UNIX|STAT: Data-analysis programs

GARY PERLMAN

AT&T Bell Laboratories, Murray Hill, New Jersey

UNIX|STAT is a set of UNIX¹ data-manipulation and data-analysis programs developed at the University of California, San Diego. Although many of the people involved with UNIX|STAT have left San Diego, and development has stopped, the software is still available. This article contains a summary of the programs in the current version, some of which have been discussed previously (Perlman, 1980, 1981, 1982, 1983, in press).

UNIX|STAT is easy to use and has a negligible training period. It features simple input formats (free-format field-oriented) and flexible data manipulation. Data-validation programs allow range and type checking. Documentation support includes on-line manual entries and tutorials. The source code is written in C (Kernighan & Richie, 1978), and many modular C functions allow extensions to the system. UNIX|STAT is small enough for microcomputers because the programs are in the 10-25KB range. The programs run on most versions of the UNIX system (V6, V7, 2.8BSD, 4BSD, III.0, 5.0, and others) dating back to 1975. UNIX|STAT is public-domain software and has been in use in 12 countries at about 200 UNIX system sites since its introduction in 1980.

Functionality often built into statistical packages (e.g., graphics, sorting, and other data manipulation) is not reinvented in UNIX|STAT, which delegates such responsibility to standard UNIX system tools. The programs included in the distribution tape are summarized in Table 1.

Distribution. The programs have been released to the public and are distributed to anyone who wants them. A distribution fee is charged to cover costs (\$20.00 U.S.). Documentation for the package is available free of charge. The distribution includes a magnetic tape with the C source files for all the programs, the documentation source files, and a collection of test examples. Special tape format requirements should be indicated; otherwise, tar 1,600 bpi format is assumed. Requests for the distribution should be sent with a check payable to Gary Perlman at AT&T Bell Laboratories 5D-105, 600 Mountain Avenue, Murray Hill, New Jersey 07974. Other information can be obtained by calling (201)

Many people have contributed source and suggestions to UNIX|STAT. Jay McClelland, Mark Wallen, Greg Davidson, Don Gentner, and Don Norman have all made substantial contributions. The author's mailing address is: AT&T Bell Laboratories 5D-105, 600 Mountain Avenue, Murray Hill, NJ 07974.

Table 1
Data-Transformation and Analysis Programs

Program	Purpose
Data-Transformation Programs	
abut	abut data files
dm	column oriented data manipulator
io	control and monitor input and output
maketrix	create matrix type file from free-form file
perm	randomly permute lines in a file
repeat	repeat a pattern or file
reverse	reverse lines and characters
series	print a series of numbers
transpose	transpose matrix type file
Data-Analysis Programs	
anova	multifactor ANOVA (repeated measures and unequal cells)
calc	interactive algebraic modeling calculator
critf	critical F ratio given probability and degrees of freedom
desc	descriptive statistics, histograms, frequency tables
dprime	signal detection d' and beta calculations
pair	paired data statistics, linear regression, scatterplots
pof	probability of F ratio given degrees of freedom
regress	multivariate linear regression
ts	time-series analysis and plots
validata	verify data file consistency
vincent	time-series comparison

582-3624 or by using electronic mail to the following addresses:

{research|ucbvax}!ulysses!gsp, sdcsla!perlman {@nprdc}.

REFERENCES

- KERNIGHAN, B. W., & RICHIE, D. M. (1978). *The C programming language*. Englewood Cliffs, NJ: Prentice-Hall.
- PERLMAN, G. (1980). Data analysis programs for the UNIX operating system. *Behavior Research Methods & Instrumentation*, 12, 554-558.
- PERLMAN, G. (1981). An example of cooperating compact data analysis programs. *Behavior Research Methods & Instrumentation*, 13, 290-293.
- PERLMAN, G. (1982, July). *UNIX|STAT: Compact data analysis programs for UNIX*. Paper presented at the USENIX Conference, Boston, MA.
- PERLMAN, G. (1983). Data analysis in the UNIX environment: Techniques for automated experimental design specification. In K. W. Heiner, R. S. Sacher, & J. W. Wilkinson (Eds.), *Computer science statistics: Proceedings of the 14th Symposium on the Interface*. New York: Springer-Verlag.
- PERLMAN, G. (in press). UNIX|STAT: Compact data analysis programs for the UNIX operating system. *American Statistician*.

NOTE

1. UNIX is a trademark of AT&T Bell Laboratories. UNIX|STAT is *not* a product of Bell Laboratories. UNIX|STAT is unsupported, although known bugs have been removed. UNIX|STAT may not be distributed for personal or material gain.

(Manuscript accepted for publication February 24, 1984.)