Computer Based Statistics: Introduction in SAS

Oliver Wuensche

25.04.2006

What is SAS?

- SAS stands for **S**tatistical **A**nalysis **S**ystem
- Powerful and flexible tool for many purposes: data management, estimation, optimization, visualization...
- Mixture of 'easy-to-use' procedures and manual programming
- Macro language for recurring code sequences

The SAS User Interface

1. Log window:
   - Blue is ok (in general)
   - Green is a warning (should be checked)
   - Red is an error message (**do not ignore**!)
2. Output window: Results are printed in this screen
3. Editor window: Write your program commands in here.

A few general concepts of SAS

- Data steps: create data sets, create new variables, subset existing datasets,...
- Procs: analyze your data, estimation, visualization,...
- `Proc SQL`: gives access to SQL commands, merging different data sets,...
- `Proc IML`: Interactive Matrix Language, access to matrix operations (similar to MATLAB or GAUSS),...

- Macros: often recurring pieces of code can be written as macros, very efficient programming!!!

Basic handling of data

To use a SAS-dataset saved on your harddisk or to create a permanent dataset, assign a library pointing to the respective folder:

`libname` *name* "*drive*:\ *folder*";

Now, you can find all datasets already in *folder* or newly created datasets in the respective library *name* in your SAS-Explorer.

Working with temporary files vs. permanent files

The name of a SAS dataset consists of two parts:

library.dataset

Permanent dataset: choose *name* as your library, if you want to write the dataset permanently into *folder*

Temporary dataset: skip the library or choose `work` as your library to create temporary datasets.

Temporary datasets will be lost when you terminate SAS.

Getting data into your SAS system

Reading Ascii data into SAS via Data Steps and a first introduction into `Proc SQL`

Example: `SAS_session1.sas`

Creating a table with SQL code

```
Proc SQL;
create table name as select
variables
from dataset;
quit;
```

Merging two tables with SQL code

```
Proc SQL;
create table name as select
a.variables
,b.variables
from dataset1 a dataset2 b
on a.idvar1=b.idvar2;
quit;
```

Some procedures (Procs) for descriptive statistics

- Proc MEANS: calculates mean, median, standard error, confidence limits for the mean,...
- Proc CORR: calculates several correlation coefficients
- Proc FREQ: one-way frequency table, n-way contingency tables, several test statistics
- Proc UNIVARIATE: several descriptives, histograms, kernel densities,...

Example for Proc MEANS

`Proc MEANS data=`*dataset*;

`var` *variables*;

`output out=` *outdataset* `mean=`*meanvar* ...;

`run`;

⇓

Example: `SAS_session2.sas`