

STAT 342

INTRODUCTION TO STATISTICAL COMPUTING AND
EXPLORATORY DATA ANALYSIS USING SAS

What is SAS?

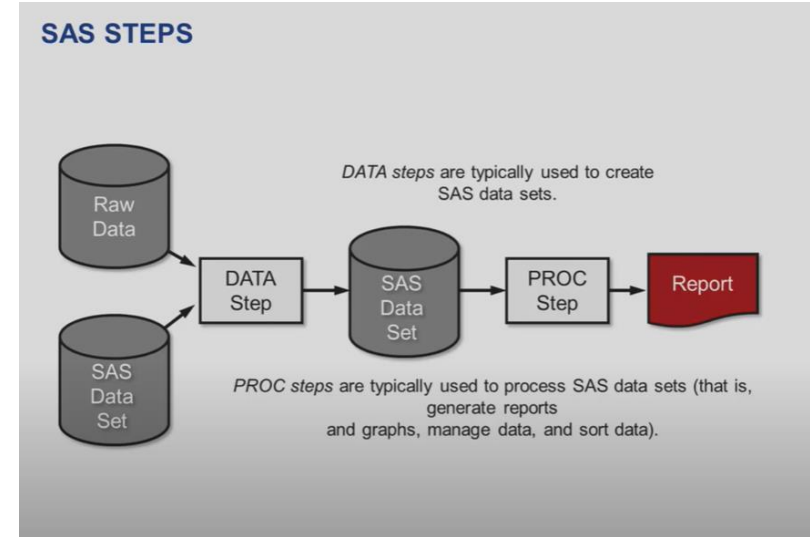
- SAS stands for **Statistical Analysis System**.
- SAS is a **command-driven software package** used for **statistical analysis** and **data visualization**.
- One of the most widely used statistical software packages in both industry and academia.
- Powerful tool for running SQL queries and automating user's task through macros.
- SAS offers descriptive visualization through graphs and there are various SAS versions provides reporting of machine learning, data mining, time series etc.
- Moreover, SAS is platform independent which means you can run SAS on any operating system either Linux or Windows.

SAS popular products

Name	Description
Base SAS	This offers hardware agility and integrates into all kind of computing environment.
SAS/GRAPH	This tool helps you to represent structured data into graphs.
SAS/STAT	This tool helps you to perform different types of regression, statistical analysis variance, regression, and psychometric analysis.
SAS/ETS	It is used for forecasting. Helps you to perform the time series analysis.
SAS/IML	Interactive Matric language is known as IML. This tool helps you to translate mathematical formulas into an innovative program.
SAS/QC	Use for quality control
SAS/PH	Clinical trial analysis
SAS/AF	Offers GUI applications facility

How to use SAS?

- To use SAS effectively, you need to follow these steps:



- **Access data:**
 - You can access data that is stored anywhere.
 - Ex: XLS/CSV file, SAS data base file, oracle file, raw database file etc.
- **Manage data (DATA step):**
 - Subset data on certain conditions, create variables, clean the data etc.
 - SAS has well-defined libraries and processes that makes the data managing process easy. This saves you from writing complex algorithms.

How to use SAS? Contd..

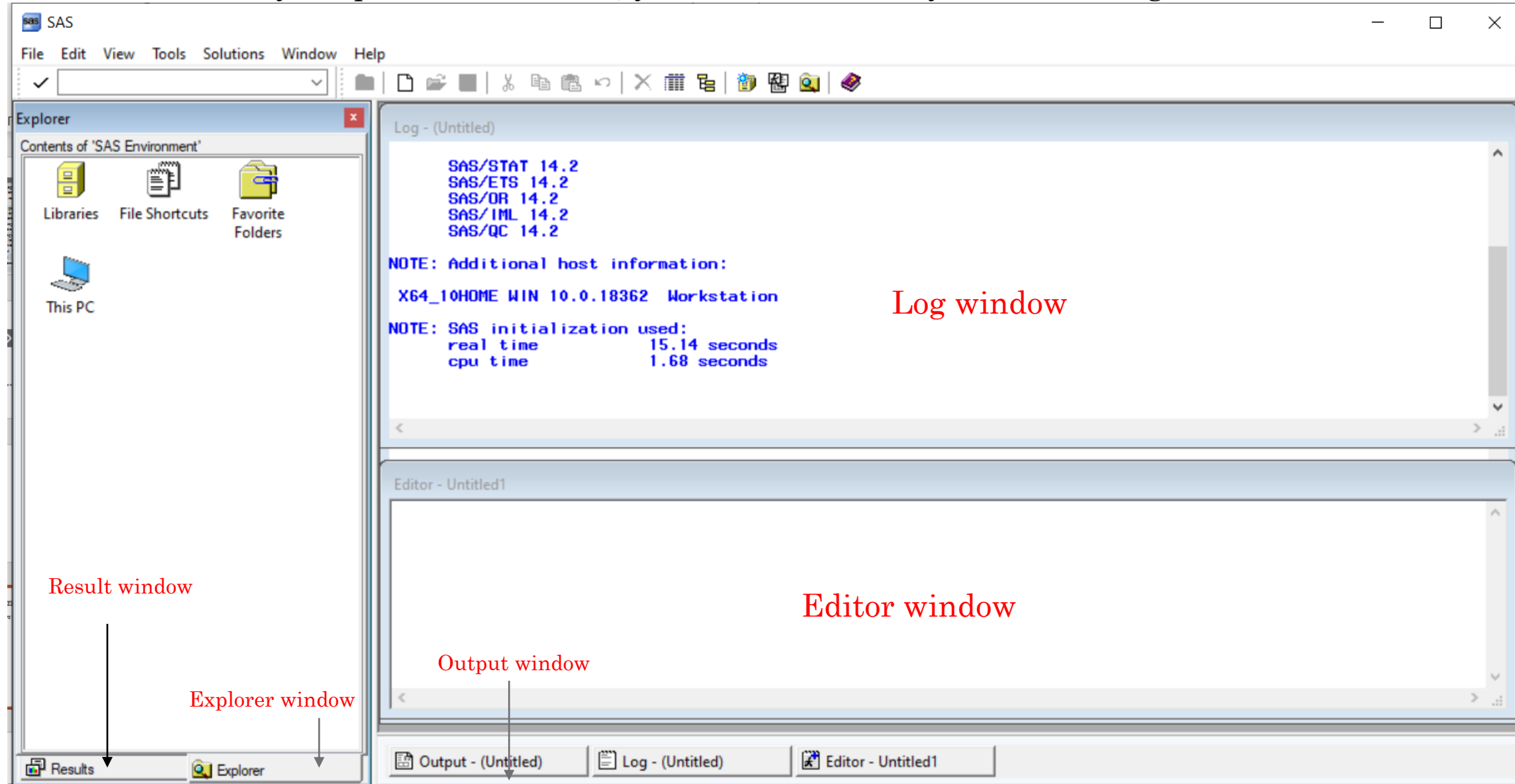
- **Analyze (PROC step):**
 - Do various kinds of analyze using SAS.
 - Ex: Regression, forecasting
- **Present report:**
 - Visualize data correctly with excellent presentation capabilities.
 - Ex: List reports, Summary reports, Graph reports, print reports.

Downloading and installing

- Please attend to the first tutorial next week
- For Windows and Linux users, follow the instructions at:
<https://www.sfu.ca/itservices/technical/software/sas-94.html>
- For Mac users: https://www.sas.com/en_ca/software/university-edition/download-software.html

SAS Windowing Environment

- When you open SAS software, you will immediately see something like this:



SAS Windowing Environment

- **Log window:**
 - Enables you to view messages about your SAS session and your SAS programs
 - Log helps you identify the error
- **Editor window:**
 - Enables you to enter, edit, submit, and save SAS programs.
 - After typing a SAS program, you can run it by clicking on the "running man" icon.
- **Output window:**
 - Enables you to view output from your SAS programs
 - Two types of output
 - Listing output
 - HTML output (default output)
- **Result window:**
 - Enables you to view specific output from a SAS program(s)
 - The Results window uses a tree structure to list various types of output that might be available after you run SAS
- **Explorer window:**
 - Enables you to manage your files in the windowing environment
 - View lists of your SAS files, create new SAS files, view, add, or delete libraries, move, copy, and delete files, etc.

Basic steps in SAS programs

- SAS program consists of three necessary steps
 - DATA (or data) step
 - PROC (or proc) step (short for procedure)
- DATA step:
 - Typically used to read or write the data, manipulate the data, and perform calculations.
 - Begins with DATA statement.
- PROC step:
 - Typically used to process SAS datasets in analyzing.
 - Run one or more of its procedures to produce reports, summarize the data, and much more.
 - Begins with PROC statement.
 - Display the data.
- The end of the DATA or PROC steps are indicated by:
 - RUN statement – most steps
 - QUIT statement – some steps

Basic steps in SAS programs contd...

- Example

Example1.sas

```
/* Lecture 1, example 1: Read the dataset and print */  
  
data studgrade; /* data step: assign a name for the dataset and enter the data */  
input StudID Midterm Final Grade $; /*Variables in the dataset and enter the data*/  
datalines;  
101 98 86 A  
102 49 60 C  
103 98 80 A  
104 90 98 A+  
105 60 80 B+  
106 . 80 C-  
;  
run; /*end the data step*/  
  
proc print data = studgrade; /* proc step: perform any analysis and print the dataset.  
Here we are only printing the dataset*/  
run;
```