Note to SPSS Users

Data were developed in SAS. Using software for data transfer, these SAS data were converted to SPSS. The conversion works well but there are some differences between SAS and SPSS that cannot be completely addressed by the conversion routine. Missing values and associated formats are two areas that require special attention. Please read the questions and answers below for issues related to these differences between SAS and SPSS.

To avoid addressing concerns with missing values and associated formats, simply use the SAS files in SPSS; SPSS can open SAS files directly. All distinct missing values will appear as a single system missing value and associated value labels in SAS will not be associated in SPSS.

As with any data, it is important to do some preparation before using these data. Read the codebook for details about each variable, there are many sources of data and similar types of items are asked of many respondents. With the interview data in particular, there are complex skip patterns that are defined in the codebook. It is unadvisable to rely upon variable labels and value labels for choosing which items to use in analysis. It is highly recommended that users run a frequency distribution on the selected variables and check for values, distribution of values, and n's.

SPSS User Frequently Asked Questions

Q: What special missing values have been defined in SPSS?
A: SAS allows up to 27 discrete missing values, SPSS only allows 3 discrete missing values or a range of missing values. Data are developed in SAS and exported to SPSS. In order to retain discrete values, we convert the SAS user defined missing values to negative numbers. These negative numbers need to be defined as missing to exclude these values from analyses.

Q: How do I define special missing values in SPSS?
A: By default, missing values are included unless they are defined as missing values. To define missing values, click on the missing column for the variable. Specify the range as -999 [low] to -980 [high].

Q: How can I get missing values automatically?
A: If you open up the SAS file from within SPSS all missing values will automatically default to special missing values.

Q: What are the missing values in SPSS?
A: SAS allows for more discrete user-defined missing values than SPSS allows for. In order to retain the values, SAS missing values are converted to negative numbers. By defining a range of negative numbers as missing, reasons for missing data are retained but values are not included in analysis.

Missing Values		
SAS	**SPSS**	**Missing Value**
.a	-999	Not applicable
.b	-988	Not asked this section
.c	-987	Incomplete
.d	-998	Do not know
.g	-986	Sparse data
.i	-993	In secondary school: returned to HS
.m	-997	Mail
.n	-985	Not graded
.p	-996	Partial
.r	-995	Refused
.s	-994	Skipped
.u	-991	Ungraded
.v	-990	Version
.w	-984	Withdrawn
.x	-992	Not in NCES data
.z	-989	Not ascertained

Q: In variable view, there are values listed in the values column. Why are all the values listed in a variable not in the data?
A: Data are developed in SAS which has a different file structure. In SAS, a variable is associated with a user defined format which is essentially a look-up table. Many variables may be associated with a single format, so that format has to have all possible values of those variables. However, a variable may not have all the values listed in the associated format. When data are converted to SPSS, the procedure looks at the associated format table and lists all possible values in the format table as opposed to listing just the values present in the variable.

For example, a yes/no variable might list "0" and "2" for "no" in SPSS because the user defined format in SAS allows for either possibility. In SPSS variable view, it may look like "no" is both "0" and "2" but in fact the variable itself has only "0" for the no response. We frequently recode "2" to "0" but, by allowing for both possibilities, it makes it unnecessary to have to change the format association as well as changing the value itself. This has no effect on analysis. When a frequency is run, only the values that are in the variable will be displayed.

Q: How do I change the values listed in SPSS?
A: In variable view, click on the values column. Values can be added, removed, or modified.

Q: How do I collapse categories or categorize continuous variables in SPSS?

A: In SAS, applying a user defined format with a range or discreet values with the same value label will automatically collapse values for purposes of doing analyses. A continuous variable, such as number of months youth has attended his or her school, can be analyzed as a categorical variable by specifying a range of values in the user defined format and doing a frequency distribution with the associated format. Categories can be collapsed in the same way.

SPSS requires that collapsed values be recoded into a new variable. When value labels are converted from SAS to SPSS, ranges are ignored and no value statements are imported. If discrete values are assigned in the user defined format rather than ranges, i.e. "7,8,9" rather than "7-9," the label will be imported for each category within that range but the SPSS frequency distribution will list each category separately.

The example below shows what happens in SPSS and SAS using a range format and a discreet format to collapse categories. Variables X and Y have the same values but have different user defined formats: one with a range assignment and one with a discrete value assignment. A frequency distribution was run on these variables in SAS and SPSS:

User Defined Value Formats with Range of Values Defined

SAS

x	Frequency	Percent	Cumulative Frequency	Cumulative Percent
(1) Mostly As	1	10.00	1	10.00
(2) Mostly As and Bs	1	10.00	2	20.00
(3) Mostly Bs	1	10.00	3	30.00
(4) Mostly Bs and Cs	1	10.00	4	40.00
(5) Mostly Cs	1	10.00	5	50.00
(6) Mostly Cs and Ds	1	10.00	6	60.00
(7-9) Mostly Ds and Fs	4	40.00	10	100.00

SPSS

x

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	(1) Mostly As	1	10.0	10.0	10.0
	(2) Mostly As and Bs	1	10.0	10.0	20.0
	(3) Mostly Bs	1	10.0	10.0	30.0
	(4) Mostly Bs and Cs	1	10.0	10.0	40.0
	(5) Mostly Cs	1	10.0	10.0	50.0
	(6) Mostly Cs and Ds	1	10.0	10.0	60.0
	7.00	1	10.0	10.0	70.0
	8.00	1	10.0	10.0	80.0
	9.00	2	20.0	20.0	100.0
	Total	10	100.0	100.0	

User Defined Value Formats with Discreet Values Defined				
SAS				

```
                                             Cumulative   Cumulative
   y                     Frequency   Percent  Frequency    Percent
(1) Mostly As                1       10.00        1         10.00
(2) Mostly As and Bs         1       10.00        2         20.00
(3) Mostly Bs                1       10.00        3         30.00
(4) Mostly Bs and Cs         1       10.00        4         40.00
(5) Mostly Cs                1       10.00        5         50.00
(6) Mostly Cs and Ds         1       10.00        6         60.00
(7-9) Mostly Ds and Fs       4       40.00       10        100.00
```

SPSS

y

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	(1) Mostly As	1	10.0	10.0	10.0
	(2) Mostly As and Bs	1	10.0	10.0	20.0
	(3) Mostly Bs	1	10.0	10.0	30.0
	(4) Mostly Bs and Cs	1	10.0	10.0	40.0
	(5) Mostly Cs	1	10.0	10.0	50.0
	(6) Mostly Cs and Ds	1	10.0	10.0	60.0
	(7-9) Mostly Ds and Fs	1	10.0	10.0	70.0
	(7-9) Mostly Ds and Fs	1	10.0	10.0	80.0
	(7-9) Mostly Ds and Fs	2	20.0	20.0	100.0
	Total	10	100.0	100.0	

Although the collapsed label prints for the range when discreet values are assigned, the values themselves are not collapsed for analyses as they are in SAS.

Ordinal variable categories with too few to analyze in a given category are typically collapsed with adjoining values. For example, if few report being "Very dissatisfied," "Very dissatisfied" might be collapsed with "Somewhat dissatisfied" and reported as "Somewhat/very dissatisfied". For SPSS users, recoding into a new variable is necessary to collapse values or categories.

It is not practical to use discreet values for large ranges, such as hours worked or hourly pay rate. Other than the lower boundary of collapsed ranges, user defined formats are in ranges and hence will not be carried over to SPSS. Generally, continuous variables are collapsed into quartiles for analysis. SPSS users will need to recode and collapse continuous values into new variables for categorical analyses, such as frequency distributions.

Examples of how variables are collapsed can be found in the NLTS2 tables published on http://nlts2.org/. Example SPSS programming code for collapsing variables is included in the document SPSS QuickCode.pdf in Appendix C. This SPSS programming code references the SAS format ranges to create variables with the same collapsed ranges SAS users have available by default.

Q: How do I calculate weighted standard errors in SPSS?

A: SPSS will calculate weighted standard errors in procedures that allow weights to be specified, however it is important to note that SPSS procedures assume a simple random sample. The NLTS2 database is not a simple random sample, it is the result of a multi-stage sampling process. While these procedures do calculate the correct estimates for weighted means and percentages, the estimates for weighted standard errors are not calculated correctly for these data. Do not report standard errors from these procedures.

To produce the correct standard error estimates for NLTS2 data it is necessary to use programs or procedures that calculate standard errors for sampled data. SPSS has a "Complex Samples" module that calculates the estimated weighted standard errors using Taylor Linearization. Complex Samples is not included with base release of SPSS and must be purchased separately. The Complex Samples module is just one of many options, such as SUDAAN, Stata, or SAS Survey procedures. Any of the basic procedures will produce the correct estimates for weighted means and percentages. However, it is important to note that procedures or algorithms that calculate weighted standard errors must account for the sampling design.

For additional information relevant to both SPSS and SAS users, please refer to the NLTS2 database documentation.