EU-Labour Force Survey

Guide to Generating a Stata Dataset (.dta) from a SPSS dataset (.sav)

Content

I. Overview

This document aims at assisting you in generating a Stata .dta file from the SPSS .sav files you have generated with the help of setups from GESIS German Microdata Lab. It is recommended that you generate a single .sav file in SPSS (with the `add files` command) which contains all countries of interest before carrying through with this exercise. In order to generate a .dta file you will have to carry out these three steps:

1) Recode user missing values to valid cases in SPSS
2) Export .sav file to .dta file
3) Recode missing values in Stata

By carrying out these three simple steps you will generate a .dta file complete with variable and value labels as well as missing values which follow a uniform coding scheme. Note however that in Stata labels may not be longer than 80 characters, so some variable labels will be cut off.

II. Step by Step description

1) Recode user missing values to valid cases in SPSS

   Once you have generated a .sav file open this file with SPSS and then open a Syntax editor. In the Syntax editor type the following command line:

```
Missing values all ().
```

Execute the command and save your data file. Now all user missing values have been recoded to valid cases.

2) Export .sav file to .dta

There exist a number of options for transforming a .sav file to .dta format.

a) You can either save your data in Stata format in SPSS by selecting `File Save as` and choosing Stata 8 format.
b) Or use conversion software such as StatTransfer.
c) Or load the .sav file directly into Stata with the `usespss` command (`findit usespss` to install). Then save it as .dta file.

3) Recode missing values in Stata

After loading your newly created .dta file into Stata this file will now have only system missing values (.) defined as missing. The following command line will help you to convert the remaining missing value categories.

```
foreach x of varlist VAR1 VAR2 {
recode `x' -1=.a -2=.b -3=.c -4=.d -9=.e
}
```

You will have to enter the names of all non string variables included in your dataset manually, replacing the `VAR1 VAR2` placeholders.[1] Also since `STARTIME` and `LEAVTIME` have valid cases on values -1 through -3 in some instances they must be recoded as follows:

```
foreach x of varlist STARTIME LEAVTIME {
recode `x' -101=.a -102=.b -103=.c
}
```

Additionally you will have to consider string variables separately. The only string variables with missing values are `REGION`, `REGIONW` and `REGION1Y.` Executing the command line below will recode these variables, however you will have to install strrec (ssc install strrec) beforehand.[2]

```
foreach x of varlist REGION REGIONW REGION1Y {
strrec `x' ("-1"=".a") ("-2"=".b") ("-3"=".c"),sub copyrest string
replace
}
```

[1] Note that when using StatTransfer or usespss all or some variable names will be capitalised. And as Stata commands are case sensitive variable names will have to be capitalized. In case you find this pesky you can convert your variable names to lower case (renvars, lower) before running the recode command.
[2] Note however that Stata does not recognize these values as missing in string variables. The only valid missing category for string variables is an empty string ("").

After executing the above commands all missing values will be recoded according to the scheme below:[3]

```
.a    no answer
.b    not applicable
.c    not available
.d    not specified
.e    other
```

GESIS - Leibniz-Institute for the Social Sciences
German Microdata Lab
Square B2, 1
D-68159 Mannheim
Tel: 0621-1246-265
Fax: 0621-1246-100
http://www.gesis.org/gml
E-Mail: gml@gesis.org

[3] Stata can not attach value labels to missing values so you will have to store this list of value labels somewhere.