Guided Data Analysis with SAS/LAB® Software

Wayne E. Watson, SAS Institute Inc., Cary, NC
Kathy Roggenkamp, SAS Institute Inc., Cary, NC
Randall D. Tobias, SAS Institute Inc., Cary, NC

Abstract

SAS/LAB® software is a data analysis advisor for engineers and scientists who lack an extensive statistical background. The software is implemented as a menu system, making it very easy to learn and use.

As an analysis proceeds, the software explains and interprets the results, makes suggestions for further analysis, and issues warnings about violated assumptions. It also provides convenient methods for converting raw data to a usable form, producing graphical display of results, and converting the results into a report.

This software benefits two audiences: the engineer or research scientist who wants to analyze data as well as the consulting statistician who supports these professionals. End users can perform much of their analysis without waiting for the availability of a consultant, and statistical consultants are freed from routine questions about data analysis and how to use software.

Introduction

In the past, statistical software resembled a set of building blocks that users had to assemble themselves to perform an analysis. Those who regarded themselves as statisticians, analysts, and programmers, all rolled into one, did not mind this state of affairs. However, it has produced barriers for those who only need to analyze data, such as research scientists and engineers.

These people want and need more intelligent software. They want a set of tools that is easy to use, reliable, and geared to the task at hand — and does not have to be programmed. SAS/LAB software fills the bill for this audience.

SAS/LAB software serves as a data analysis advisor for those who lack an extensive statistical background, guiding them through a statistical exploration of their data. The software provides convenient methods for converting raw data to a usable form, quick and accurate analysis of the data, graphical display of results, the flexibility to explore different analyses, and a convenient facility for producing a report from the results.

Features

With SAS/LAB software you can explore your data graphically, perform common statistical analyses, and generate a report of the results using an online journal facility. It is implemented as a series of nested windows, each focusing on a different aspect of the data analysis task. Textual results, graphical displays, and selection mechanisms (such as pushbuttons and pull-down menus) are integrated into a single window.

The system is easy to use and provides guidance as you perform your work. Ease-of-use is exhibited in several ways. It is a point-and-click menu-based system. You provide information about an analysis by selecting from choices that the software presents. Text entry is minimized, speeding up the task and reducing the possibility of error. You are prompted for required information.

You do not need to know SAS® syntax in order to use SAS/LAB software. You quickly learn how to make selections from pull-down menus and selection lists and to maneuver from window to window.

Another principal feature of the SAS/LAB menu system is the intelligent guidance that it provides the user. The guidance comes in several forms: assistance in defining analysis characteristics, assumption checking, and recommendations for refining an analysis.
Getting started

You invoke SAS/LAB software by issuing the lab command from any SAS display manager window. The first time that you bring up the menu system, you are presented with a list of existing SAS data sets. You select the data set you want to analyze, or you can elect to create a SAS data set if your data are not already in that form. In later invocations of SAS/LAB software, the menu system remembers which data set you last analyzed. You can change to a different data set by using a pull-down menu selection.

When you select a data set to analyze, the software displays information about the data, such as the number of observations and the names and number of unique values of each variable. You can choose to manipulate the data by transforming variables or editing values, summarize the data by plotting histograms or bivariate scatter plots, or analyze the data.

Data summary

You may choose to explore your data before proceeding with a statistical analysis. The data summarization facilities emphasize graphical display. Histograms, normal probability plots, box-and-whisker plots, and single and multiple bivariate scatter plots are available. Univariate descriptive statistics, correlations, and one- and two-way frequency tables can also be displayed. You can customize most displays by selecting optional characteristics from the pull-down menus.

Data analysis

A statistical analysis has several characteristics: the type of analysis, the dependent (response) variable, the predictor variables (factors), and the type of model. There are several ways to specify these characteristics; the easiest is to press the [Analyze] button and let the software prompt you for the appropriate information. For example, a list of available analyses is displayed from which you select one. If you select Not sure, the software assists you in making a decision.

Four types of analyses are currently available: one-way analysis of variance, simple linear regression, multiple regression and analysis of variance, and analysis of covariance.
Similarly, you provide the response variables, factors, and model information. You also indicate which factors are classification variables. You can choose from three models: main effects only, main effects and interactions, and quadratic, or you can compose your own model using a model-building facility. The software prompts you for the information; complete online help is always available.

The software guides you through the decision-making process. You are only presented with those choices that are appropriate for your situation. For example, in the window pictured above, Analysis of covariance does not appear because it is not appropriate for the data set. Similarly, if only one choice is appropriate, the software makes that decision for you — if the data set contains two numeric variables and you choose one to be the response, the remaining variable is automatically set as the factor. Or, if you have one numeric and one character variable, no decision is necessary; the numeric variable is made the response and the character variable is made the factor. Because the character variable must be a classification variable, the analysis is automatically set to One-way analysis of variance. The software displays any selections it makes.

When the analysis characteristics have been specified, the software performs the requested analysis. It checks the assumptions underlying the validity of the analysis and warns you if any violations are detected. The assumptions that are checked may be specific to the analysis. For instance, a test of equal variances among groups is performed for one-way ANOVA while a test of constant variance is performed for all other analyses. Some checks are made for all analysis types. For instance, the data are always checked for outliers, and if the response values are all positive, the response variable is checked to see if a power transformation improves the fit of the model.

The results of the analysis are displayed graphically along with a natural-language summary of their interpretation. Various windows can be accessed from the main analysis window, each presenting a tabular and/or graphical display of a particular aspect of the analysis.

If you investigate an assumption, a window displays information allowing you to assess the nature and degree of the violation. It also displays an interpretation which explains the violation and possible remedies for you to take.

In some instances, applying a correction for one assumption remedies other violated assumptions. On the window above, applying the suggested log transformation to the response variable also corrects problems detected by the checks for constant variance, outliers, and curvilinearity. Choosing a remedy causes the analysis to be rerun and the assumptions to be checked again for the new analysis.
The software also checks for outliers—observations which differ from their predicted values by a large amount. Possible outlying observations are listed on the window and a scatter plot can be requested which highlights the problem observations. You are informed that an outlier may be due to erroneous data or an inappropriately specified model. Suggestions for resolving the problem include checking the data for correctness, specifying a different model, or deleting the observation.

No computer program can make all decisions for you, and SAS/LAB software does not try to do so. You must be the final judge concerning decisions about the analysis. Specifically, not all assumption violations are problems. SAS/LAB software provides suggestions and assistance; you must decide how best to use this guidance. Thus, SAS/LAB software is not intended to replace statisticians but to address their clients’ routine questions.

After reconciling potential assumption violations, you may wish to view the results of the analysis. The analysis window contains a graphical display of the data and a short interpretation of the overall findings. As shown in the window below, a more detailed interpretation is available from the pull-down menu.

More detailed results are also available. As with the assumptions, the results are specific to the type of analysis. Generally, for each type of analysis, the overall model fit and an ANOVA table are presented, and a facility for analyzing group means or calculating predicted response values is available.

The Overall Fit window displays information about the overall fit of the model, such as the overall F statistic, its associated probability, and the model R-squared.
The ANOVA window displays the significance of each effect in the model. You can choose between the usual ANOVA table, shown below, and a Pareto chart version.

For simple regression the Parameter Estimates window offers more appropriate information than the ANOVA table window: the parameter estimates, standard errors, and t statistics for each term in the model.

Storing results

SAS/LAB software allows you to store the results of the analysis in an online journal facility. Journaling is performed on a window-by-window basis. The user can request that the text or the graph on a window be saved. Text and graphs are saved separately. Information is automatically included in the journal to help you identify what the analysis consisted of: the date and analysis characteristics such as the response, factors, model, and omitted observations. The software maintains a separate journal file for each data set it analyzes.

You can edit the contents of the journal, adding your own comments and tailoring the software-generated text for your report needs. Journaling of information is cumulative. The contents of the journal file or the contents of an individual window can be printed. Graphics can also be printed immediately or stored for later review and output.
Help

Online help is available for all windows and can be saved to the journal. Help includes information on how to use the pushbuttons and pull-down menus on a window as well as technical information on the role of the window in the analysis. For example, help for the outlier window explains what an outlier is, how the software detects it, what you can do about outliers in general, and how you use the window to assess and remedy the outlier problem.

Data entry and manipulation

If your data are not in the form of a SAS data set, you can choose from among several methods to create a SAS data set. Available methods include entering data manually in several different tabular forms and reading an existing file. The software prompts you for the layout of the file. One of the tabular forms is pictured in the window below.

Customizing your SAS/LAB environment

You can customize the window sizes that the software uses and optimize graph characteristics for your devices.

SAS/LAB software also allows you to configure its printing and plotting facilities to use the particular text printers and graphics hardcopy devices at your site. After an initial setup, printing your results on these devices is effortless.

Conclusion

SAS/LAB software enables you to perform a data analysis without knowing much about statistics or software. The software guides you through the analysis; it prompts you for necessary information, offers suggestions for further analysis, interprets results, warns you of problems, and simplifies the task of producing a report containing text and graphic results.

SAS/LAB software fills an important role in the information delivery system provided by the SAS System. It makes it easier than ever to turn your data into information, as it combines the reliable statistical methods for which the SAS System has always been known with graphical displays and guidance in using those methods.
Product Information

- SAS/LAB software is a separately licensed product.

- Base SAS and SAS/GRAPH® software are required; SAS/FSP® software is recommended to take full advantage of the data management facilities.

- SAS/LAB software is available on all Release 6.08 and 6.09 platforms.


The type of graphic display used by SAS/LAB software is not supported for all types of monitors. Because this feature is an intrinsic part of the product, SAS/LAB software is only recommended for certain monitors. The currently supported monitors are:

- VT 286     - TEK 4105
- VT 340     - TEK 4205
- IBM® 3179  - workstation
- IBM 3270   - PC
- IBM 3279

SAS, SAS/FSP, SAS/GRAPH, and SAS/LAB are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. IBM and OS/2 are registered trademarks or trademarks of International Business Machines Corporation.

Other brand and product names are registered trademarks or trademarks of their respective companies.