SPSS WITH SYNTAX

English versions 0.1 (May 1 2012)

Harry B.G. Ganzeboom

FILE MANIPULATIONS

Get file
Add files
Match files
Save outfile
Save translate /outfile=

VARIABLES: NAMES, LABELS, LIST, MACRO

Var Labels
Value labels
Add value labels
Delete variables
Display labels / dictionary
Varlist
Define … !enddefine.

CREATING NEW VARIABLES AND CHANGING EXISTING VARIABLES

Recode
Recode .. into
Autorecode /into

Compute
- Often used numerical and statistical functions
- String variable manipulations

Count

Rank /prop
Desc /save

Aggregate /break=varlist /newvar=mean(var).
Aggregate /outfile="file-location"
  /break=varlist /aggvar=mean(var).

SELECTING AND FILTERING CASES

There are two syntax statements for (temporary) omitting cases from your active data-file:

FILTER BY filter.

In which <**filter**> is a 0/1 variable that denotes the parts of the file that you want to omit temporarily. E.g.

FILTER BY female.

would analyze only women, if female=1 for women and 0 otherwise. The omission is temporary and stays active until you state:

```
FILTER OFF.
```

The alternative is:

```
SELECT IF (condition).
```

This will permanently omit cases from the file that do not meet the condition. Selecting can be made temporary by:

```
TEMP.
SELECT IF (condition).
```

In this case the selection is only valid for the next procedure.

Examples are:

```
TEMP.
SELECT IF (AGE GE 25).
SELECT IF (AGE LE 74).
SELECT IF (EDUC GE 1).
Freq INCOME.
```

This would display INCOME for persons between 25 and 74 years of AGE, with EDUC >= 1.

Permanently **SELECT**ing cases reduces the file size, which makes processing quicker. **FILTER** just scratches them out, but they keep being there, eating up processor time.

FILTER is more convenient to use than **SELECT IF**, but creating the filter-variable can be complicated. Multiple **FILTER** statements are not allowed.

LOGICAL CONDITIONS

In **IF**, **DO IF** and **SELECT IF** we need to specify logical conditions. Logical operators in SPSS can be either word of symbols, in particular:

GE	**>=**	Greater or Equal
EQ	**=**	Equal
LE	**<=**	Lower of Equal
NE	**<>**	Nor Equal
GT	**>**	Greater Than
LT	**<**	Lower Equal
OR	**|**	Or
AND	**+**	And

Note that the words are reserved expressions in SPSS – as a consequence it is not a good idea to call a variable LT (.. leeftijd ..).

SORTING FILES

Sorting files by a variable can be very helpful to inspect data, look for outlying cases etc. Sorting is mandatory before three procedures:
- Splitting files
- Matching files
- Aggregating files

If you want to sort by one column it is useful to do it by a mouseclick. Put your cursor on the column header in the data sheet, variable view, right click and sort ascending / descending.

The appropriate syntax is:

```
Sort cases by V1 V2.
Sort cases by V1 (a) V2 (d).
```

Sorting is by default ascending (from low to high).

SPLITTING FILES

The basic syntax here is:

```
Split files by V1 V2.
.. procedures ..
Split file off.
```

SPSS will not do the <procedures> for all distinct groups define by V1*V2 and assemble the output in a single table. This is very useful (and fast) when working on data with comparative structure, such as country comparisons.

Note that the file needs be have been sorted beforehand. **Split file** will process groups as encountered sequentially in the file and the output can become quite large if you fail the sort before splitting!

REGRESSION IN SPSS

SPSS offers several options to obtain a regression analysis:

> ➢ Analyze > Regression > Linear > REGRESS
> ➢ Analyze > General linear model > UNIANOVA
> ➢ Analyze > Generalized linear model > linear response > GENLIN

These programs should give the same results. Important differences in use are:

REGRESS
- Can use data with pairwise deleted missing values (but does this not very adequately).
- Can do stepwise modelling and compute F-change.
- Has better regression diagnostics than the other programs.
- Does provide standardized coefficients.

UNIANOVA
- Is oriented towards analyses of analysis of (co)variance and experimental design. Note that printing the coefficients is an option, not the default.
- No standardized coefficients. You should standardize variables yourself to get the standardized coefficients.
- Can easily shift between discrete factors (BY) and continuous covariates (WITH). Automatic dummy variables.
- Does automatic interaction expansion.
- Can print and graph adjusted means.
- Does not allow you to specify the reference category in a set of dummy variables. The default is the <u>last</u> category. However, you can use your own dummy variables or order your categorical variable to your convenience.

GENLIN
- For OLS basically the same as UNIANOVA; extensions to binomial, multinomial, ordered logistic and other types of models.
- For regression the possibilities are basically the same as UNIANOVA, but GENLIN does not allow you to specify the place of reference category (last or first) in a set of dummy variables.

Basic syntaxes:

```
REGRESS /dep=Y / enter=X1 X2.

REGRESS /miss=PAIR /stat=DEF TOL COLLIN CHANGE.
  /dep=Y /enter=X1 /enter=X2.

UNIANOVA Y by X1 with X2
   /print=parameter
   /design X1 X2.

UNIANOVA Y by X1 with X2
   /print=parameter descriptives
   /emmeans=tables(X1)
   /design X1 X2 X1*X2.

GENLIN Y by X1(order=ascending/descending) with X2
   /model X1 X2 X1*X2.
```

COMMON TASK IN DATA ANALYSIS

Creating a multiple indicator index variable
- Factor analysis
- Reliability analysis

Regression with dummy variables

Regression with interactions

Missing values analysis and imputation

Graphs