

## The SAS system

SAS stands for the Statistical Analysis System, a software system for data analysis and report writing. SAS is a group of computer programs that work together to store data values and retrieve them, modify data, compute simple and complex statistical analyses, and create reports. SAS can be used through the SAS Analyst drop down menus, or by writing one's own code for maximum flexibility in performing complex analyses.

## Preparing data for SAS

Data must be in a SAS or Excel data set in order to use SAS. It is important to understand how the data table must be arranged so that it can run on SAS.

As an example, maize grain yield data is shown from an experiment with five levels of nitrogen (N) and 3 replicates, for a total of 15 observations. The experiment was installed in a Randomized Complete Block Design (RCBD). The response was grain yield collected at harvest. Proper arrangement of the data is shown Data sets on the web page.

Each item in the table is a data value. Each row represents one observation. Each column represents one type of information or variable. Variables can include treatments, replicates or blocks, covariates, and responses. The column headings are the names of the variables. A variable name can contain from 1 to 32 characters consisting of numbers and letters, but must begin with a letter. It is a good idea to keep the names informative, but short to minimize the amount of typing. This data set has 15 observations and 3 variables.

## Starting SAS

When SAS is opened, three windows appear. On the left is the Explorer/Results window. The main window includes the Editor, Output and Log windows. The tabs along the bottom are used to toggle among the windows.

## The Explorer/Results Window

The Explorer/Results Window on left side of the screen is used to browse SAS libraries, program files, and the results of statistical procedures. The Explorer Window is used to locate SAS data sets and other SAS-related files. Data sets that have been imported into SAS will be found under Libraries in Sasuser or Work. Data in Sasuser is stored until deleted by the user, while data in Work is automatically cleared when the program is closed. The Results Window is used in connection with the Output Window to look at the results of statistical procedures.

## The Editor Window

The Program Editor is used to create, edit, and execute SAS programs. Here, any data manipulation and analytical tasks can be performed through the writing of code that instructs SAS how to process the data. While the Program Editor is a critical feature for advanced users of SAS, it is easier to use the SAS Analyst drop down menus and dialog boxes to perform an

analysis. Not all data manipulation and analytical tasks are available through the drop down menus. The Program Editor can be used to save the code created by SAS Analyst for editing, to provide a record of an analysis, or to re-run an analysis in the future. While this course will not teach SAS programming, the Program Editor can be used to modify and run programs previously generated by SAS Analyst.

To open a SAS program, make the Program Editor the active window and select File: Open Program. When the file type SAS Files (\*.sas) is selected, any SAS programs in the currently selected directory should appear in the dialog box. After a program has been opened, it can be submitted from Program Editor by clicking on the running man icon located toward the right of the toolbar. The output from the program will appear in the Output window.

### The Output Window

Output is used to view the results generated by the submitted program. Information in Output can be viewed, printed, or saved. The Output can be saved as a text file and then edited in an application such as Microsoft Word.

The Results window works with the Output window to organize the information contained in the Output. To display a particular section of Output, simply double-click the appropriate page icon.

### The Log Window

The Log is used to check a program for errors. It is a good idea to check the Log after submitting a SAS program in order to ensure that the program did not encounter any errors. The Log also contains important summary information. In the Log window comments appear in blue, while error messages appear in red.

### Getting data into SAS

The easiest way to enter data into SAS is to make an Excel file of the data as described in Data sets. Open the Calculating machine data from the web page in Excel, convert it into the form required by SAS, and save it. Each row will be a single observation, and the columns will contain information on the machine (A or B), the replicate (1 through 10), and the observed time for the calculation. The saved data can then be imported into SAS by selecting File: Import Data and following the instructions. Make sure that the file is not open in Excel or any other program before attempting to import it into SAS.

### SAS Analyst

This class will use SAS Analyst as the primary tool for data analysis. Analyst provides drop down menus and dialog boxes to select the desired analysis. To open SAS Analyst, select Solutions: Analysis: Analyst. To open the imported data, select File: Open by SAS name. The data should appear in a table.

To obtain descriptive statistics, select Statistics: Descriptive: Summary Statistics from the tool bar across the top. A dialog box will open with the variable names in a box on the left. Click on Time (the response variable) and then on the Analysis to move the variable to the box. Click on Machine (the treatment variable) and then on Class to move it to the box. Next move to the bottom row of buttons. Click on Statistics and select everything that looks interesting and/or useful, then click Ok. Use the choices under the Plots button to obtain a visual check for normality of the distribution. The Output button modifies output appearance and is not needed for this class. The Save Data button saves the results to a file for further analysis and again is not needed for this class. The Titles button should be used to enter your name and an identifier for this analysis. Entering your name in to the title is the only reliable way of distinguishing your results from those of your classmates. Finally the Variables button allows you to calculate a separate analysis for each Rep and will not be used at this time.

After looking through all the options under the buttons, return to the original box and select Ok. A window will pop up with the summary statistics requested.

To save the results, select Edit from the top tool bar and then Copy to Program Editor. The results will appear with color in a new window. Next select File: Save As. In the bottom Save as type box select RTF files. The file can now be saved to your jump drive or emailed to yourself (or to the TA for assignments) as an attached file.

To view the plots, go to the Results window on the left and select the desired plot under Univariate for the Histogram and under Boxplot for the Boxplot. The plots can also be found by minimizing the Summary Statistics window and checking the tree to the immediate left of the data table.

Next, select Code from the tree to the left of the data table to see the program code that Analyst has created for you. Select Edit: Copy to Program Editor, then click on the running man icon in the top tool bar to run the program. Go to the Output window to view the results. The code disappears from Program Editor, but, if needed, it can be retrieved by selecting Run: Recall Last Submit.

Finally, check the log window for a list of all the steps performed by SAS. Hopefully there are no errors indicated in red.