y	Y
Comma delimited	Y	N	Y	Y	Y	N	Y	N	N	N	Y
Organized by either variable or observation	N	N	N	Y	N	N	N	N	N	N	N
Nonnumeric & variable names	Y	N	Y	Y	Y	N	N	Y	N	Y	Y
Import DIF	N	Y	N	N	Y	N	N	N	N	N	N
Import Lotus WKS	Y	N	N	N	Y	N	N	Y	N	N	N
Import dBASE III	Y	N	N	N	Y	N	N	N	N	N	N
Ease of importation	3	2	5	5	5	5	1	5	5	4	5

Key: 5 = Best, 1 = Worst. Y = Yes, N = No, S = Some.

*Allows internal editing of all exported files.

**Available in optional Report Writer.

entry screens so that they can be used by those unfamiliar with the rest of the program. This might be particularly useful for keying in questionnaire data. Unfortunately, the StatPac Gold data editing capabilities are as primitive as those offered by the majority of statistics packages.

We have yet to find a statistics package data editor that is as convenient for data entry and editing as a Lotus style spreadsheet. Spreadsheets allow freedom to enter data in any order, view it by simply moving the cursor across the screen, move data, and insert columns and rows to make room for new variables and cases. They also allow for simple data transformations (such as adding, subtract-

ing, dividing, and multiplying columns) through their formula commands.

Data keyed into a spreadsheet can be saved in its own format (e.g., with Lotus 1-2-3 a WKS file) or in ASCII in a space delimited format (e.g., a Lotus PRN file). Most statistics packages can read a space delimited ASCII file; some can read a WKS file as well. On the other hand, most statistics packages cannot read each other's files unless they can create ASCII files. So if more than one statistics package is to be used, data entry on a spreadsheet is essential. It is also desirable for a statistics package to be able to read variable names as well as numeric data, although this is a much less common feature. ASCII files

may be organized by variable or observation, and a statistics package should be able to read either, but this is an even more unusual feature. ASCII files may also be in a DIF format or a comma separated format (Miller, 1983). The greater the variety of files that can be imported the better.

In terms of the variety of file formats they can import and the ease with which importation can be performed STATGRAPHICS and ABstat are the clear winners. They can import an ASCII space delimited format or comma separated file. They can handle both numeric and non-numeric data as well as variable names, and they can also import Lotus WKS and dBASE III files. STATA also offers several file format choices.

File importation should be only a matter of keying in the name of the file to be imported and perhaps informing the statistics package how many columns and rows are contained in the data base. In fact, STATS-2, NCSS, SYSTAT, STATGRAPHICS, STATA, STATISTIX, and StatPlan III import ASCII files this easily. At the opposite extreme, ABstat and StatPac Gold require detailed specifications such as the columns at which each variable begins and ends. This is time consuming and the possibility for making mistakes is very high. We must hasten to add that ABstat can read a Lotus WKS file as well as other formats with much less difficulty.

Table 6 also lists data handling capacities. The maximum number of variables and observations that each program can handle is listed. These are absolute limits. Thus, for example, DYNA-STAT can recognize no more than 100 variables and no more than 16,000 observations. Sometimes the total number of data points that can be handled is the maximum number of variables multiplied by the maximum number of observations, but this is not always the case. For example, DYNA-STAT will recognize only 16,384 data points and not the 1,600,000 that might be expected, so it would never be possible to meet DYNA-STAT's 100 variable and 16,000 observation limits simultaneously. STATGRAPHICS is by far the worst program in this category. It can handle a total of only approximately 8,500 data points

(e.g., 17 variables \times 500 observations).

Few of the programs can perform all statistical operations on the maximum number of variables that can be stored in their data bases. For example, STATS-2 and STATGRAPHICS can perform a multiple regression analysis with no more than 14 independent variables.

Data Base Transformations and Utilities

In addition to being superior tools for data entry and editing, spreadsheets can be conveniently employed to perform many data transformations. However some transformations such as those listed in Table 7 are more easily performed using the statistics packages that offer them than through programming a spreadsheet. These include:

1. $\langle \rangle =$. Used to transform data according to logical expressions.
2. Logit— $\text{Log}(X/(1-X))$ where X is a value between 0 and 1.
3. Probit—Given a value, X , from the unit normal distribution, this transformation gives the probability of being less than or equal to X . As with the Logit transform process this simply prepares the data for regression analysis.
4. Standardize. First the mean and standard deviation of each selected variable are calculated. The result is then calculated using Mean X /Standard Deviation.
5. Z-Score. The result is the expected value of the observation if it had come from the normal distribution.

These transformations should be usable in combination or in a single command. SYSTAT and NCSS offer the greatest variety of transformations. Most of the other programs offer all the transformations that most users would require, and none is difficult to implement.

Conclusions

One can make a statistics package purchase that will be more a source of frustration than assistance. It is easy to buy programs that are far too difficult if not impossible to operate or that crash frequently. Furthermore, the enormous variety of

Table 7. Transformations and Data Base Utilities

	ABstat	DYNA- STAT	NCSS	STATA	STAT- GRAPHICS	STATIX	StatPac Gold	StatPlan III	STATS-2	Stat I	SYSTAT
()	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y
<>=	Y	N	Y	Y	Y	Y	Y	N	Y	Y	Y
If-Then-Else	N	N	Y	Y	Y	Y	N	N	N	N	Y
LOGIT— Log(X/1-X)	N	N	Y	N*	N	N	N	Y	N	Y	N*
PROBIT (inverse normal, probability =X)	N	N	Y	N*	N	N	N*	N*	N	N	N*
Moving average smoothing	Y	Y	Y	N	N**	N	N	Y	N	N	Y
Running medians smoothing	N	N	Y	N	N	N	N	N	N	N	Y
SELECT IF or equivalent	Y	Y	Y	Y	Y	Y	Y	N	Y	N	Y
Standardize (mean X/ standard deviation)	Y	N	Y	N	N	Y	N	N	N	N	Y
Transpose (exchange rows and columns)	N	N	Y	N	N	N	N	N	Y	N	Y
Z-Score, expected normal, inverse normal, quartile	N	Y	Y	N	N	Y	Y	Y	N	N	Y

*Handled directly from raw data. No transform needed.

**Available as option in time series analysis.

Key: Y = Yes, N = No.

data handling and statistical techniques and terminology make it impossible to evaluate every important aspect of a package's operation via articles such as this one, manufacturers' specifications, or even long conversations with a manufacturer's technical staff. The only sure way to evaluate a statistics package is to do so oneself, but this is not easily accomplished unless testing can be done in a software retailer's establishment (difficult because retailers rarely stock statistics packages) or through a money back guarantee (offered only by the manufacturer of STATISTIX among these programs). Despite these problems, we can give some relatively clear guidance.

SYSTAT offers substantially more statistical computing power than the other programs analyzed here, and the difference between SYSTAT and the others is even greater than is suggested by the data presented in our tables. Within nearly every category SYSTAT has more features and allows the user to make more choices. This power is achieved using a clear and

logical command language that can be supplemented with SYSTAT BASIC (a standard feature). Furthermore, SYSTAT has a solid array of two-dimensional graphics although they are not of presentation quality. SYSTAT is also the most accurate of the programs and is capable of performing calculations on larger and smaller numbers than any of the other programs described here. Finally, SYSTAT operates with reasonable speed on an all-floppy system.

Despite SYSTAT's computing power superiority, it is inferior to STATGRAPHICS in several ways. First and foremost, STATGRAPHICS contains an array of presentation quality graphics probably unequaled in the entire statistics package market. Indeed, it would be an impressive tool if all it did was graphics. Its three-dimensional plots give the analyst a view of data that is difficult to imagine with comparable two-dimensional plots. Its graphics editor allows titles and labels to be modified in many ways. Similarly, its text editor

allows screens containing raw data and results to be modified for presentation quality printing. STATGRAPHICS also offers the second greatest variety of statistical techniques among the programs discussed here. It can read the greatest variety of file formats more easily than the others. The fact that it can read a WKS file directly means that it is not necessary to save a spreadsheet in ASCII format thus saving disk space and time. Highly experienced users can also write procedures for STATGRAPHICS in the manufacturer's version of the language APL (a separate purchase). But with all its strengths, this program's severely limited data handling capacity might remove it from consideration for many.

We would recommend SYSTAT or STATGRAPHICS to anyone whose applications fit these programs' specifications, and this would include a large majority of political scientists. However, at over \$400 they might exceed some budgets.

If its inability to handle large numbers without their being down scaled is not a problem, NCSS at \$99 represents an outstanding value. It exceeds the others by most standards, and it is one of the more likable. Its menu screens are very clear, and it contains many unusual extras, such as the ability to plot raw data or calculated results without leaving a particular part of the program. Its greatest operational weakness is its frequent accessing of the disk drive on which data are stored. This can be ameliorated by using RAM disk software. Given the performance of this package and the appearance of the presentation quality graphics in earlier versions, we expect that the new NCSS graphics supplement which includes three-dimensional plots will make this an even more useful program.

Except for the nonstarters, all the other programs are useful tools and two have features worth mentioning here. StatPac Gold's data entry customizing feature could be quite useful for some applications, and its statistical and transformation capabilities are quite substantial. The StatPac Gold manual is not as clearly written as most represented here, but a newly written tutorial is quite helpful.

STATA's commands constitute a lan-

guage that can be used to develop highly impressive statistical functions beyond those provided with the program. Many of its other capabilities, discussed above, also make this another good choice.

References

- Grafton, Carl and Anne Permaloff. 1983. "Statistical Packages for Personal Computers," *PS*, 182-88.
- Lehman, Richard S. 1987. "Statistics on the Macintosh," *Byte*, 207-14.
- Miller, David. 1983. *IBM Data Files: A BASIC Tutorial*. Reston: Reston Publishing Company, 136-45.
- Wampler, R. 1978. "Test Problems and Test Procedures for Least Square Algorithms." *National Bureau of Standards*.
- Wilkinson, Leland. "Statistics Quiz," Systat, Inc., 1800 Sherman Ave., Evanston, IL 60201.

Finding Jobs: Placement of Political Scientists in 1985-86

J. Brinton Rowdybush

American Political Science Association

The placement success rate of new political science Ph.D.s continued to be in 1986 what it has been since the mid 1970s: approximately seven out of ten new Ph.D.s were successful in finding employment with over 90% of them taking positions in academia. What is remarkable about the job market for political scientists is its continuing stability, with only marginal changes during the past decade in the placement success of firm candidates (Table 1). There have not yet been any major shifts in the numbers of men or women looking for employment or the number of employment opportunities.

It is true, however, that there have been some changes in the job market. Over the past five years more than two-thirds of job placements have gone to Ph.D.s and the 1986 figure of 69% confirmed this trend (Table 1). This is contrasted to the 1960s and 70s when at least half of new faculty