

Self Learning Assessment

Program Code: CO3KB

AY(2024-25)

Course:Database Management System (DMS) – 313302

Title : How SAS is related to SQL

Introduction to SAS programming

Statistical Analysis System (SAS) is a software suite that has been developed by SAS Institute, one of the leaders in analytics. It is useful for performing advanced analytics, multivariate analyses, business intelligence, data management functions, and also for conducting predictive analytics.

1.Use of SAS:

SAS is used by many top organizations which include Google, Facebook, Twitter and Accenture for answering the many questions which persist at a business and organizational level and is also used for generating informative reports. It helps in business forecasting, quality improvement, operations research and project management. It is particularly useful in applications development and graphics design. This article should help you get an intuition about the working of SAS applications and a hint about its raw power.

2.Types of SAS software:

There are various types of SAS software that are available in general there are 4 types of SAS software which are given on next page.

- 1. SAS for Windows**
- 2. SAS EG (Enterprise Guide)**
- 3. SAS Enterprise Miner (EM)**
- 4. SAS Stat Software**

The most used software among the above types is SAS for Windows. It can be easily deployed and allows for modernization of your SAS environment by enabling grid processing.

3.SAS Libraries:

Libraries are the storage mechanisms in SAS. It can store the programs which can be similar in nature. Broadly speaking there are 2 types of libraries viz. temporary and permanent. Temporary libraries are those that exist only during the current SAS session or job. Permanent libraries, as the name suggests are stored in the external storage and are not deleted at the end of a session.

4.SAS programming workflow:

SAS programming applications are characterized by the flow control specified in the diagram.

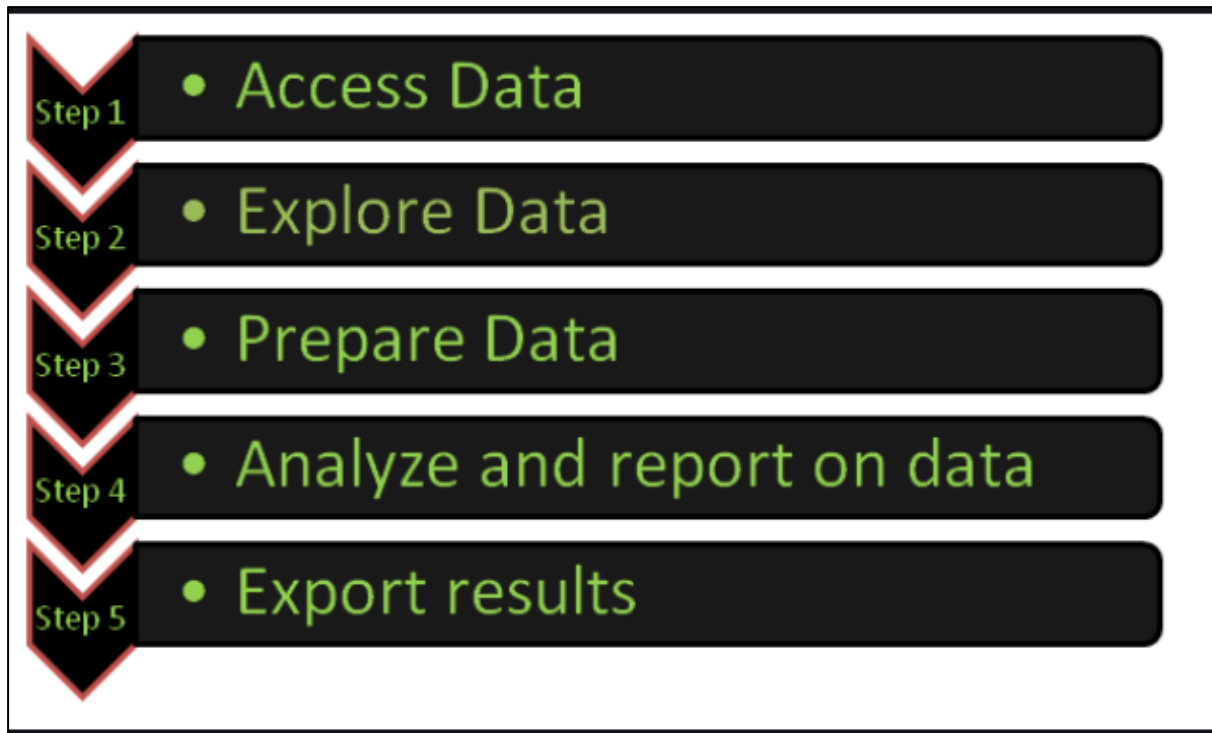


Figure :1

In the first phase of the above diagram a SAS program will read the data which is usually stored as worksheets in an Excel Workbook. A connection to the workbook is made in this phase and the data is imported in a format suitable for usage in the following phases.

5.Explore Data:

In the second phase we check for inconsistencies or incorrect values. Here frequency reports and summary statistics may also be generated. Looking at these reports and summaries at first glance we might find some inconsistencies right off the bat. For example all the label names of a categorical attribute except for one might be in capitalized letters. Another example of an incorrect value is a negative value for the minimum number of runs scored by a batsman during a string of matches. The first 5 rows are also shown in this phase.

6.Prepare Data:

In this phase the issues found in the earlier phase are fixed. Here we create new columns with existing values or concatenate them. Conditional processing is also done in this phase. Certain

tables are also joined to provide more consistent representations of data. These tables have clean and validated values.

7. Analyse and Reporting on data and exporting the results:

The last two phases are combined into one in SAS programming applications. Here we create maps based on geographical coordinates, summary statistics etc. which are some of the output models used for displaying results. We can also save the output models created in an Excel Workbook and download it for later offline use.

Also, there are 200+ components present in SAS, Some popular components are:

- **Base SAS**
- **SAS/GRAPH**
- **SAS/STAT**
- **SAS/INSIGHT**
- **SAS/PH**
- **SAS/ETS etc.**

This is the overall framework of a SAS programming application. Hope this article helps you to get a clear view of the workflow of a SAS program.

8.SAS Refresher

We will start by setting our library name, which opens a directory to the location where our SAS data files are stored.

```
*assign libname  
LIBNAME idre 'C:';
```

Figure : 2

SAS also allows you to clear a particular libname or use the `_all_` keyword to clear all assigned libnames.

```
*clear libname;  
LIBNAME idre CLEAR;  
LIBNAME _ALL_ CLEAR;  
  
* reassign library;  
LIBNAME idre 'C:';
```

Figure : 3

As we will be using several different datasets in the seminar today, let's also cover how to create new permanent and temporary datasets from the data files you have been provided.

```
*permanent dataset;  
DATA idre.new;  
SET idre.charities;  
RUN;  
  
*temporary dataset;  
DATA new;  
SET idre.charities;  
RUN;
```

Figure : 4

9.SAS options

SAS includes a large suite of system options that will affect your SAS session. Specific options are invoked by default when you open SAS. The options can vary depending what computing environment you are using (e.g. Windows, Unix). The `OPTIONS` procedure lists the current settings of SAS system options in the SAS log.

9.1. Autocorrect

The AUTOCORRECT option is turned on by default and allows SAS to correct syntax with small mistakes like a misspelled keyword. In the first example below, the DATA keyword is misspelled to DATE. When the option is invoked, you will see that in the Log (shown below), SAS issues a warning it assumed that the keyword was misspelled and continues executing the procedure.

```
*autocorrect option;  
OPTIONS AUTOCORRECT; /*default*/  
PROC FREQ DATE=idre.charities;  
TABLE code;  
RUN;  
  
OPTIONS NOAUTOCORRECT;  
PROC FREQ DATE=idre.charities;  
TABLE code;  
RUN;
```

Figure : 5

9.2. FMterr

The FMterr option controls whether SAS will issue a warning for incorrect formats being used for variables. In this case, the default is for SAS is to error and stop processing the executed procedure. In the first example, the default option is invoked and as you can see below SAS issues a warning that the format used could not be found. However, in the second example where we tell SAS to not issue an error (NOFMterr), SAS ignores the incorrectly used format and will the execute the command without the format.

```
*format error;  
OPTIONS FMterr;/*default*/  
PROC PRINT DATA=idre.charities;  
FORMAT code $code.;  
RUN;  
  
OPTION NOFMterr;  
PROC PRINT DATA=idre.charities;  
FORMAT code $code.;  
RUN;
```

Figure : 6

10.Color-Coded Syntax.

When executing code in SAS in the Enhanced Editor you will notice some color coding. Color coding program components will help you more easily diagnose syntax errors, and when you first start with SAS you will make many mistakes. Take a look at the example syntax below copied from the Enhanced Editor window. Here you will see 5 different colors automatically generated by SAS. For example you will see that keywords like DATA, CLASS, MODEL are all highlighted in blue. If you use the wrong keyword with a procedure, the keyword will often remain black like the variable names because SAS does not recognize it. Options like SOLUTION are also considered keywords.

```

TITLE "REGRESSION";
PROC GLM DATA = newdata;
CLASS female prog ;
MODEL read = write female math prog /SOLUTION;
FORMAT female female. prog prog.;
RUN;

TITLE "REGRESSION;
PROC GLM DATA = newdata;
CLASS female prog ;
MODEL read = write female math prog /SOLUTION;
FORMAT female female. prog prog.;
RUN;

```

Figure : 7

Application of the Self Learning Assessment

- Core SQL Skills: Test basic SQL operations (library functions) using PROC SQL in SAS. Evaluate your ability to retrieve and manipulate data.
- .Advanced SQL Techniques: Practice creating/modifying datasets and optimizing query performance.
- SAS Integration: Incorporate SAS functions and macros into SQL queries. Assess how well you use these for complex tasks.
- Reflection: Review areas of difficulty and seek better solutions to improve efficiency.

References

1. <https://stats.oarc.ucla.edu/sas/seminars/sas-programming-basics/>
2. <https://www.geeksforgeeks.org/introduction-to-sas-programming/>

.....

Course coordinator: Vidya S Lunge

Date of Submission:

Submitted By:

Name: Amaan Abrar Kazi

Rollno: 55

Enrollment no:23150920711

Class:P10