

POINT AND CLICK YOUR WAY TO BUILDING ANALYSIS DATASETS

Betty J. Anderson and Sara L. Osborn
Mayo Clinic Foundation, Rochester MN.

Abstract

Working with numerous and individualized databases is a continuous element in our daily research environment. Investigators wish to examine and interrogate their data from many different aspects and end points. There are also frequent times when the final end point can be obtained by programmable calculations using the variables already existing in the database. SAS ASSIST is a task-oriented, menu-driven database system designed to make end users access to the SAS system quick and easy. It's point and click method allows the end user to do complicated calculation with ease, while behind the screens SAS ASSIST is generating the necessary code. It also allows subsetting of the data to various necessary end points to answer the ever reoccurring questions of the investigators. Using SAS ASSIST provides the novice with a comfortable management environment of the data. This tool allows us to adequately manage some 80 or more unique and separate databases at the same time. We will demonstrate simple step through instructions on data management and manipulation using SAS ASSIST. These tasks would otherwise need to be performed by a SAS power user, but with the aide of SAS ASSIST the novice user can manage and manipulate his/her own data.

Introduction

The medical environment of research is a fast moving ever changing area where there is a need for data management which can be conducted with ease, speed, and accuracy. SAS ASSIST supplies the study coordinator with total control of the data. It provides for data updating and construction of data subsets all which can be done at the work station personal computer. Preliminary conclusions can be tested with the ability to view the data from many distinctly different aspects or end points. An investigator can test his hypothesis and immediately see the results of his idea. This program allows the novice to perform all of these functions. This saves valuable time and decreases the expense for both the investigator and other support personnel. In this tutorial we will show you how to subset your data to allow you to view it from different aspects, and to modify and define variables that do not exist in the original database using some of the already preexisting variables.

Acquiring SAS ASSIST

To begin with SAS ASSIST your organization will need to secure SAS ASSIST software, license, and disk space to hold all of the data that is to be collected. We will not go into the specifics of setting up the screens which includes attaching a variable name to each piece of data, setting up formats, codes, edits, and the entering of the data.. A SAS dataset will allow for each piece of information to have a SAS variable name which will be used in SAS ASSIST to locate the necessary data point as pictured in Figure 1.

Figure 1 Sample Dataset

FSEDIT MYDATA.SAMPLE	
IC:	
DOB:	
GEND:	
DATEAVH:	
REOP:	
REOPDT:	

When all the data is entered into the correct space for the corresponding variables you will be ready to use the functions available in the SAS ASSIST software package.

Operation of SAS ASSIST

We will demonstrate the necessary operation to answer the question : What are the ages of all the patients requiring a second operation for failure of the original surgery? We will need to calculate the age from our preexisting variables, and identify all of the patients who have had a second operation.

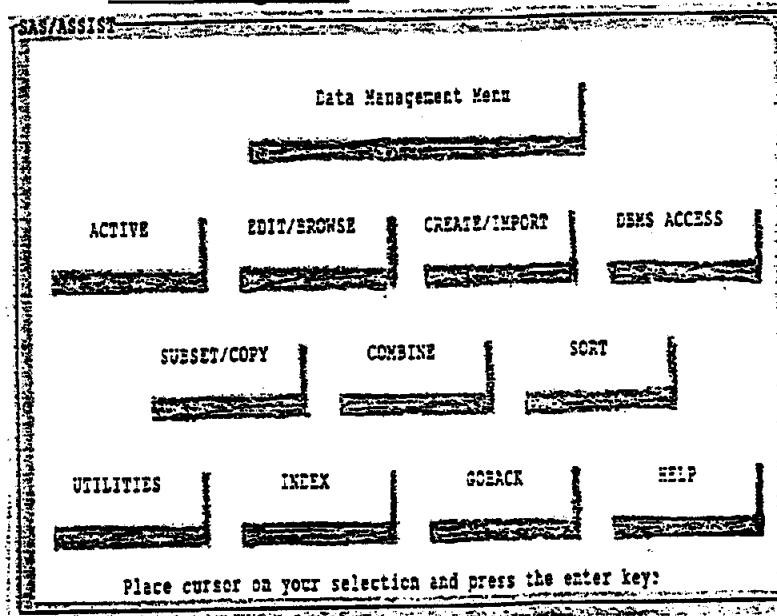
SAS ASSIST is a point and click system, and is the necessary operation to open windows or pull down menus. Upon entering the SAS ASSIST program you will arrive at the Primary Menu as pictured in Figure 2.

Figure 2 Primary Menu

SAS/ASSIST			
Primary Menu			
TUTORIAL	DATA MGMT	REPORT WRITING	GRAPHICS
DATA ANALYSIS	PLANNING TOOLS	EIS	
RESULTS	SETUP	INDEX	EXIT
Place cursor on your selection and press the enter key.			

SAS ASSIST is a menu driven program, and each of the icons on the Primary Menu has a continuation of selections under them. There are multiple boxes located on this screen that lead to many various functions available with SAS ASSIST. To perform the function of subsetting the data we will need to click on the box titled Data Management. Upon clicking on the Data Management icon another screen of icons will appear as pictured in Figure 3.

Figure 3 Data Management



From these selections we will pick the icon titled Subset/Copy. This will bring us to the screen necessary for creating a subset copy of our original dataset. See figure 4.

Figure 4 Subset/Copy

SAS/ASSIST: Subset or Copy a Data Set

<Active data set:>

<Output data:> -REQUIRED- Will be created as a: DATA SET

Subset and transform the data:

<Currently selected observations:> -ALL-

<Currently selected variables:> -ALL-

<Transform existing variables:> -NONE-

<Define new variables:> -NONE-

<Run> <Results> <Setup> <Go back> <Help>

SAS datasets are identified by the library name where they are stored, and the individual name of the dataset. Example: MYDATA (library name) SAMPLE (individual dataset name) .

Point and click on the box labeled Active dataset. This will produce a list of datasets that are available to you. Choose the dataset containing the data that you wish to work with and click on it. It will then appear on the screen. This will then direct SAS ASSIST to work with this dataset.

The next box Output data requires the new subset dataset being created to be named so it can be activated for use in other areas of SAS. Clicking on this box opens another window to name the dataset, using any combination of eight characters. There is also the choice to make the dataset being created temporary or permanent. A temporary dataset will be available for use only during the immediate session of SAS. It will be located in the WORK library, and will be deleted when the current session of SAS is closed. To choose to make a permanent dataset, point to permanent and click. This will then open another window which is requiring the assignment of a library in which to store the new dataset. This will allow access to this dataset during any session of SAS, and will be permanently located in the selected library until it is removed . Your choice of library will be directed by the library allocated at this time. Point and click on the chosen library , and then click on OK to return to the subset dataset screen. Now the dataset to be worked with has been chosen, and the dataset to be created has been named. In this example we named our new dataset REOPAGE , and chose to make it temporary.

When creating a subset of the original data there are times when all of the observations, patients, or selections in the original dataset are not needed in the newly created subset. The next box , Currently selected observations , allows the selection of observations to be contained in your new dataset, eliminating the unwanted observations which do not pertain to the new subject. SAS has provided for the times when all of the variables are needed in the subset by setting the default to select all of the variables. Point and click on this box. See figure 5.

Figure 5 Currently selected observations - Where Clause

SAS/ASSIST: Build a WHERE Clause to Subset the Current Data

To make a selection, place cursor and press ENTER.
Use OK to save selections. Cancel to return without saving selections.

Build a WHERE Clause to Subset the Current Data

Select:	WHERE clause being built:
☐ Variable...	
☐ Constant...	
☐ Function...	
☐ Comparison operator...	
☐ Logical operator...	
☐ Arithmetic operator...	
☐ Concatenate	
☐ Opening parenthesis	
☐ Closing parenthesis	
☐ Undo last operation	

<edit the WHERE clause >

<OK> <Cancel> <Reset> <Help>

This screen assists in building a Where clause. The where clause is composed of variable names to identify which variables are significant to the subject, and the value located in these fields with arithmetic, logical, or comparison operators to indicate the necessary function. In our example this would be where we wish to identify only the patients who had had an additional operation following their first surgery to appear in our new dataset. The variable for reoperation is called REOP, and requires a 1 if a reoperation was performed, and a 0 if none was performed. The where clause would read : REOP=1.

This is formulated by pointing and clicking to the necessary areas to select these characters. Point and click on variable, and scroll down the list to the appropriate variable and point and click on the selection. Example: REOP.

The screen will immediately go back to the where clause screen, and the highlighted choices are the selections available at this time.

The choices:

1. Comparison operators - such as equal too, less then etc.
2. Logical operators - and /or
3. Arithmetic operators - add, subtract etc.
4. Concatenate.

There is also a choice of adding the sign for NOT.

We would need to choose the comparison operator to have equal added. Point and click on the = sign, and it will appear in the box where the where clause is being created, and the screen will immediately return to the where clause screen once again. Example: =

At this point the choice is highlighted to pick another variable, or do constant. Constant allows the user to type in the value that will be found attached to the variable that will signify the observations to be saved. In this example we need to type in 1 to signify the patients who had indeed had a reoperation. Point and click on OK, to be back at the where clause screen. Example: 1

When you become proficient at writing where clauses you may choose the Edit where clause button, click on it , and it then allows the user to type the where clause directly into the box. There are also choices to use opening and closing parenthesis in where clauses with multiple variables. At anytime there is a change to be made to the where clause click on Undo last operation, and this will delete the last choice and go back one step in the where clause formulation to allow the last operation to be redone. When the where clause has been completed click on OK to return to the subset screen.

We have now subsetting the dataset eliminating the patients who do not meet the criteria, as in our example, of having had a reoperation.

The next box , Currently selected variables ,when clicked on opens a list of all the variables found in the original dataset. Point and click to select the variables to be in the new subset. The default in this box is all of the variables in the original dataset, and if not

otherwise specified they will then all be transferred into the new dataset. When choosing the variables this function has a limit of 22 variables which can be selected for transfer. We will limit our selection of variables to only the name, reoperation, date of first surgery, date of birth, and date of reoperation.

The Define a new variable section assists in creating a new variable using the existing variables in the dataset. Point and click on this box, and it will open a new screen that is prompting the creation of the characteristics for the new variable. See Figure 6.

Figure 6 Define or Modify a Variable

SAS/ASSIST: Define or Modify a Variable			
Define a new variable or modify an existing variable's attributes. To make a selection, place cursor and press ENTER. Use OK to save selections. Cancel to return without saving selections.			
Define or Modify a Variable			
Add new variable Name:	Variable:	Length: 8	
	Type:	* Numeric o Character	
	Label:		
	<Format:>	-NONE- <Informat:> -NONE-	
	<Initialize:>	-NONE-	
	- Delete this variable.		
OK	OK	Cancel	Help

The first blank calls for a name for the variable. This can be up to eight characters long. It is necessary to note if the variable will be character or numeric. The numeric is default, and to be noted as character just point and click on the button in front of the selection character. It is also required to label the variable, and this is the name which will appear on the printouts to specify what information the variable contains. There is also a place to chose a format or informat. In our example we will name the variable AGEREOP. It is a numeric value. The label will read age at time of reoperation. Our value will require no informat or format to be selected.

The next operation is to point and click on the initialize box. This will open a window which has the function of creating the operation necessary for SAS to know how to create the variable, and the value which it will produce. See Figure 7.

Figure 7 Initialization

SAS/ASSIST: Enter Numeric Expression

Select the desired data value or operator.
 To make a selection, place cursor and press ENTER.
 Use OK to save selections. Cancel to return without saving selections.

Enter Numeric Expression

Initialization for:

Select:

☐ Variable... ☐ Function... ☐ Constant...

☐ + ☐ - ☐ * ☐ / ☐ < ☐ =

☐ > ☐ ** ☐ <= ☐ ^= ☐ >=

Expression being built:

<Undo last operation> <Edit the expression>

<OK> <Cancel> <Reset> <Help>

The initialization screen is just another format similar to the where clause. Point and click on the button appearing before each selection as needed to build the necessary SAS expression. In our example the expression will read : REOPDT - DOB . This statement will produce the results in number of days. To produce the list in years add to the expression / 365.25 which means divided by 365.25. To produce the list in months add / 30.5. The resulting expression would read ((Reopdt-DOB)/365.25) Parenthesis are necessary to direct operators to perform the calculations in proper order. When the expression is written to direct the computer to do the operation desired point and click on OK. The previous screens will each be revisited for review. Point and click on OK until the subset screen returns once again.

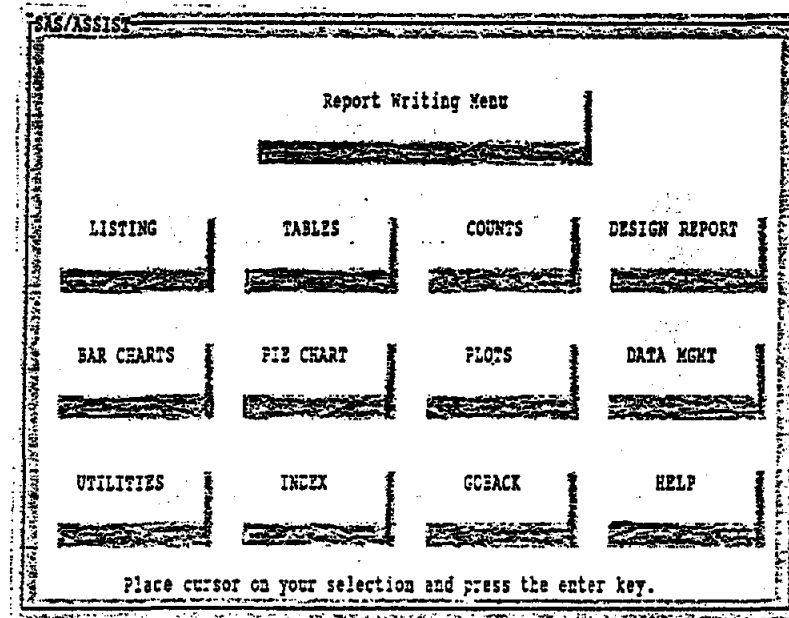
We have now supplied the necessary information to create a subset where only the patients who have had a reoperation will appear. Only the variables containing the information of name, date of first surgery, the variable indicating that they did have a reoperation, the date of reoperation, and the date of birth will appear. Also now there will be a new variable containing the age of the patient at the time of reoperation.

Depending on the version of SAS being used the Run command will be found in the menu bar on the bottom of the subset screen or under locals on the pull down menu . This will direct the program that was created to run, and thus create the new subset we have named REOPAGE. The data will appear on the screen in a spreadsheet format. The data can be viewed in this format, or be moved to other areas in SAS ASSIST to produce reports or statistics as needed.

We wish to return to the Primary menu. See figure 2. On this menu point and click on the icon Report Writing. Another screen of icons will appear. See figure 8.

We wish to return to the Primary menu. See figure 2. On this menu point and click on the icon Report Writing. Another screen of icons will appear. See figure 8.

Figure 8 Report Writing Menu



On this screen point and click on the icon Listing. This will open the screen to produce a list of the data . See figure 9

Figure 9 Listing Data

SAS/ASSIST: List a Data Set

<Active data set:> <Subset data:> NO

<Variables to appear in report:> -ALL-

Obs	Var 1	Var 2	Var 3	Var 4	Var 5
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

<Additional options>

<Run> <Customise> <Results> <Setup> <Goback> <Help>

If your active dataset is not selected, click on Active dataset , and a list of available datasets will appear. If the dataset was permanent look for that library name with the name

— given to the new dataset . If the dataset was made temporary it will be under the Work library. The libraries are listed in alphabetical order in the window. Click on the dataset. We will click on : Work AGEREOP. We will not need to subset the data again being we did subset it before we created the new dataset. Chose the variables to be listed by pointing and clicking on them while scrolling through the list. We will chose to print all of the available variables. This is the default and as noted in the window Variables to appear in report . Also available are additional options. Under this we will chose to click on Print number of observations. This will appear on our report as the number of patients who meet the criteria.

We will need to prepare our report to appear readable with titles and footnotes as desired. On the menu bar located at the bottom of the screens click on customize This will list the options available to prepare the report.

We will click on titles, and point and click to fill as many title lines as necessary. We will title ours :

Endocarditis Aortic Valve

Age at reoperation

Point and click on OK to return to the customize choices.

At this point you will also need to chose page dimensions. The default page dimensions are set up for the screen display on your computer. Printing options vary from platform to platform and you will need to setup the dimensions as required by your printer. When this is all completed we will point and click on Run, and the report will appear on our screen. PF3 will take us to the end of the report, and it is ready for printing.

To print use the menu bar at the top of the screen. Point and click on print. Another window will appear. Point and click on Print file. Another screen will appear requiring the name of the printer. Enter your form name or printer name, and do OK. A message indicating the number of lines printed will appear. This means the report has been created and is ready for printing, and only the final operation of freeing the print is necessary to link the report with the printer. Then point and click on file again and chose print and then free print. This will allow the printer to print the lines that have been ordered. It will say print form has been freed.

We have now produced a report where each of the patients who has had a reoperation will appear, and the created variable of their age at the time of this reoperation is also listed. See figure 10.

Figure 10 Printed Report

ENDOCARDITIS AORTIC VALVE AGE AT REOPERATION IN DAYS				
Patient identification number	Date of birth	Reoperation	Date of reoperation	Age at time of reoperation in days
807459	01/31/1939	1	01/24/92	19351
862693	05/18/1912	1	11/26/85	26855
903486	06/21/1936	1	08/25/92	20519

Conclusion

The examples shown here are only a small set of the tasks available in the SAS ASSIST program. SAS ASSIST is a very powerful program which is versatile and user friendly. The methodology used carries through all functions of SAS ASSIST, and many of the templates and selections we have covered are similar throughout the SAS program. It's pull down menus and on line help make it functional for the novice end user. It allows us to perform many of the data edit checks, report preparation, data manipulation, and basic statistics necessary to complete our job. SAS ASSIST can be comfortably used by the novice user or the power user, and produce the necessary results.