

HOW TO CREATE YOUR OWN DATASET USING SPSS SYNTAX FILES

Editing SPSS Data Definition Files

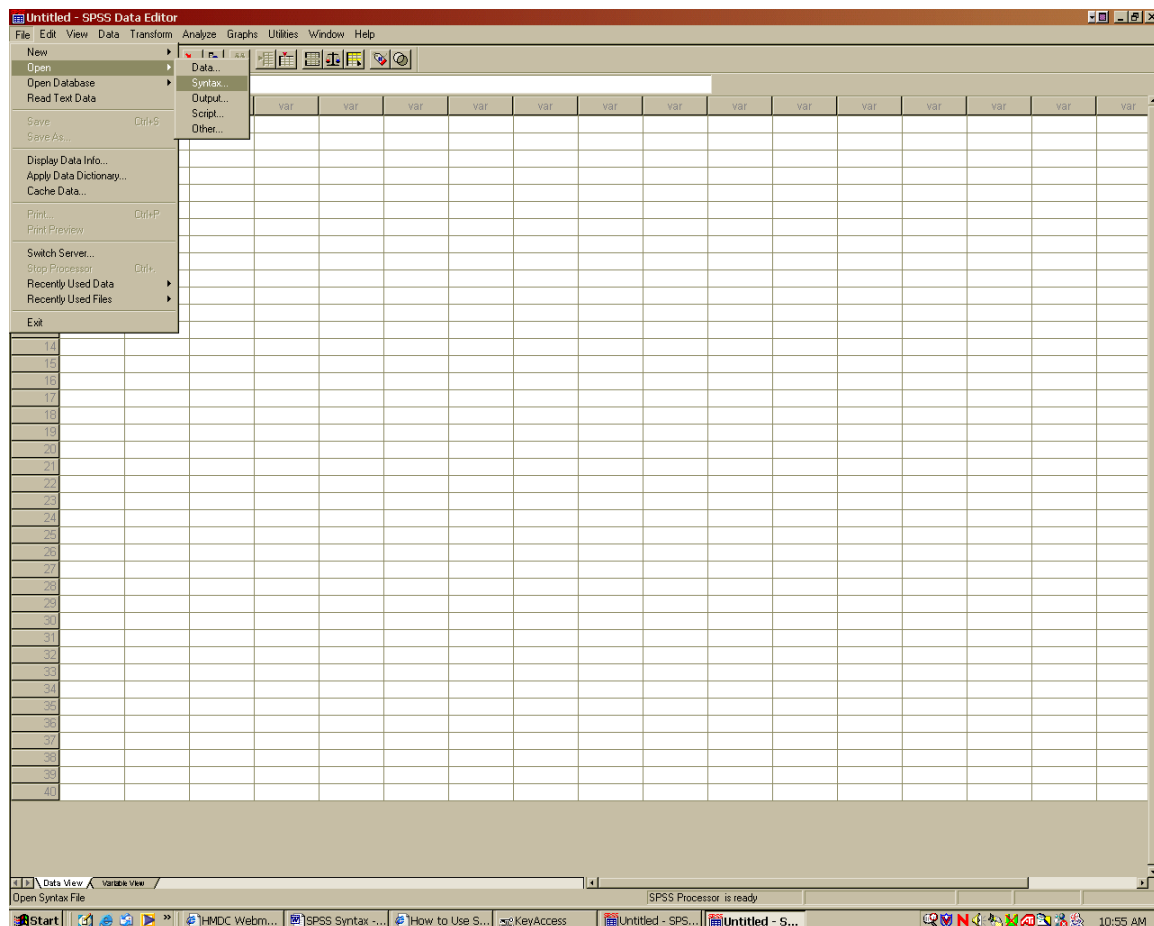
Many datasets currently available for download at the virtual data center are already in SPSS or STATA format. This means that students and faculty can use SPSS and STATA to manipulate and analyze these datasets immediately after downloading them. Other datasets on the website are subsetable, which allows users to download only the variables that are of interest to them in the format they desire.

However, others are neither subsetable nor preformatted. Instead, many simply appear to be “a pile of numbers” without any variable names or identifiers. Fortunately, in many cases these datasets are accompanied by what are known as “SPSS Syntax Files” or, more generally, data definition files. You can recognize these files because they have the “.sps” extension.

Data definition files, after some easy manipulation, can format the datasets for the user. In other words, the syntax file can divide the “pile of numbers” into separate cases and variables. Furthermore they can also label the variables to make use of the dataset easier.

Note that SPSS Syntax files are text files that can be edited using any standard word editor, but we recommend using the syntax editor in SPSS. This document will show you how to do this.

In order to open your syntax file open SPSS on your machine and then go to File – Open – Syntax... as shown in the figure below. Locate the syntax file on your computer wherever you saved your dataset files.



When you open your data definition file in the SPSS editor it will probably look something like this

```
SPSS DATA DEFINITION STATEMENTS FOR ICPSR 2790
WORLD VALUES SURVEYS AND EUROPEAN VALUES SURVEYS,
1981-1984, 1990-1993, AND 1995-1997
1ST ICPSR VERSION
FEBRUARY, 2000
FILE GENERATED: 04FEB00

This SPSS setup file contains the following statements:

DATA LIST: Specifies the input data file to be read and assigns the
name, type, decimal specification (if any), and identifies variable
beginning and ending column locations. Users must replace
"file-specification" with a complete statement of the location of the
data file.

VARIABLE LABELS: Assigns descriptive labels to variables. Variable
labels and variable names may be identical for certain variables.

VALUE LABELS: Assigns descriptive labels to codes in the data file.
Value labels may not be present for all variables in the data file.

MISSING VALUES: Declares user-defined missing values. These values
can be treated specially in data transformations, statistical
calculations, and case selection. Missing values may not be present
for all variables in the data file.

NOTE: Users should modify this SPSS setup to suit their specific
needs. The MISSING VALUES section has been marked by SPSS comment
statements. To include this section in a final SPSS setup, users
should remove the SPSS comment statements.

* SPSS DATA LIST COMMAND.

DATA LIST FILE="file specification" /
V1 1-1      V2 2-3      V3 4-7
V4 8-9      V5 9-9      V6 10-10
V7 11-11    V8 12-12   V9 13-13
V10 14-14   V11 15-15  V12 16-16
V13 17-17   V14 18-18  V15 19-19
V16 20-20   V17 21-21  V18 22-22
V19 23-23   V20 24-24  V21 25-25
V22 26-26   V23 27-27  V24 28-28
V25 29-29   V26 30-30  V27 31-31
V28 32-32   V29 33-33  V30 34-34
V31 35-35   V32 36-36  V33 37-37
V34 38-38   V35 39-39  V36 40-40
V37 41-41   V38 42-42  V39 43-43
V40 44-44   V41 45-45  V42 46-46
V43 47-47   V44 48-48  V45 49-49
V46 50-50   V47 51-51  V48 52-52
V49 53-53   V50 54-54  V51 55-55
V52 56-56   V53 57-57  V54 58-58
V55 59-59   V56 60-60  V57 61-61
V58 62-62   V59 63-63  V60 64-64
```

There are five primary commands that you are likely to need in editing an SPSS Syntax File. They are 1) File Handle, 2) Datalist, 3) Variable Labels, 4) Value Labels, and 5) Missing Data. This manual will show you how to use each of these commands to either modify an existing SPSS syntax file, or create your own.

1. File Handle

The file handle command is probably the most important command in the entire syntax code. This command instructs SPSS as to where to find the data you need it to format. You will need to edit this command to use your dataset. In order to do so you should scroll down to the command beginning with DATA LIST FILE, which is usually the first command in the data definition file after the information about the dataset itself.

Change this command to indicate where your dataset is located. For example, if your dataset has been saved in the "My Documents" directory of your computer you could have something that reads

```
DATA LIST FILE="C:\My Documents\dataset.dat"/
```

where dataset.dat is the actual name of the dataset

Alternatively you could also manually open the data in SPSS and then retrieve the SPSS code for the location of the file by using the output window.

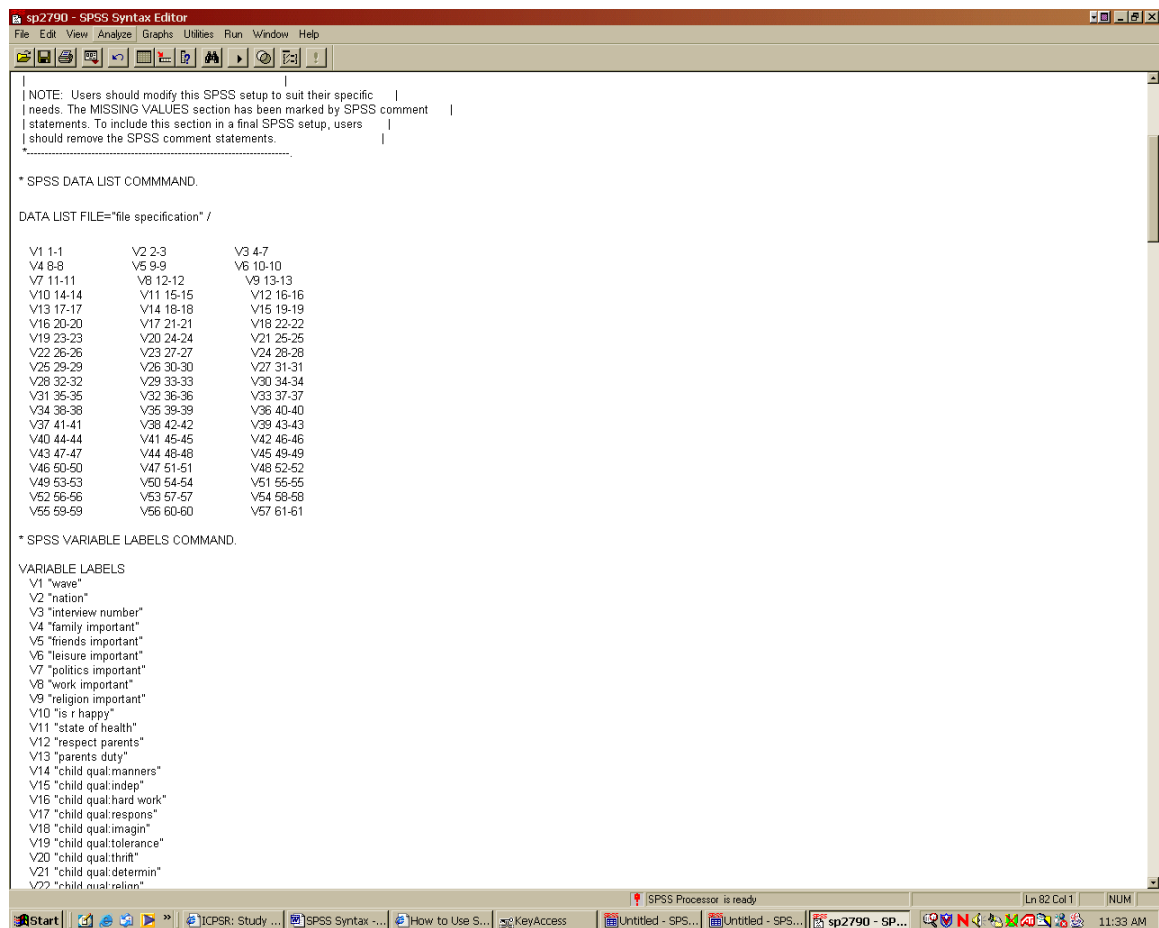
Often this is all you need to do. Once you go into the preexisting SPSS syntax code and identify where the dataset is, you are ready to go. After this simply use the run command or highlight the syntax and click the small arrow on the tool bar.

Important! Make sure you end this command with a period.

2. Datalist

This command tells SPSS how to organize the data into separate variables, which can be used and manipulated later.

The syntax looks something like this and usually immediately follows the file location command discussed above.



```
sp2790 - SPSS Syntax Editor
File Edit View Analyze Graphs Utilities Run Window Help
NOTE: Users should modify this SPSS setup to suit their specific
needs. The MISSING VALUES section has been marked by SPSS comment
statements. To include this section in a final SPSS setup, users
should remove the SPSS comment statements.
* SPSS DATA LIST COMMAND.
DATA LIST FILE="file specification" /
V1 1-1      V2 2-3      V3 4-7
V4 8-8      V5 9-9      V6 10-10
V7 11-11   V8 12-12   V9 13-13
V10 14-14  V11 15-15  V12 16-16
V13 17-17  V14 18-18  V15 19-19
V16 20-20  V17 21-21  V18 22-22
V19 23-23  V20 24-24  V21 25-25
V22 26-26  V23 27-27  V24 28-28
V25 29-29  V26 30-30  V27 31-31
V28 32-32  V29 33-33  V30 34-34
V31 35-35  V32 36-36  V33 37-37
V34 38-38  V35 39-39  V36 40-40
V37 41-41  V38 42-42  V39 43-43
V40 44-44  V41 45-45  V42 46-46
V43 47-47  V44 48-48  V45 49-49
V46 50-50  V47 51-51  V48 52-52
V49 53-53  V50 54-54  V51 55-55
V52 56-56  V53 57-57  V54 58-58
V55 59-59  V56 60-60  V57 61-61
* SPSS VARIABLE LABELS COMMAND.
VARIABLE LABELS
V1 "wave"
V2 "nation"
V3 "interview number"
V4 "family important"
V5 "friends important"
V6 "leisure important"
V7 "politics important"
V8 "work important"
V9 "religion important"
V10 "is r happy"
V11 "state of health"
V12 "respect parents"
V13 "parents duty"
V14 "child qual:manners"
V15 "child qual:indep"
V16 "child qual:hard work"
V17 "child qual:respons"
V18 "child qual:imagin"
V19 "child qual:tolerance"
V20 "child qual:thrift"
V21 "child qual:determin"
V22 "child qual:religion"
```

Note the list of commands in three columns that begin with V and are followed by some numbers. Here we are basically instructing SPSS to create variables using our dataset .A command VX M-N(Z) creates a variable named VX out of columns M through N and would have Z numbers after the decimal place. Note if Z=A, then the variable constructed is a string variable.

Some examples:

The command "V1 1-1" instructs SPSS to construct a variable V1 out of the first column alone

The command "V2 2-3" instructs SPSS to construct a variable V2 out of the second and the third columns in the dataset.

If we had a command that reads V2 3-19(A) instead then we would be instructing SPSS to construct a string variable named V2 out of the 3rd through 19th columns in the dataset.

3. Variable Labels

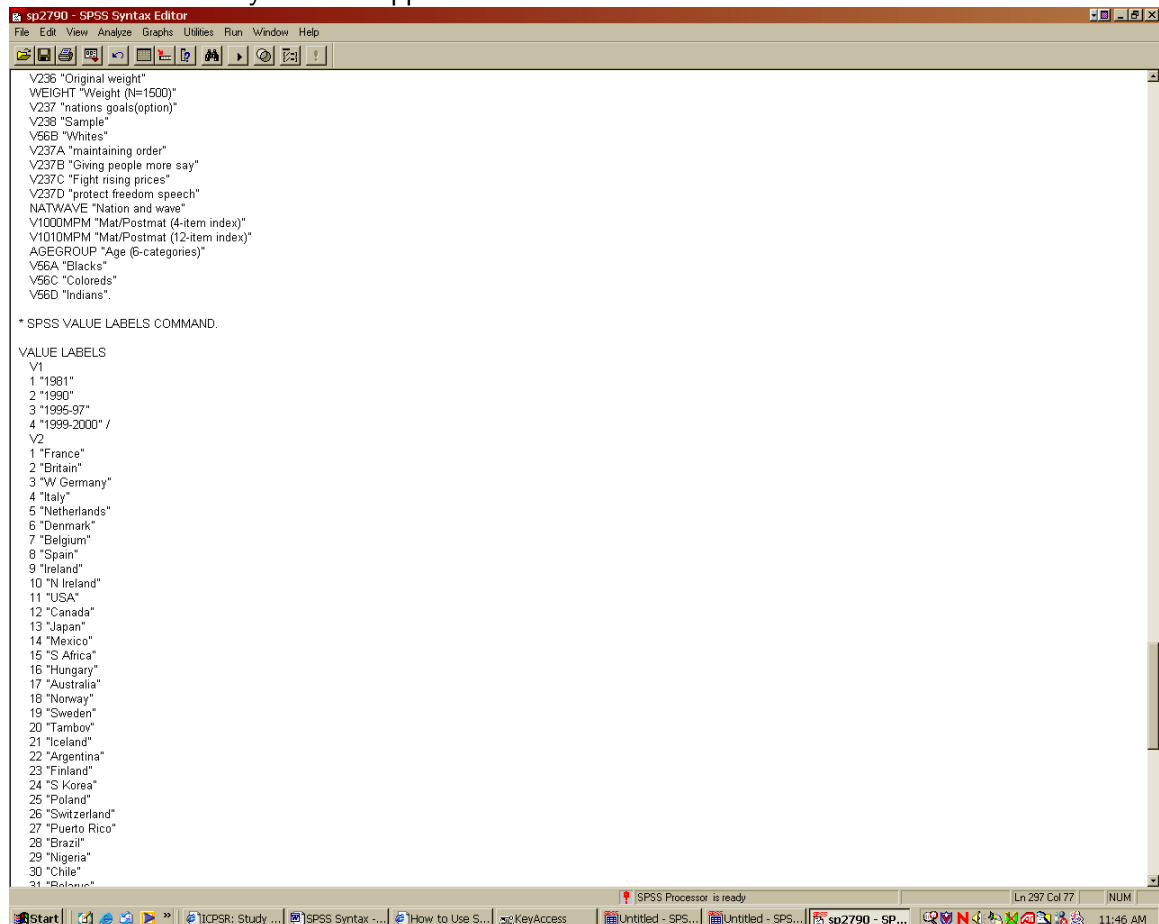
In the figure above, just below the variable definitions you will see a number of commands that are in effect used for assigning labels to the different variables.

So in this example the command of V1 "wave" gives the first variable a label of "wave" A command V11 "State of Health" gives variable V11 a label of "State of Health".

4. Value Labels

This command assigns labels to the different values within each variable. This can be very useful. Suppose you are working with international survey data like in this example and you have assigned each country a number. Assigning a label to each number will make it easier to remember the state that is represented by each number. Furthermore the label will appear when running crosstabs and other analyses of the data. This will make the output easier for the user to understand.

This is how the syntax will appear -



```
sp2790 - SPSS Syntax Editor
File Edit View Analyze Graphs Utilities Run Window Help
V236 "Original weight"
WEIGHT "Weight (N=1500)"
V237 "nations goals(option)"
V238 "Sample"
V56B "Whites"
V237A "maintaining order"
V237B "Giving people more say"
V237C "Fight rising prices"
V237D "protect freedom speech"
NATWAVE "Nation and wave"
V1000MPM "Mat/Postmat (4-item index)"
V1010MPM "Mat/Postmat (12-item index)"
AGEGROUP "Age (6-categories)"
V56A "Blacks"
V56C "Coloreds"
V56D "Indians".

* SPSS VALUE LABELS COMMAND.

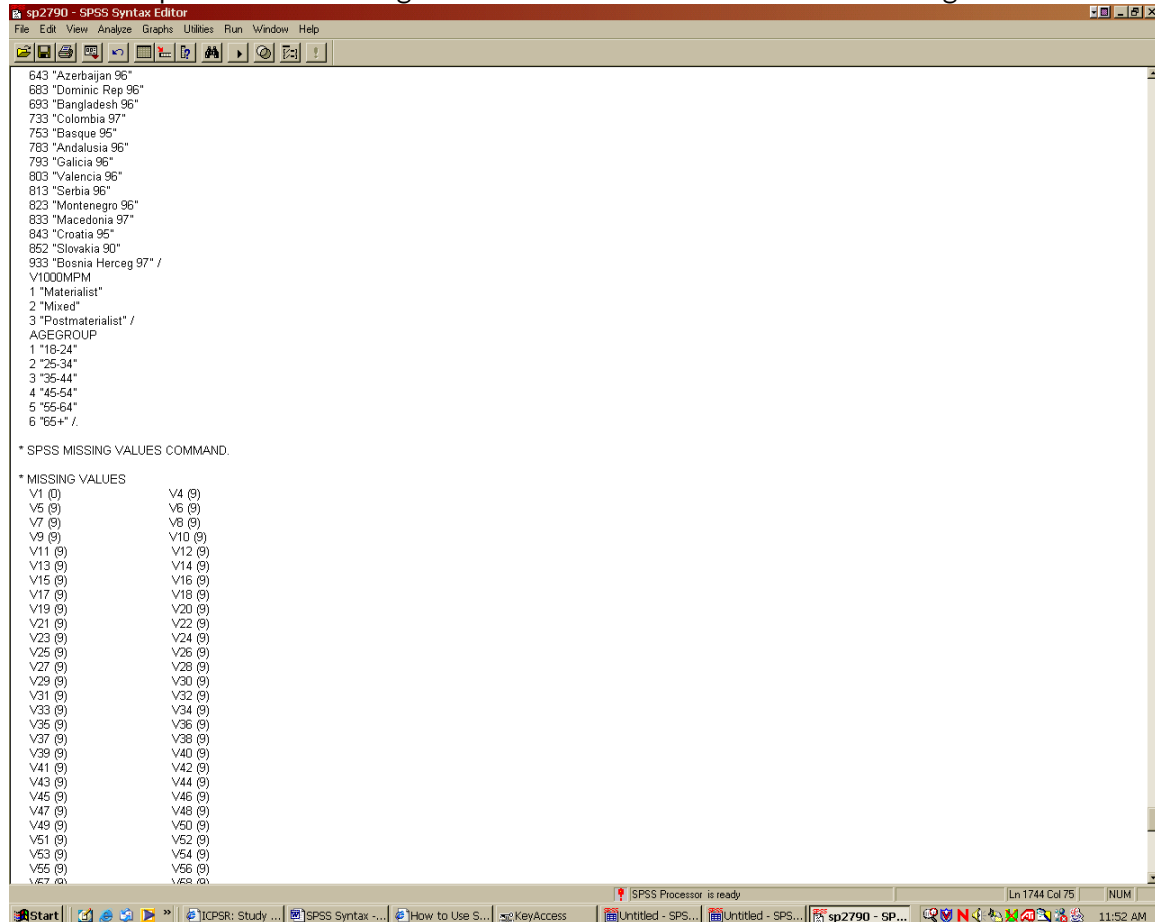
VALUE LABELS
V1
1 "1981"
2 "1990"
3 "1995-97"
4 "1999-2000" /
V2
1 "France"
2 "Britain"
3 "W Germany"
4 "Italy"
5 "Netherlands"
6 "Denmark"
7 "Belgium"
8 "Spain"
9 "Ireland"
10 "N Ireland"
11 "USA"
12 "Canada"
13 "Japan"
14 "Mexico"
15 "S Africa"
16 "Hungary"
17 "Australia"
18 "Norway"
19 "Sweden"
20 "Tambov"
21 "Iceland"
22 "Argentina"
23 "Finland"
24 "S Korea"
25 "Poland"
26 "Switzerland"
27 "Puerto Rico"
28 "Brazil"
29 "Nigeria"
30 "Chile"
31 "Bolivia"
```

This is rather self-explanatory. For instance, a value label of V1 1 "1981" is telling SPSS to label all data scored as 1 within variable 1 "1981". Similarly, a value label of V2 1 "France" instructs SPSS to label all data scored as 1 within variable 2 as "France".

5. Missing Values

The missing values command will list certain values within each variable as missing. You can see this when the data is put in "variable view" format. This command will not actually change the values to "missing" but simply inform SPSS that the entry is missing in the dataset.

Our example has the following entries in the data definition file for missing values.



```
643 "Azerbaijan 96"  
683 "Dominic Rep 96"  
693 "Bangladesh 96"  
733 "Colombia 97"  
753 "Basque 96"  
763 "Andalusia 96"  
793 "Galicia 96"  
803 "Valencia 96"  
813 "Serbia 96"  
823 "Montenegro 96"  
833 "Macedonia 97"  
843 "Croatia 95"  
852 "Slovakia 90"  
933 "Bosnia Herceg 97" /  
V1000MPM  
1 "Materialist"  
2 "Mixed"  
3 "Postmaterialist" /  
AGEGROUP  
1 "18-24"  
2 "25-34"  
3 "35-44"  
4 "45-54"  
5 "55-64"  
6 "65+" /  
  
* SPSS MISSING VALUES COMMAND.  
  
* MISSING VALUES  
V1 (0)          V4 (9)  
V5 (9)          V6 (9)  
V7 (9)          V8 (9)  
V9 (9)          V10 (9)  
V11 (9)         V12 (9)  
V13 (9)         V14 (9)  
V15 (9)         V16 (9)  
V17 (9)         V18 (9)  
V19 (9)         V20 (9)  
V21 (9)         V22 (9)  
V23 (9)         V24 (9)  
V25 (9)         V26 (9)  
V27 (9)         V28 (9)  
V29 (9)         V30 (9)  
V31 (9)         V32 (9)  
V33 (9)         V34 (9)  
V35 (9)         V36 (9)  
V37 (9)         V38 (9)  
V39 (9)         V40 (9)  
V41 (9)         V42 (9)  
V43 (9)         V44 (9)  
V45 (9)         V46 (9)  
V47 (9)         V48 (9)  
V49 (9)         V50 (9)  
V51 (9)         V52 (9)  
V53 (9)         V54 (9)  
V55 (9)         V56 (9)  
V57 (9)         V58 (9)
```

For instance, here we have instructed SPSS to note that any entry of 0 for variable 1 should be noted as missing. Similarly an entry of 9 is an indicator for missing information for variable 5.