



SPSS to SAS Grid

Mark Ezzo

VINCI SAS Administrator



April 13, 2017





Terminology (SPSS – SAS)

Some vocabulary differences exist between SAS and SPSS. To help you translate from one language to the other, here is a brief dictionary of analogous terms:

SPSS term Analogous - SAS term

active file - *no analogous term in SAS programs*

no analogous term - temporary SAS data set

case - observation

command - statement

Data Editor window - Viewtable window

file handle - libref

function - function

input format - informat

numeric data - numeric data

output format - format

procedure - procedure

save file - permanent SAS data set

SPSS data file - permanent SAS data set



Terminology (Continued)

SPSS portable - file SAS transport file

SPSS Viewer window - SAS Output and Log windows

string data - character data

syntax programming - statements or code

Syntax Editor window - Enhanced Editor window (Display Manager SAS only)

syntax file - a program

system file permanent - SAS data set

value label user-defined format

variable - variable

variable label - label

no analogous term - DATA step

no analogous term - PROC step



Concepts

Active files The concept of an active file in SPSS has no precise equivalent in SAS programs. When you read data in an SPSS program, SPSS creates an active system file. This active file is similar to a temporary SAS data set because, by default, it exists only for the duration of the SPSS session, just as temporary SAS data sets exist only for the duration of a SAS session. However, SPSS has only one active file at a time, while SAS can have any number of temporary or permanent data sets. When you run an analysis in SPSS, the data must come from the active file. When you run an analysis in SAS, by default SAS will use the data set most recently created. But you can easily use any other SAS data set including the permanent SAS data set you created last year and haven't touched since. In SAS programs, all SAS data sets are always active.

DATA and PROC steps The SAS language has some concepts that have no parallel in SPSS, such as DATA and PROC steps. All SAS programs are divided into these two types of steps. Basically, DATA steps read and modify data while PROC (short for procedure) steps perform specific analyses or functions such as sorting, writing reports, or running statistical analyses. SPSS programs do the same types of operations but without distinct steps.



Concepts (Continued)

DATA and PROC steps

The SAS language has some concepts that have no parallel in SPSS, such as DATA and PROC steps. All SAS programs are divided into these two types of steps. Basically, DATA steps read and modify data while PROC (short for procedure) steps perform specific analyses or functions such as sorting, writing reports, or running statistical analyses. SPSS programs do the same types of operations but without distinct steps.





Live Demo

- We will be using Enterprise Guide on the Grid to display how to use SAS most seamlessly.
- We will also demonstrate the Wizards you can use to create and keep SAS code via parameter entry.
- Finally, we will view SAS Enterprise Miner to demonstrate Grid-Enabled rapid Data Mining and Modeling





SAS EG 7.1

- **Standard for coding and Grid Access**
- **Enhanced and automatic features**
- **Easy to Configure and Use**
- **Wizards for many tasks**
- **Able to develop SQL Pass-Through Queries (recommended for the DB)**
- **Automatically adapt for parallel processing**
- **Live Demonstration!**





Parallel Process

In EG, via the Analyze Program feature, you have the ability to adapt a program for Parallel Processing. This allows much of the program to run simultaneously in multiple jobs slots. This is much faster and more efficient than a Linear Program utilizing one job slot.

Can be used for ETL or Analysis!



Parallel Process

Considerations:

- There are costs to parallel processing so the more complex the program/data the greater the benefit.
- Small programs do not have the need
- You may have to do a few modifications, depending upon the original program (libnames, etc.)
- Can greatly reduce processing time!
- Runs in both EG and GSUB (batch)
- Example



SAS Gsub Example

- Here is the command we will be using for batch processing on the Grid:
- **sasgsub -GRIDSUBMITPGM**
/data/prod/admin/VHASLCEZZOM/Programs/Program.sas
-
- It breaks down as follows:
-
- **sasgsub** is the command
- **-GRIDSUBMITPGM** is the option to submit a program
- **/data/prod/admin/VHASLCEZZOM/Programs/**is the directory where the program is located
- **Program.sas** is the program you wish to run
- **Live Demo!**



SAS Enterprise Miner

- **SAS/EM is the state of the art Data Mining and Modeling Software.**
- **Friendly User interfaces allows less programming and more analysis.**
- **Access via IE through this link:**
- **<http://vhacdwdwhsasmid3.vha.med.va.gov:7980/SASEnterpriseMinerJWS/Status> or**
- **<http://vhacdwsasmid1.vha.med.va.gov:7980/SASEnterpriseMinerJWS/Status>**
- **Quick Demo**



Good Gridding!

Thank you for attending.

Please contact VINCI SAS Administrators:

VINCI SAS Admins

VINCISASAdmins@va.gov

with any questions or comments.