



Mu Sigma

MSU SAS Content

Introduction to SAS

Do The Math

Chicago, IL

Bangalore, India

www.mu-sigma.com

Proprietary Information

"This document and its attachments are confidential. Any unauthorized copying, disclosure or distribution of the material is strictly forbidden"

SAS is very powerful tool providing a wide variety of application in the field of data analysis

- ▶ SAS (Statistical Analytical System) was developed to analyze the Agricultural data for United States Department of Agriculture in 1966

- ▶ SAS is a set of solutions for enterprise-wide business users and provides a powerful fourth generation programming language for performing tasks such as these:
 - Data entry, retrieval, and management
 - Report writing and graphics
 - Statistical and mathematical analysis
 - Business planning, forecasting, and decision support
 - Operations research and project management
 - Quality improvement
 - Applications development

- ▶ With Base SAS software as the foundation, user can integrate SAS with many business solutions that enable user to perform large scale business functions, such as data warehousing and data mining, human resources management and decision support, financial management and decision support, and others

- ▶ SAS is used at more than 65,000 sites in 140 countries, including 91 of the top 100 companies on the 2013 Fortune Global 500® list

Numerous software are used in the academia and industry for data management, statistical analysis and optimization

Statistical Analysis				Win Cross
Predictive Modeling				
Decision Trees & Segmentation			Knowledge Seeker	CART
Forecasting & Simulation				
Optimization		LINDO	Evolver	Risk Optimizer
Campaign Management	Unica			

Base SAS provides the user capability to play around with the data in all the possible ways

- ▶ Base SAS software provides you with essential tools for basic data-driven tasks that are performed by a programmer:
 - Data access
 - Management
 - Analysis and presentation

Data Access:

You can access data that is stored almost anywhere, whether it is in a file on your system, or data that is stored in another database system. In almost any format, including raw data, SAS datasets and files created by other vendors' software.

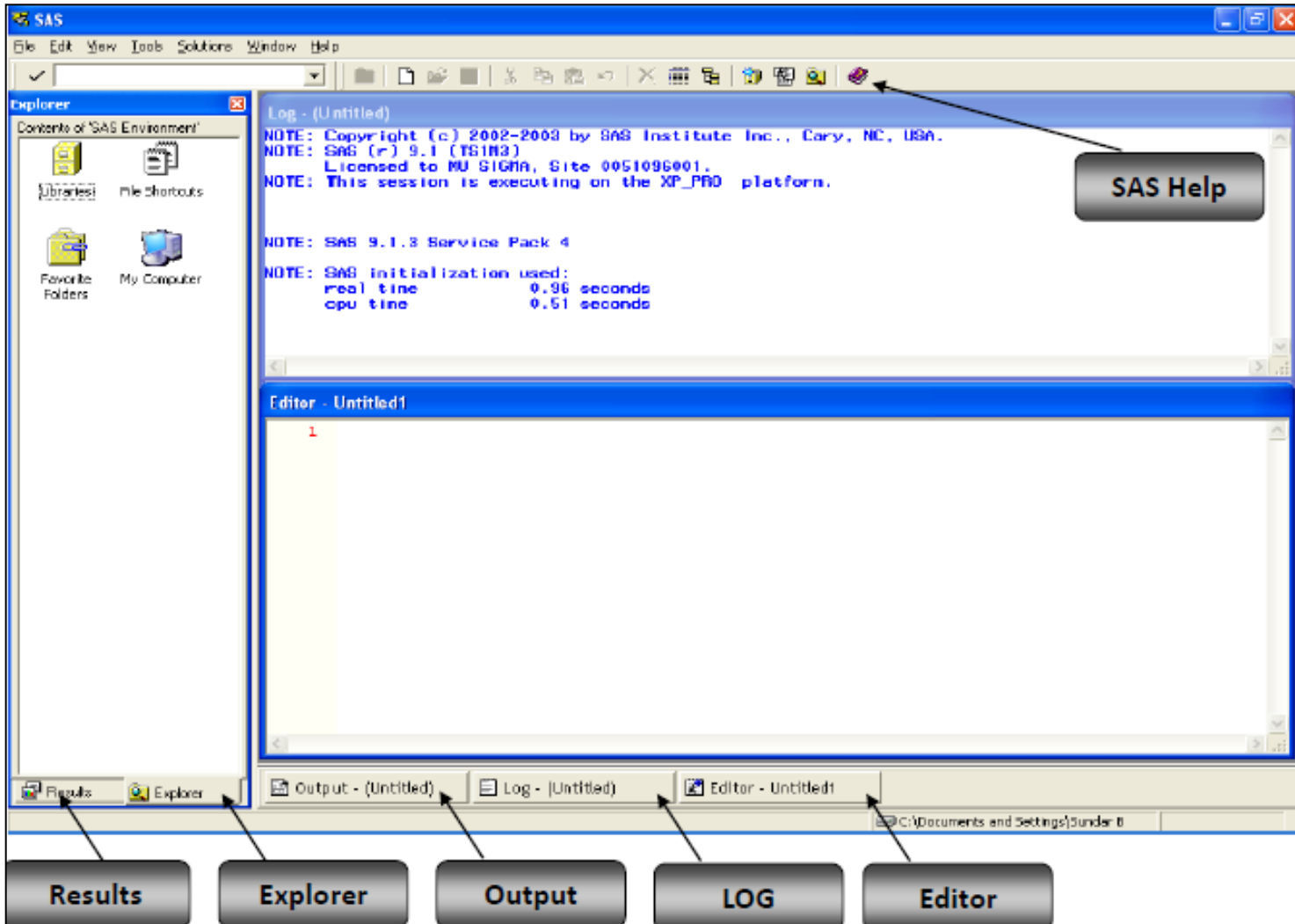
Managing data:

After you have accessed your data, you can use the SAS language to manipulate it. Format your data, create variables, use operators to evaluate the data values, use functions to create and recode the data values, subset the data and perform conditional processing, merge data sources, create, retrieve and update database information.

Analysis and Presentation:

Once your data is in shape, you can use SAS to analyze it and produce reports. Your SAS outputs can range from a single listing of datasets to customized reports of complex relationships.

There are five basic windows available in the SAS software – irrespective of whether its Windows SAS or Unix SAS



- ▶ Editor- to write the code
- ▶ Log- to check the code processing
- ▶ Output- to see the output, esp. in Proc step
- ▶ Explorer- to navigate and see the libraries and datasets
- ▶ Results- to see the previous outputs

Library is just an address to a physical location where a file can be stored

- ▶ SAS needs 2 things to locate a dataset
 - Where the dataset has been kept (known as Library/ Libref)
 - What is the name of the dataset (Dataset name)

- ▶ Syntax- Libref.Datasetname
 - E.g. Output.XYZ → This means that Output is the library name and XYZ is the name of the dataset present in the library
 - Output is pointing to a physical location on a drive/server (e.g. [E:\Mu sigma material\Drive\SAS](#))

- ▶ Since it is just an address so many libraries can point to the same location but not vice versa

- ▶ Libraries can be seen in Explorer window and which will display Datasets, views, catalogs

- ▶ SAS libraries can be categorized in two parts
 - Permanent library: Data exists even after the session is closed
 - Temporary library: Data gets deleted after session is closed

SAS programming can have 3 types of code elements



S.no.	Data Step	Proc Step	Macro
1.	Start with Data key word	Start with Proc key word	Start with % symbol
2.	For ETL purpose	For analyzing the data	For automation
3.	Can have a lot of variety while coding	Pre defined routines and syntax	Is a combination of Data &/or Proc steps
e.g.	<ul style="list-style-type: none"> Put your data into a SAS data set Compute values Check for and correct errors in your data Produce new SAS data sets by subsetting, merging, and updating existing data sets 	<ul style="list-style-type: none"> Create a report that lists the data Produce descriptive statistics Create a summary report produce plots and charts 	<ul style="list-style-type: none"> Any combination of Data and Proc steps can be used To automate the process To make the reusable code