submitted to *Human and Ecological Risk Assessment*

Review of Systat for Windows, Version 7.0

David E. Burmaster
Alceon Corporation
POBox 382669, Cambridge, MA 02238-2669
tel: 617-864-4300; fax: 617-864-9954; email: deb@Alceon.com

Program:	Systat for Windows, Version 7.0
Source:	Systat, a Division of SPSS, Inc.
	444 North Michigan Drive
	Chicago, IL 60611
	Tel: 312-329-2400; fax: 312-329-3668
	http://www.spss.com
System:	Windows 3.1, Windows 95, or Windows NT 3.51 or higher
	80386 processor or higher, 8MB RAM or more recommended,
	25MB available disk space with a swap file,
	VGA or higher graphics adapter
Cost:	Commercial price, $995. for program plus manuals
	Academic price, $695. for program plus manuals
	Upgrade prices available directly from the company

Abstract

With 13 new procedures and many new features just added in the latest upgrade, Systat 7.0 is an extremely powerful software package for research-quality graphics and statistics. With a standard Windows "graphical user interface," Systat gives users such an extensive library of integrated graphics and statistics that no single person is likely to use all the features or techniques.

Features

Think of Systat as a world-class library of statistical modules and algorithms integrated seamlessly. Most people start using Systat because it is a premier tool for exploratory data analysis as pioneered by Tukey (1977) and a popular tool for

making presentation-quality graphs in the spirit of Tufte (1983; 1990; 1997). This new version (V7.0) has just added (i) bootstrapping and sampling, (ii) classification and regression trees, (iii) conjoint analysis, (iv) correspondence analysis, (v) logistic regression, (vi) perceptual mapping, (vii) probit analysis, (vii) signal detection analysis, and other features for power users. Well, those are impressive additions to the more traditional and widely used features including: exploratory data analysis using graphics as the principal tool, descriptive statistics, univariate and multivariate linear and nonlinear regression, time series analysis, t-tests and all other standard tests, principal components analysis, ANOVA / MANOVA, discriminant analysis, probability plots, scatterplot matrices, 15 types of smoothers along with kernel estimators for bivariate data, Voronoi diagrams, cluster analysis, and more. Since each feature has options accessed by dialog boxes, sometimes with hierarchical dialog boxes, Systat has the power to tackle large and difficult datasets.

Performance

Systat is a compiled product written by people who know how to write fast and efficient code. It is fast, remarkably fast, especially when run under Windows NT ® on a PentiumPro ® chip. The graphics modules allow truly interactive data exploration; the "Dynamic Explorer" can transform x- and y-axes to replot, in real time, several 2D scatterplots in a "scatterplot matrix," along with the best-fit regression lines confidence ellipses, and univariate border densities. However, even with a fast PentiumPro chip, the program is slow at rotating a 3D scatterplot, a feature first brought to Macintosh ® computers in 1987 in a pathbreaking program called "MacSpin."

Documentation

In a word, superb. Systat comes in 5 volumes (weighing some 7 pounds!): Graphics (and introduction), Data, Statistics, New Statistics, and Command Reference. Each volume gives clear presentations with many sample problems keyed to data files distributed with the software. The style appeals to novice and

to veteran users, and the style conveys advanced statistical and computational instructions without intimidating the reader. Throughout, the manuals give full citations to books and articles that explain the methods, techniques, algorithms, or point of view.

Systat also has extensive context-sensitive help files available on-line at any time with the click of the mouse on an icon in a menu bar. In my experience, the on-line help files do not fully replace the printed manuals, but they are better than those found in many other competing software packages. For a person working on a laptop computer while flying in an airplane, the on-line help files usually carry the day.

Ease of Learning and Ease of Use

Anyone experienced in using software running under the Windows ® operating system can easily start the program and use the sample files to work the example problems in the manuals. Anyone with a semester or two of college statistics can easily begin productive work in 30 minutes or so. In my experience, new users can produce useful results almost immediately, especially by starting with the graphics modules for exploratory data analysis. Since a user can access the program's features from menus and from dialog boxes, I have not had to use the Basic-like command language included in the program in a decade of using the program. Overall, Systat offers a user the pleasant combination of easy and immediate access to many features combined with industrial strength for the serious statistician.

Error Handling

Systat has not crashed once during hours of work. Given that I make many mistakes, Systat has excellent error handling and recovery. I have not explored every module in the program, but Systat gets good marks for error handling.

Support

Systat provides strong technical support on the telephone for issues related to installation. The company also provides admirable technical support (with appropriate limitations) on statistical questions. Of course, Systat cannot provide tutorials for people wanting to learn statistics. The company assumes, I think quite properly, that people who buy and use this powerful product have had one to more semesters of probability and statistics.

Value

Not the least expensive product on the market and certainly not the most expensive product on the market, Systat can meet most if not all an analyst's needs for years to come. Most people will never outgrow this package. I started using Systat approximately 10 years ago on Macintosh computers, and I have not outgrown the software. On the contrary, Systat has added new features faster and upgraded the package more frequently than I have been able to master the techniques and find data sets suitable for analysis.

Examples

In 1996, Fawcett et al. published a study on the distribution of blood lead concentrations in a birth cohort of New Zealanders at age 21 years. After receiving a copy of the data files from the authors, I used Systat to graph LogNormal probability plots for the concentrations of blood Pb in the 368 females and the 411 males in the sample. Figure 1 shows the results -- with the z-score shown on the abscissa and the $\log_{10}$(concentration of blood Pb in μg/dl) on the ordinate. From Figure 1, we see immediately (i) that each distribution is approximately LogNormal in shape and (ii) that the distribution for females is systematically lower (is dominated by) the distribution for males, i.e., the two distributions have different geometric means but similar geometric standard deviations. Upon closer inspection of the lower tail of the distribution for females, we see evidence suggests that there may be two distinct subpopulations of females, each with different LogNormal distribution of blood Pb concentrations. In

particular, one could fit two piecewise linear splines to the plot for females in Figure 1 -- one for z less than ~1 and one for z greater than ~1.

As a second example, Burmaster (1997) has analyzed data collected by the US Veteran's Administration for the height (Ht in cm), body weight (Wt in kg), and body fat (F in percent of body weight) for 646 men from age 50 yr through age 79 yr. Interesting patterns and insights emerge from the data, as shown in the scatterplot matrix in Figure 2. This figure shows the data for height and body fat in the measured units and the data for body weight transformed by the natural logarithm function. These bivariate marginal scatterplots -- and many other analyses in the original publication -- support the conclusion that Ht, ln(Wt), and F are well modeled by a trivariate Normal distribution with these Pearson correlations: $\rho(Ht, \ln(Wt)) = 0.486$, $\rho(Ht, F) = 0.074$, and $\rho(\ln(Wt), F) = 0.613$.

Acknowledgments

Alceon Corporation funded this review.

Trademarks

Systat ® is a registered trademark of Systat, a division of SPSS, Inc. Windows ® and Window NT ® are registered trademarks of Microsoft Corporation. Macintosh ® is a registered trademark of the Apple Computer Corporation. Alceon ® is a registered trademark of Alceon Corporation.

References

Burmaster & Murray, 1997
Burmaster, D.E. and D.M. Murray, 1997, A Trivariate Distribution for the Height, Weight, and Fat of Adult Men, Risk Analysis, in review

Fawcett et al, 1996
Fawcett, J.P., S.M. Williams, J.L. Heydon, T.A. Walmsley, and D.B. Menkes, Distribution of Blood Lead Levels in a Birth Cohort of New Zealanders at Age 21, Environmental Health Perspectives, Volume 104, Number 12, pp 1332 - 1335

Systat, 1997
Systat, 1997, Users Manual and Reference Manual for Systat V7.0 for Windows, SPSS, Inc, Chicago, IL

Tufte, 1997
Tufte, E.R., 1997, Visual Explanations, Graphics Press, Cheshire, CT

Tufte, 1990
Tufte, E.R., 1990, Envisioning Information, Graphics Press, Cheshire, CT

Tufte, 1983
Tufte, E.R., 1983, The Visual Display of Quantitative Information, Graphics Press, Cheshire, CT

Tukey, 1977
Tukey, J.W., 1977, Exploratory Data Analysis, Addison-Wesley, Reading, MA

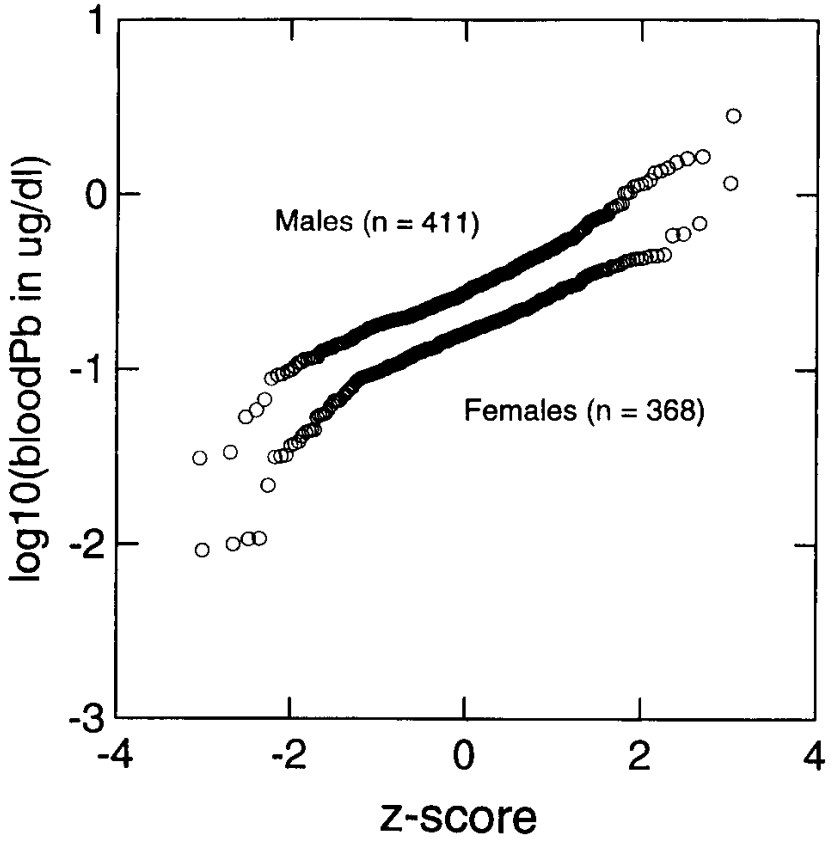

Figure 1: LogNormal Probability Plots for Concentrations of Blood Pb

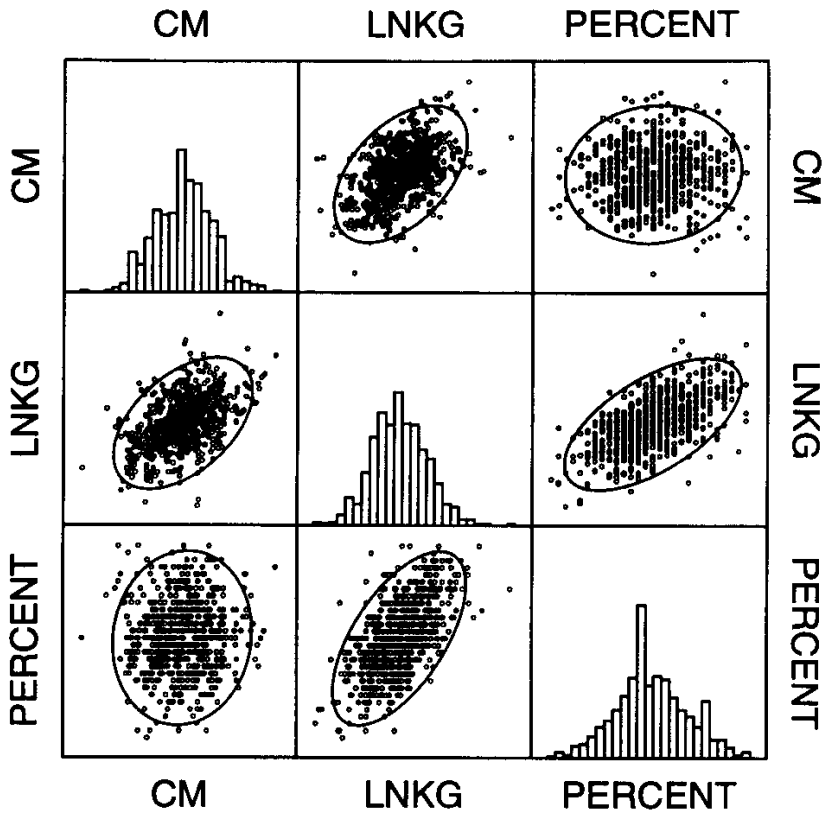

Figure 2: Scatterplot Matrix for Ht (cm), ln(Wt(kg)), and F(percent)