INTRODUCTION

This workshop offers hands-on experience with the analysis environment provided by SAS/LAB software. The rationale for this environment is as follows.

In the past, statistical software resembled a set of building blocks that users had to assemble to perform an analysis. Those who regarded themselves as statisticians, analysts, and programmers all at the same time 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 that does not have to be programmed. SAS/LAB software fulfills these requirements.

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.

Scope of the Workshop

You will work through an extended example that illustrates many of the important features of SAS/LAB software, from data management to data exploration to data analysis to report generation.

You were invited to bring your own data to investigate with SAS/LAB software. You will be able to do that after we work through the example together. Those of you who did not bring your own data may want to see what SAS/LAB software can do with the sample data sets that are provided with the software.

SAS/LAB Software under Microsoft Windows

At the workshop, you are running Release 6.08 of the SAS® System under Windows. You should familiarize yourself with the following operating system details before you proceed.

You can operate SAS/LAB software with either pop-up menus or menu bars, according to your preference. Pop-ups are activated in a window by pressing and releasing (or "clicking") the right mouse button; pop-up selections are then made with the left mouse button. Menu bars are displayed across the top of each SAS window, and selections are made using the left mouse button.

If you want to change the menu style, select Preferences from the AWS menu bar. Then select Options to bring up a dialog box and indicate your choice of pop-ups or menu bars. Select the Save button after you make your selection.

This handout uses the > symbol to indicate pop-up or menu bar selections. For example, the selections described in the previous paragraph would be indicated as follows:

Preferences > Options

ANALYSIS OF MINING DATA

In this example, you are a mining engineer who has performed an experiment to determine whether drilling time is faster for wet drilling or dry drilling (Penner and Watts, 1991). Each drilling method was used in three different test holes, each of which was drilled to a depth of about 400 ft. Your data set contains one observation for each five-foot length of each hole.
Invoke SAS/LAB software by typing the following in the command dialog box:

```
lab data=sasuser.mining
```

The window that appears is the software's main window, from which all other activities originate. The software is ready to work with the data set you specified.

The information in the bottom half of the main window summarizes the data set. The MINING data set has 480 observations and contains four variables. Information about each variable is provided, including the number of unique values.

For a closer look at the data, select

File > Edit/Browse

to see a listing of the variables and data values.

Drilling time may become greater as the hole becomes deeper. To see the relationship between these two variables, select [Summarize], then DRILTIME and DEPTH, then [OK]. Drilling time does not appear to change much with depth until the hole is well over 200 feet deep.

You are mainly interested in how drilling time is affected by the method of drilling, wet or dry. To add this information to the plot, select

Locals > Variables > Grouping variable > Select grouping variable

*If there is no command dialog box at the bottom of your screen, make these selections to bring up the dialog box:

Globals > Command > Command...
Select METHOD from the Variables window.

There are so many points that it is difficult to interpret the resulting plot. To see what is going on between drilling time and drilling method, without the complication of depth, select [Univariate statistics] and DRILTIME by METHOD.

In general, wet drilling appears to take longer, although there is greater variability among dry drilling times. To see which observations have the very high drilling times, make these selections:

- Locals > Plot options...
- Identify outliers...
- DEPTH [OK]

The high drilling times all occurred at depths greater than 300 feet.

Appropriate statistics accompany every summary plot. Select [Statistics] to see descriptive statistics for DRILTIME by METHOD. As the box plots showed, the mean drilling time is lower for wet drilling, but the standard deviation is greater for dry drilling.

Select

File > Return to main window

What have you learned so far? The summarization suggests that dry drilling is faster than wet drilling, but the information is ambiguous enough (especially at greater depths) that you need to make sure that the difference is significant. Analyzing the data can fill in that piece of the puzzle.

To begin the analysis, select [Analyze]. The software automatically prompts you for the information needed to perform the analysis. Detailed help is available in each of the prompter windows.

Because you want to see how DRILTIME is related to both DEPTH and METHOD, select Multiple Regression and ANOVA as the analysis type.

Each prompter window looks similar to the one above.

- When you are prompted for the response variable, select DRILTIME.
- When you are prompted for the explanatory variables, select DEPTH and METHOD.
Next you are prompted for any classification variables. The software knows that any variable with character values, such as METHOD, must be a classification variable. It offers you the chance to treat DEPTH as a classification variable, but you want to treat DEPTH as a quantitative variable in this case. Simply press [OK] to go on.

Finally you are prompted for the type of model to be fit. You suspect that there may be interaction between DEPTH and METHOD. That is, the comparative effectiveness of the drilling methods may change with the depth of the hole. Therefore, you should select to fit a Main effects and interactions model.

This is all the information needed, so the analysis begins. The window that appears next has several noteworthy features. First, you are informed if any important statistical assumptions were violated by the analysis. Second, you are shown a graphical display of the results. Finally, the results are briefly interpreted for you.

The type of graphical display presented depends on the type of analysis performed and the number and types of factors involved. In this case, the plot shows two regression lines, one for each method.

The interpretation indicates that the explanatory variables appear to be related to the expected value of DEPTH, but assumptions were violated as the analysis was performed. Select [Assumptions] to see which assumptions were a problem. Three assumptions are flagged.

In general, you should investigate the assumptions in the order in which they are listed. This is usually the most effective way to clean up the analysis. In this case, first select Response scaling.

The software tries a series of transformations and recommends that you use the log (base 10) of the response for a better fitting model.
Select

File → Apply
to redo the analysis using the recommended transformation.

Notice that the response variable (on the Y axis) is now LOG10(DRILLTIME). But once again, assumptions were violated.

Two outliers are listed. To see a plot of the residuals versus predicted values, select

Locals → Plots → Residual plot

Notice the two points marked by a +. These are the two observations that are flagged as potential outliers.

Press End (F3) to return to the previous window. Looking at your data collection sheets, you see that there were drilling problems at these outlying points. Thus, you feel justified in deleting them from the analysis.* Click on each row to put an asterisk (*) in the 'Omit?' column for each observation. Then press End (F3) to redo the analysis.

*You should not delete outliers simply because they are flagged by the software. Outliers do not invalidate a statistical model.
No assumption violations are noted this time, so you can begin to focus on the results. The plot shows that the regression line for dry drilling is consistently below the line for wet drilling, although the lines tend to converge at greater depths.

To save a copy of the graph for a report, select
   Journal > Save > Current graph
To save a summary of the analysis for a report, select
   Journal > Save > Current text
To review what you saved and optionally add a comment, select
   Journal > Review > Saved text

Press End (F3) to leave the journal.

Select [Results] and Overall model fit. The overall model is significant. About 47% of the variability in the response is explained by the model. If you saved these results to the journal, the overall fit table would be appended to the information already there.
Since the regression lines for dry and wet drilling converged at greater depths, and since the interaction of DEPTH and METHOD was significant, you may want to investigate more directly the impact of DEPTH on the relative effectiveness of the two drilling methods. One way to do this is to form groups based on DEPTH.*

Select

File > Modify > Recode a variable > Recode ranges of values...

In the Variables window, select DEPTH as the variable to be recoded. Then, as shown below, enter depthgrp as the name for the new variable. Add (grouped) to the variable label, and enter 3 as the number of groups to be in the new variable. Then select [OK].

To group together depths of <= 100, > 100 to <= 300, and > 300, fill in the subsequent window as shown below. Then select [OK].

The new variable DEPTHGRP is added to the data set, and the main window is updated with the new information. Notice that DEPTHGRP has three unique values, as you would expect.

Now you want to repeat the analysis using DEPTHGRP rather than DEPTH as an explanatory variable. To switch variables, click on Factors:. When you are prompted to select explanatory variables, click on DEPTH to unselect it (the * should disappear), and then click on DEPTHGRP to select it. Select [OK]. When you are prompted to select classification variables, again click on DEPTHGRP, and leave METHOD also selected. Select [OK]. Your main window should now look like the one below.

Select [Analyze] to redo the analysis. Notice that the software presents a different type of plot this time, one suitable for two classification variables.

To fix the assumption violations that occurred during this analysis, select [Assumptions]. First select Response scaling and apply the suggested LOG10 transformation. Then select Outliers and delete the listed observation. You should get the following results.

*This example is not intended to recommend any particular analytic approach but rather to emphasize some convenient features of SAS/LAB software.
Select [Results] and Analysis of variance to see that all terms in the new model are significant. Press End (F3) to return.

Since you have a good model for the data, you may want to be able to predict what the drilling time would be for a hole drilled by a certain method at a certain approximate depth. From the Results window, select Predicted values. Then select Examine LS Means.

You are shown the predicted mean LOG10(DRILLTIME) for the levels of the first classification variable, METHOD. As expected, the value for dry drilling is lower. The two predicted means are significantly different.

Notice that the response means are presented in their transformed (logged) version. To see the predicted drilling times in minutes, select

Locals > Use original response units

You can also examine LS means for each DEPTHGRP level and for the groups formed by the interaction of METHOD and DEPTHGRP. To see the latter, select METHOD*DEPTHGRP after selecting

Locals > Change class variable...

Notice that dry drilling is definitely faster at depths less than 300 feet. At depths greater than 300 feet, dry drilling does not appear to be significantly faster than wet drilling.

Select

File > Return to main window

To focus more closely on what is going on at greater depths, you can perform an analysis only for that group.* In the main window, select

Locals > Select observations...

Fill in the window as follows, click on Apply, and select [OK]

*This example is not intended to recommend any particular analytic approach but rather to emphasize some convenient features of SAS/LAB software.
The information at the bottom of the main window now applies only to the subset of the data that you have specified: there are 120 observations in the subset, DEPTHGRP has only one unique value among these observations, and so forth.

In the next analysis, DEPTH and DEPTHGRP need not be explanatory variables since you have already based the subset on DEPTHGRP. Instead, you want to do a one-way analysis of variance using METHOD as the classification variable. To set up this analysis, click on Analysis: and select One-way ANOVA.

Select [Analyze] to see the results.

There is one outlier, and you can delete it using the Assumptions button and the Residuals and Outliers window as you have done in previous analyses. The redone analysis produces the results below. The interpretation tells us that there is only weak statistical evidence of a difference among drilling times for wet and dry drilling.

For a one-way ANOVA where the classification variable has two levels, the software performs a t-test for the difference between the two means. Select [Results] and T-test.
The \( t \)-test interpretation again notes the weak statistical evidence for a difference between drilling times for the two methods. For more information about the statistical evidence in this particular analysis, select Help > Detailed interpretation

A detailed interpretation of the results is available from every analysis window.

On the other hand, you may want help on generic information presented in a window. For example, in the \( t \)-test window you may wonder about the difference between the rows for equal and unequal group variances. For this type of information, select Help > Window help

In any help window, highlighted words and phrases can be selected for more information about those topics.

Select [Exithelp] and then

File > Return to main window

CONCLUSION

SAS/LAB software enables you to perform a data analysis without knowing a great deal 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.

REFERENCES


SAS and SAS/LAB are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. \( \copyright \) indicates USA registration.

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