Institute for Health and Society, Medical College of Wisconsin

MEDICAL COLLEGE OF WISCONSIN

Datum

NEWSLETTER
Division of Biostatistics

CTSI — ADVANCING HEALTH THROUGH RESEARCH AND DISCOVERY

SAS OnDemand for Research

Ruta Brazauskas, PhD, Division of Biostatistics, MCW

SAS OnDemand for Academics is a free "cloud-based" version of SAS® software available to faculty, researchers and students. SAS has long been a preferred statistical software package in many academic institutions and industry. Recently, certain components of SAS package became available free of charge for research purposes at academic institutions. We would like to provide you with instruction on how to obtain the program and how to use it for simple analyses. While there are several SAS products, most statistical analysis tasks are performed by SAS Enterprise Guide. Before you register, you must get an approval by agreeing to use SAS OnDemand for Academics only for research purposes. This can be accomplished by filling out a form found at the following address: http://support.sas.com/research. Wait to receive your Registration Key (via e-mail) before you begin the registration process. You will receive an e-mail if you are not eligible. SAS OnDemand runs on the SAS server which means internet connection is required while it is in use. This also means that it is somewhat slower than running a program on your personal computer. SAS OnDemand Enterprise Guide can be installed on a machine with Windows operating system or on a Mac using Boot Camp. After you install SAS OnDemand on your machine, you can perform statistical analysis by using menu driven options or by writing and running SAS code. If you are not familiar with SAS Enterprise Guide or want to learn more about using SAS Enterprise Guide, the free Getting Started tutorial provides helpful instructions on how to use the main features of SAS Enterprise Guide. After

you have completed your registration and downloaded the software, you can access this online tutorial by following these steps: 1. Select Start > Programs > SAS > Enterprise Guide 4.2 OnDemand. The Credentials Required window is displayed. 2. Enter your SAS server user ID. You can find this user ID on your SAS OnDemand for Academics home page. 3. Enter the password that is associated with your SAS profile account. 4. Select OK. The SAS Enterprise Guide window is displayed. You are now logged on to the SAS server. 5. Select Tutorial: Getting Started with SAS Enterprise Guide from the Welcome to SAS Enterprise Guide window.

Help on statistical analysis tasks is available at

http://support.sas.com/learn/statlibrary/statlib_eg4.2/top_learn.htm

Sample session Open SAS Enterprise guide and login using your credentials. When you start SAS, you will see a Menu bar, at the top, with headings File, Edit, View, Tasks, Program, Tools, and Help. Each menu heading, when clicked, reveals a list of commands. In addition, the entire workspace will be divided into three windows: Project Tree, Server List, and Process Flow. This session will show you how to open and save a data set, and how to calculate some basic descriptive statistics and display data in graphical form. For illustration purposes, we will use data obtained in a study of low birth weight children dying of sudden infant death syndrome [AM Walker et al. (1987) Diphtheria – tetanus– pertussis immunization and sudden infant death syndrome. American Journal of Public Health 77: 945-951]. A new data set can be entered directly into SAS Enterprise or an existing data set in Microsoft Excel format can be opened in SAS Enterprise. First, we will create a data set called "SIDS" containing the ages (in days) at time of death for a sample of 11 girls: 53, 56, 60, 60, 78, 87, 102, 117, 134, 160, and 277 by following these steps: 1. Select File > New > Data. A new dialog window will be opened. 2. Start by naming your data set. Write "SIDS" in "Name" space and click Finish at the bottom of the window. 3. A blank data frame, with columns and rows like a spreadsheet, will appear in front of you. Each column lists data for a single variable and each row displays the observed values for the variables for each individual case. The columns are initially referred to as A, B, C, D, E, and F. Column names can be entered or modified by right-clicking on the column's name, then

select Properties. That will open a new window where you can change the name of the column and specify the type of the data to be entered. By default, SAS expects the data in each column to be characters. If a column contains a non-numeric entry the entire column will be treated as text-type data and the column number will be marked with the red icon containing letter A. Some statistical analyses and graph constructions in SAS require the column data to be numeric. To specify a numerical variable, choose "Numeric" in the drop-down menu under Properties of the column. Suppose the first column (Column A) is named "age" and contains the ages of the girls at the time of death. An entry in a worksheet can be made by clicking on the row and column cell where the entry is to be placed. A screen shot of the data set typed in SAS is shown in Figure 1.
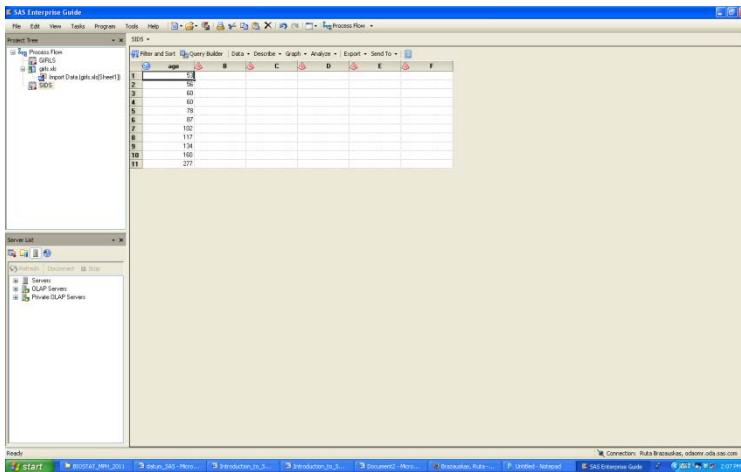

Figure 1: Screen shot of the typed data set

Alternatively, if you have a data set entered into Excel spreadsheet and saved on your local computer, say under the name "sids.xlsx", you may open it from the file by following these steps once you have SAS Enterprise Guide opened:

1. Select **File** > **Open** > **Data.** A dialog box will appear.
2. Locate the directory where you stored the file "sids.xlsx".
3. Double-click on the file name "sids.xlsx"**.** The data file will be displayed in the session window.

Regardless of which approach you used – typed the data yourself or opened them from an Excel file – the data set "sids" is now open and ready for analysis.

Now we use this dataset to illustrate some of the basic functions in SAS, such as graphical and numerical summaries of the data and how the results appear in the session window. Basic statistical summary measures and graphs can be generated by using the **Tasks** option from the menu located at the top of the session window. Follow these steps:

1. Select **Tasks** > **Describe** > **Summary Statistics**. A new dialog window "Summary Statistics for …" will be opened. On the left hand side you will see various options available to get the numerical summary measures and graphs of your choice (Statistics, Plots).
2. The first selection on the left hand side is **Data.** When you click on that, you will see the main portion of the dialog window being divided into two parts: **Variables to assign** (on the left) and **task roles** (on the right). Use it to select variables for the analysis. Move "age" to the "analysis" variable category. You can drag it to the right or use an arrow between the two sides to accomplish that.
3. Click **Statistics** on the menu on the left hand side and that will immediately take you to the list of basic statistics. Mark the ones you are interested in, perhaps Mean, Standard deviation, Minimum, Maximum, Number of observations.
4. Then click **Percentiles** and select those of interest, e.g. Lower quartile (equivalently, 25th percentile or 1st quartile), Median, upper quartile (75th percentile or 3rd quartile).
5. Go to **Plots** on the menu on the left hand side and click on "Histogram".
6. Click the **Run** button on the bottom of the dialog window.

This will display information about the current project, including the output in the session window and the history of commands executed. The largest window is now called Summary Statistics. Below you will also see four choices: **Input Data, Code, Log,** and **Results**. After SAS finishes all the tasks you requested, it will automatically take you to the **Results**. Your data set can be found under **Input Data**. The **Code** window will display the SAS code behind the tasks you chose. **Log** file will provide information on the performed tasks. Running the **Summary Statistics** task with aforementioned options produce the output shown in Table 1.

Table 1: Summary Statistics.

Analysis Variable : age							
Mean	Std Dev	Minimum	Maximum	N	Lower Quartile	Median	Upper Quartile
107.636364	66.132099	53.00	277.00	11	60.00	87.00	134.00

The output in Table 1 tells us that, among girls, the average age at the time of death from SIDS was 107.6 days. The standard deviation was 66.1 days. The median age was found to be 87 days

---

which means that 50% of the girls died before 87 days of age. The histogram allows us to visualize the distribution of age in the sample: it is not symmetric, skewed to the right with majority of SIDS deaths occurring prior to 180 days (6 months) mark.
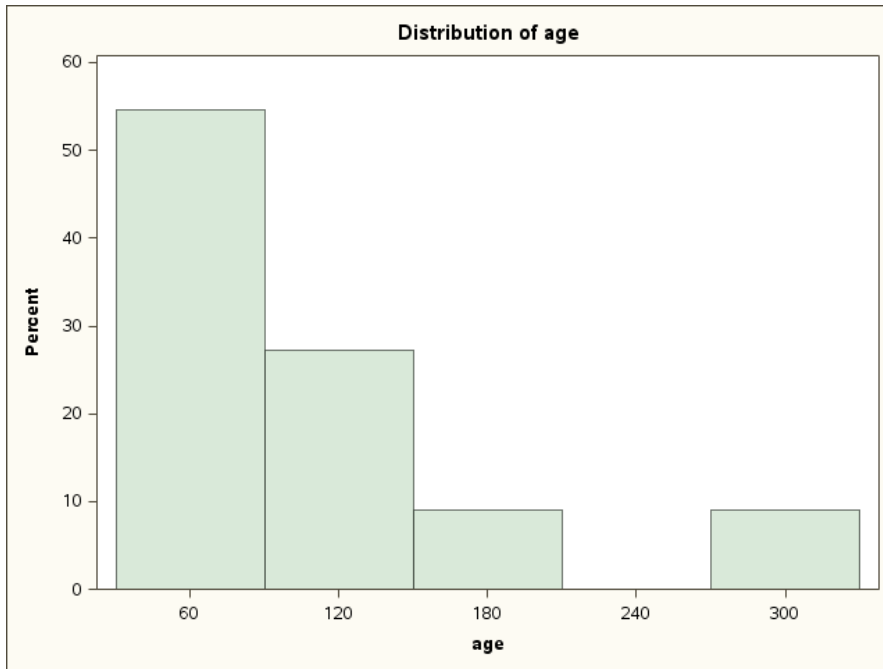

Figure 2: SAS Summary Statistics output

You can use standard copy-and-paste methods to place the numerical portion of the output and the histogram in a Word document. You may save your results as a pdf file. If you have a printer attached to your computer, you can print the contents of the results window by using the **Print SAS Report** option on the **File** menu. You can also save the entire project by selecting **Save As** option on the **File** menu. Saving an entire project will allow you to save the data set (which is particularly useful if you entered the data yourself and did not use the data previously available) and the results and it enables you to see the steps you took to accomplish the task.

When you finish, exit SAS by clicking on the **File** menu and then select **Exit**. A dialog box appears asking if you want to **Save Changes**. If you have not saved the changes or the project you were working on, it is a good time to do that now. If you already saved the project, you may skip this step by clicking "**No**" in response to this final question.

More complex statistical analyses and modeling tasks can be performed using other menu-driven options.