

Data Analysis semicolon-free; Using SAS® without the syntax
Julia Brooks, SAS Institute Inc., Chicago Regional Office

Introduction

Traditionally, SAS® Software has been known for its strengths in statistics. From ANOVA to Z-scores, SAS software has had the breadth and depth to provide researchers and statisticians with the programming tools needed to analyze complicated data and produce unique statistics.

Although SAS' prewritten procedures aid the user by eliminating the programming of these difficult algorithms, they do not eliminate the need for the user to have some knowledge of the SAS language, and the multitude of options available for customization. The nonprogrammer who has data analysis responsibility within their company or university, typically does not have time nor desire to learn a programming language.

As SAS Software has expanded its technology by providing "more than just statistics," it has also enhanced the area of data analysis. With Release 6.07 of the SAS System, two new interactive products, SAS/LAB® and SAS/INSIGHT® software, help the user get statistical results without writing SAS code. The remainder of this paper will give a brief tutorial on each product, while describing the analytical and graphical similarities and differences.

SAS/LAB Software

SAS/LAB software is a perfect fit for a user with limited statistical background. It is a menu-driven product that guides the user through the analysis of data providing useful graphics and interpretations of results. SAS/LAB software automatically checks underlying assumptions while performing commonly encountered analyses such as analysis of variance, simple and multiple regression, and analysis of covariance.

The primary window displays required fields such as the data set to be analyzed, the type of analysis to be performed, the dependent and independent variables, and the preferred model. From here, the experience level of the user determines their next selection.

The more experienced user can quickly fill in these requirements by selecting a field and choosing from a prompted list. When selecting a model, the user is not only given a list of models to choose from, but also given the option to define their own. Additionally, on this main window is a list of the variables included in the data set. By selecting individual variables from the list, SAS/LAB software will produce distribution information including a histogram, box and whisker plot, normal probability plot, and descriptive statistics. Similar graphs and statistics are produced by the Summarize button found to the right of the required fields. The Summarize button allows the user to select up to five variables at a time to analyze through a scatter plot matrix, bivariate or univariate statistics, or a three-dimensional plot.

For the user who is unsure of how to analyze their data, the software offers an alternative path through an Analyze button on that same primary window. This button can be chosen once a data set has been selected. At that point, the user is prompted with a list of analyses, similar to that prompted by selecting the Analysis field directly. However, the difference is an additional selection of "Not Sure." By selecting "Not Sure," the user allows the software to choose the most appropriate analysis for their data. The "guiding" then begins as the user is prompted for the response variable, explanatory factors, and model. SAS/LAB software then runs the analysis.

The next window produced is a plot of the predicted response variable versus the observed. (In the case of simple linear regression this would be the response variable plotted against the independent variable). A message at the top of the window along with the interpretation of results will tell the user if any standard regression assumptions have been violated. By selecting the Assumptions button found to the right of the graph, SAS/LAB software will list and highlight those areas that need further examination. This list might include response scaling, equal slopes, outliers, constant variance, influential observations, and confounded effects. For each case, the software will explain what is wrong and suggest a solution, such as transforming data or excluding observations. The user may choose to follow the software's solution or create their own. At that point, the analysis is rerun and the user is prompted to look at the Results button.

Through the Results button, the user is given four areas to explore: overall model fit, analysis of variance, parameter estimates, and predicted values. Statistics given in the Overall Fit window include the R-square, adjusted R-square, root MSE, coefficient of variation, and the p-value for the F statistic. At any point, the user can choose Help from the menu to get a detailed interpretation of each statistic and the significance of the model. The second selection in the Results is an analysis of variance, displayed in either the traditional ANOVA table format, which is the default, or in a graphical format through the Pareto option. The user can view the regression equation by choosing the third selection, Parameter estimates. Along with the model, SAS/LAB software also displays the degrees of freedom, standard error, the ratio of the estimate to its standard error (T), and the p-value for the hypothesis test that the parameter is equal to zero ($\text{Pr} > |T|$) for each term in the model. The model, itself, can be used to predict new response values through the final selection in the Results window, Predicted values. Here the user can explore predictions over a grid or make individual predictions of their own. A Prediction List window displays the estimates and a 95% prediction interval.

At any time throughout the analysis, the user can choose to save current text or graphics into an on-line journal. This journal facility allows the

storing, editing, and printing of text and graphics, software-generated interpretations, and personal comments.

SAS/INSIGHT Software

SAS/INSIGHT software provides the data analyst with a user-friendly interface for easy, in-depth, graphical data exploration. SAS/INSIGHT software is a dynamic tool for data analysis and visualization, allowing the user to graphically explore data through interactive box plots, bar charts, scatter plots, and three-dimensional rotating plots. The user can also fit explanatory models using generalized linear modeling techniques.

From a prompted window, the user can choose a data set to analyze. That data set is then displayed by the software in a spreadsheet format, columns representing variables and rows representing observations. By highlighting a column and choosing from the Edit or Data menus, the user can perform such data management tasks as moving columns, sorting by variable values, labeling, formatting, transforming variables, or extracting variables to a new table. If no columns are highlighted and the user chooses a selection from the menu, SAS/INSIGHT software will prompt with a window, called a dialog, specific to the menu task chosen. These dialogs allow for more customization by including, in addition to the variable choice, output preferences and methods (where applicable).

The main graphical and analytical features of SAS/INSIGHT software are found under the Analyze menu. Through analysis windows, the user may first choose to examine the distribution of their data through the Distribution(Y) selection. This selection provides univariate statistics (moments, quantiles, frequencies, trimmed and winsorized means, and standard deviation based on Gini's mean difference); parametric estimation based on normal, lognormal, exponential, or Weibull distributions; Kolmogorov statistics; and nonparametric density estimates. The statistics are enhanced through density plots, cumulative distribution plots, and QQ plots. Secondly, the user may choose Fit(YX) to examine the relationship between a dependent variable and a set of independent variables. SAS/INSIGHT software

enables interactive model building by giving the user the ability to add and delete model effects with very few keystrokes. Analytical methods within this selection include multiple, logistic, and Poisson regression with numerous diagnostic capabilities available through the dialog windows. An Output button within the Fit(YX) Dialog allows the user to add statistics and tables to the produced results. Some of the table options included are the X'X matrix, parameter estimates, and confidence intervals. Statistics added will then remain stored as variables in the data window even after the analysis window is closed. And finally, Multivariate (Y's) offers the user the ability to examine relationships between two or more variables, including correlation matrices and principal components. Visualization methods include two- and three-dimensional plots, as well as scatter plot matrices with confidence ellipses for mean or prediction.

All graphs and tables produced within SAS/INSIGHT software are linked to one another. Therefore, when the user selects observations from the raw data, those observations are highlighted in all corresponding plots, and vice versa. A method known as "brushing" allows the user to draw a box around a group of observations in a plot and move the box (in a brushing manner) to investigate particular observations in more detail.

To enhance the graphics and pin point observations, SAS/INSIGHT software supplies the user with an abundant tool set. Under the Edit menu, the user can find Tools, Colors, and Markers to help modify their graphs and more thoroughly explore their data. The Tools include three choices of cursors: an arrow (which is the default), a hand (for spinning plots and adjusting histogram bar widths), and a magnifying glass (to enlarge graphs). The colors and markers provide an array of visual combinations to distinguish observations. The user may either select observations directly and then choose the desired color and/or marker or may choose the color or marker and, through a dialog window, set the criteria for those observations to be highlighted. In addition, the user may choose to assign a particular variable as a label variable. Its value will then be displayed when an observation is chosen in a graph. If no variable is assigned as a label, the observation number

the point selected will be shown. Additional graphics options are found under the Edit menu.

Conclusion

From the product descriptions and tutorials above, we can see that though there are some statistical similarities, SAS/LAB and SAS/INSIGHT software are vastly different products marketed for two different groups of users.

In contrast to SAS/INSIGHT's open-ended analysis, SAS/LAB software has specific paths designed for the user to follow. In that same "guided" theme, SAS/LAB's automatic assumption checking shelters the user from producing invalid statistical results. SAS/LAB software also provides full interpretation of results and a journaling facility to document comments and paths taken so the user can more easily understand their data.

For in-depth, statistical exploration, SAS/INSIGHT software provides the tools and capabilities. With the newest additions of logistic and Poisson regression, SAS/INSIGHT software becomes more desirable for the user who has data that does not lend itself solely to simple or multiple regression. In addition to the ability to perform multiple analyses on one or more data sets at a time, SAS/INSIGHT software also allows the user to output generated statistics to SAS data sets.

Both SAS/LAB and SAS/INSIGHT software provide thorough data analysis at the click of a mouse. The choice to use either product should be based on the user's experience level and statistical knowledge.

References

SAS Institute, Inc. (1992) SAS/LAB Software: User's Guide, Version 6, First Edition. SAS Institute Inc., Cary, NC.

SAS Institute, Inc. (1990) SAS/INSIGHT User's Guide, Version 6, First Edition. SAS Institute, Inc., Cary, NC.

SAS/INSIGHT and SAS/LAB are registered trademarks of SAS Institute, Inc. in the USA and other countries.