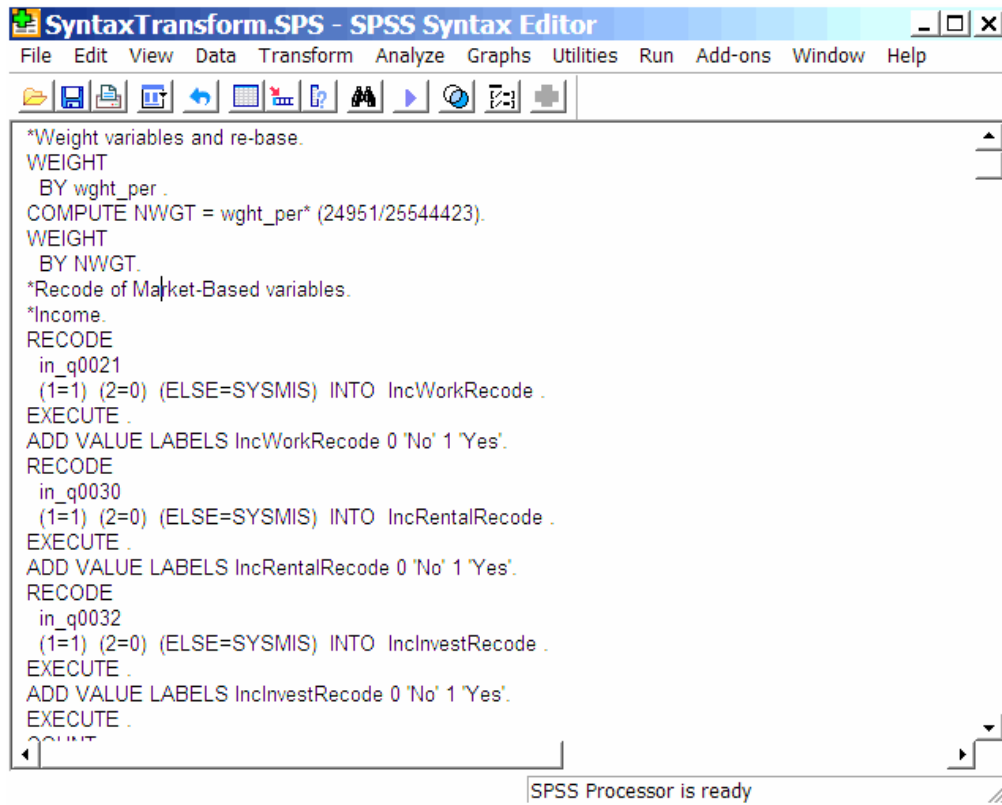


SPSS Presentation Soci-410
Vicki Beaupré-Odorico
Feb 2010

1) Syntax files

A syntax file is an efficient way to store all of your procedures, analyses, commands, and comments. Syntax files are saved as .SPS files. Below is an example of a syntax file with commands:



```
SyntaxTransform.SPS - SPSS Syntax Editor
File Edit View Data Transform Analyze Graphs Utilities Run Add-ons Window Help

*Weight variables and re-base.
WEIGHT
  BY wght_per .
COMPUTE NWGT = wght_per* (24951/25544423).
WEIGHT
  BY NWGT.
*Recode of Market-Based variables.
*Income.
RECODE
  in_q0021
  (1=1) (2=0) (ELSE=SYSMIS) INTO IncWorkRecode .
EXECUTE .
ADD VALUE LABELS IncWorkRecode 0 'No' 1 'Yes'.
RECODE
  in_q0030
  (1=1) (2=0) (ELSE=SYSMIS) INTO IncRentalRecode .
EXECUTE .
ADD VALUE LABELS IncRentalRecode 0 'No' 1 'Yes'.
RECODE
  in_q0032
  (1=1) (2=0) (ELSE=SYSMIS) INTO IncInvestRecode .
EXECUTE .
ADD VALUE LABELS IncInvestRecode 0 'No' 1 'Yes'.
EXECUTE .
COUNT
```

SPSS Processor is ready

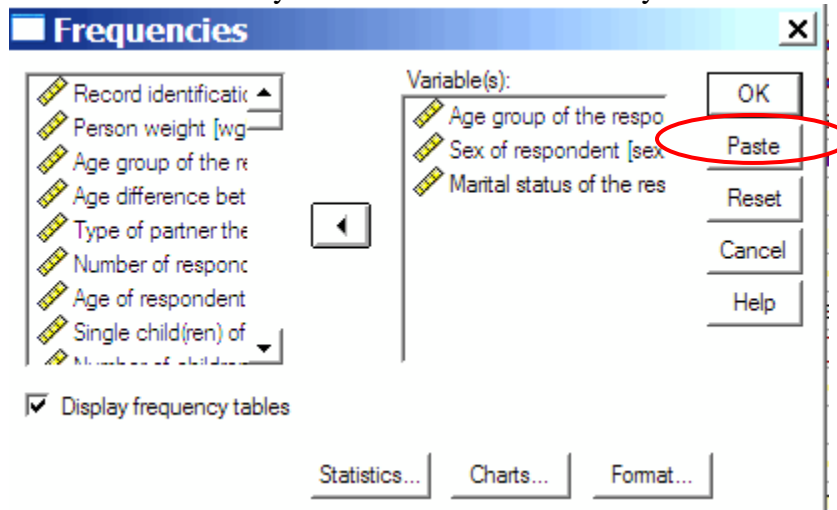
To open a new syntax file, in the data editor (SPSS data file), click on File -> New -> Syntax. To open an existing syntax file, in the data editor, click on File -> Open -> Syntax.

When you select any command in the data editor you will be given the option to paste the command. This will paste the syntax command directly in a syntax file.

If you do not have a working syntax file, SPSS will open a new one for you. If you are working with an open syntax file, SPSS will paste the command at the bottom of the syntax file. Below is an example of a “frequencies” dialog box. Before you can paste a command, you need to select the operation.

To run a frequency distribution, select which variables you are investigating, and then click on the paste box (this will paste the command in your syntax file). You can paste

any command in a syntax file. When you become proficient with syntax commands, you will be able to write your own commands in the syntax file directly.



***Note:** it is useful to work with two types of syntax files:

“V” files (e.g. SyntaxV.SPS) should be used to store all commands which transform your variables or your data file (e.g. re-codes, compute function, aggregate command, merge files...)

“P” files (e.g. SyntaxP.SPS) should be used to store any statistical procedures (e.g. descriptive statistics, correlations, regression...)

Working with a syntax file allows you to maintain the original data file, and to keep it small. Every time you open a data file, open the “V” syntax file, where all the variable and file transformations are saved and run all the commands by selecting Run -> All. Then you can open the “P” file and run your procedures.

***Useful Tip 1:** Comments can be integrated directly into your syntax file. You must place an asterisks “*” at the beginning and a period “.” at the end of the comment. I suggest you start each new section by briefly describing what analysis or variable transformations you will be performing and include the date. This will help you keep track of why you conducted transformations or analyses and it will give you appropriate keyword search terms.

***Useful Tip 2:** Always increment your syntax files. Every time you work on your syntax file, save an updated copy as Syntax01.SPS, Syntax02.SPS, Syntax03.SPS, etc... When you modify any command, the new syntax file will reflect these changes, but your old syntax file will save the original commands.

2) Frequency Distribution

Before you begin, set your output and value labels so that they contain both the name, value and label, go to Edit - > Options - > Output Labels - > and select Names and Labels and Values and Labels. This will enable you to see both the variable name and label in the output (i.e. marstat and Marital Status of the Respondent) and the value and value label (i.e. 1 “Married”, 2 “Common-Law”, 3 “Divorced”...).

To run a frequency distribution click on Analyze - > Descriptive Statistics - > Frequencies. Select a variable and click on the triangle (it should now appear on the right hand side), repeat for each variable. Click on paste.

C) Run a frequency distribution for marstat (Marital Status).

Table 1: Frequency Distribution of Respondent Marital Status

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Married	5691	45.9	46.0	46.0
	2 Living common-law	1068	8.6	8.6	54.6
	3 Widowed	1088	8.8	8.8	63.4
	4 Divorced	492	4.0	4.0	67.4
	5 Separated	866	7.0	7.0	74.4
	6 Single (Never married)	3166	25.5	25.6	100.0
	Total	12371	99.8	100.0	
Missing	8 Not stated	16	.1		
	9 Don't know	10	.1		
	Total	26	.2		
Total		12397	100.0		

What does this tell us? The total sample size is 12397, only 12371 (or 99.8%) of the respondent’s gave a valid response so we are left with a total of 26 missing cases (or .2%). 46.0% of respondent’s are married, 8.6% are living common-law...

When you look at a frequency distribution with missing cases, always interpret the valid percent since it divides the number of cases in each category into the total valid cases (i.e. married category $5691/12371 = 46.0\%$).

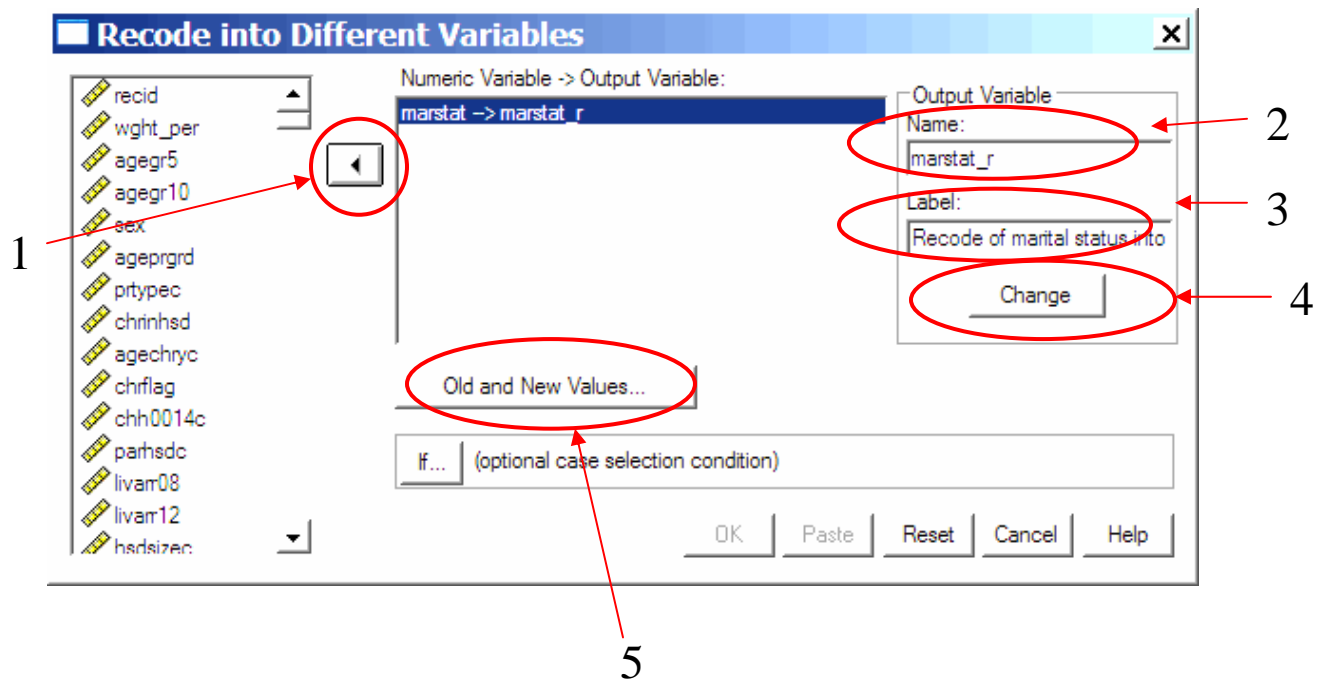
3A) Recode

Recoding allows you to create a new variable using an existent variable. To recode the values of a variable into a different variable click on Transform - > Recode - > Into Different Variables. This dialog box will appear:

***Note:** if you choose to recode Into Same Variable, it will get rid of the original values and your new variable will replace the old variable.

***Note:** When you recode a variable, always use the old variable name for the new variable but include an r or _r to indicate that it is a recode.

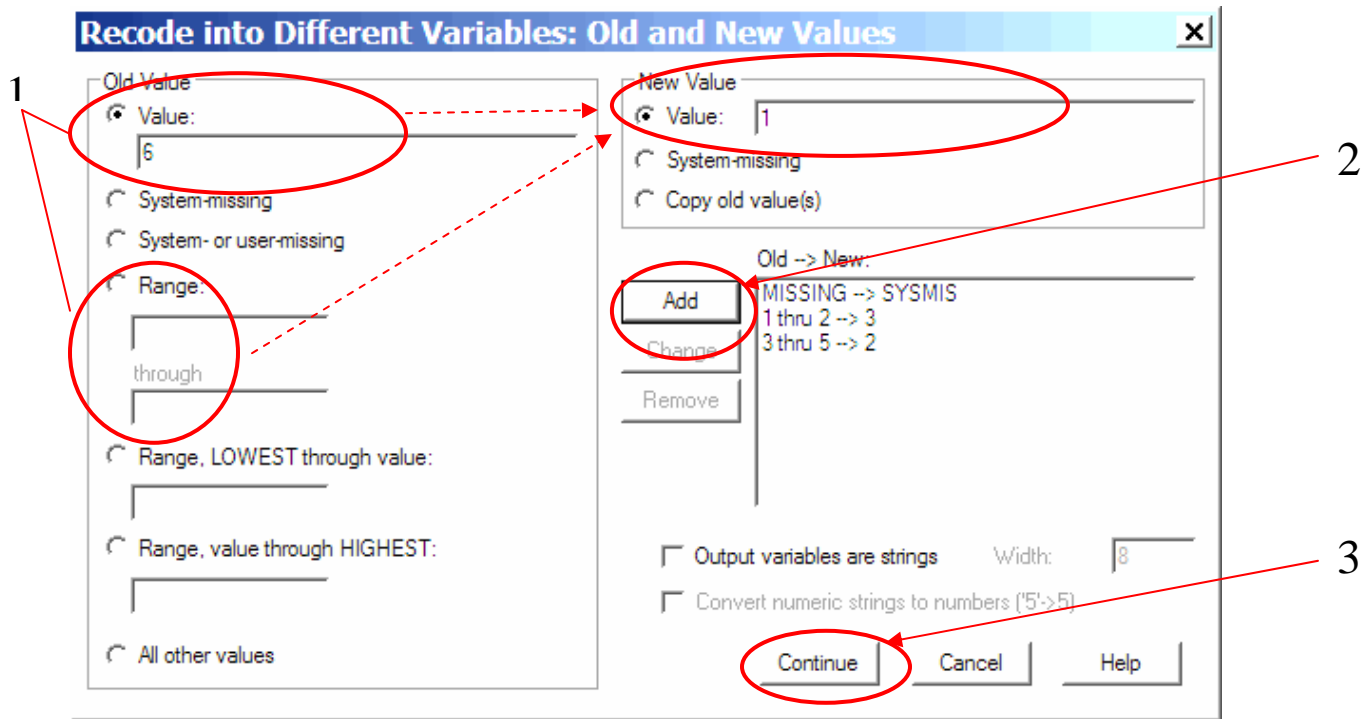
- 1) Select the variable to be re-coded and click on the triangle.
- 2) Click on the name box and give the variable a new name.
- 3) Click on the label box and provide a brief description of the new variable.
- 4) Click on change
- 5) Click on Old and New Values



When you click on Old and New Values (step 5) another dialog box will appear, please see below. Every original value has to be transformed into a new value.

***Useful Tip:** Make sure to recode your variables in the same direction (i.e. from low to high).

- 1) Enter the old values and the new values in the Old Value box and New Value box. If you are changing only one value, select the value box, if you wish to refer to a range of values, you must click on the small circle beside the range button and enter the range of values.
- 2) For each change you have entered, click on the add button.
- 3) When you have finished entering all your recodes, click on the continue button.
- 4) You will then return to the original screen (above). Click on paste. This will paste your recode commands into your syntax file. Run the command from your syntax file.



****Note:** You will have to enter the values and value labels directly into your syntax command. After you have pasted the recode command put this command below:

Add value labels marstat_r 1 “Single” 2 “Divorced/Widowed/Seperated” 3 “Couple”.

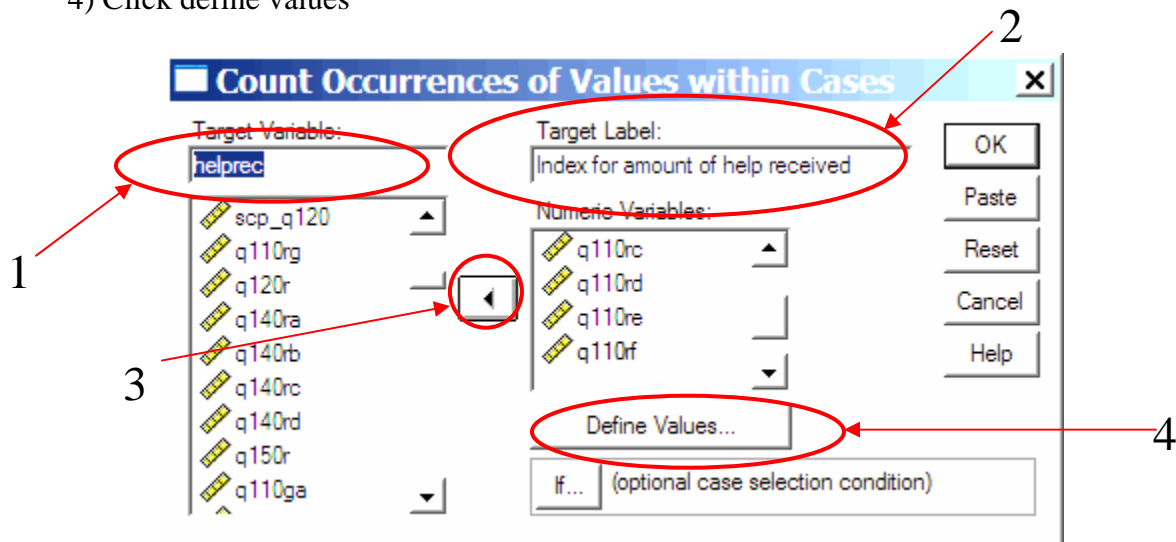
***Useful Tip:** After you have recoded your new variable, run a crosstabulation with the old and new variable to make sure your transformations are correct. Your N and missing values should be the same for the original variable and for the new recoded variable. Your measure of association (i.e. gamma, phi) should be “1”.

3B) Using the Count Function

The count function allows you to count the number of variables that have a particular value. For example, if you have six variables that ask questions about group participation (sports league, hobby group, religious affiliation, school group...), and you want to construct a scale using these six variables, you would use the count function. You need to specify which value you want to count. If the variables are coded as (1=yes, 2=no) and you want to find the number of occurrences the respondent stated yes "1" for each group, you have to specify that you want the value "1" counted. To use the count function click on Transform - > Count.

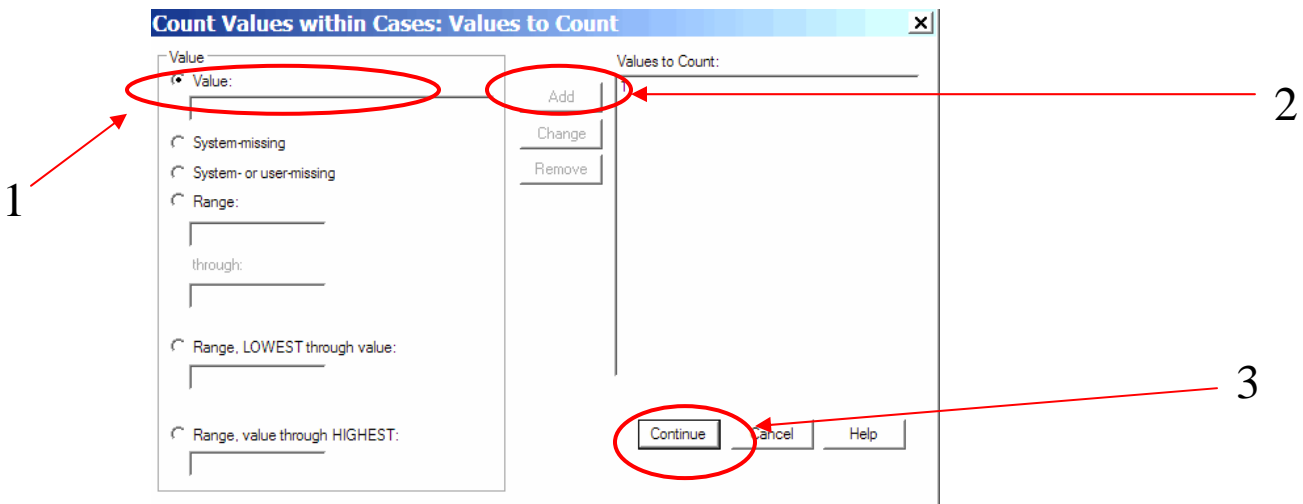
Let's compute an index for amount of help received (count "1" for these six items: q110ra, q110rb, q110rc, q110rd, q110re, q110rf):

- 1) Type in the name of the variable you want to create in target variable.
- 2) Type in a label name
- 3) Select each variable you want to be counted and click on the triangle.
- 4) Click define values



When you click on define values, a new dialog box will open. This is where you identify the values, or the range of values, you want counted.

- 1) Type the value you want counted in the value box or input the range in the range box.
- 2) Click on add.
- 3) Click on continue when you are done. This will return you to the first dialog box.



Once you have defined the values, click on paste. This will past your count command in the syntax file. You can then run the command from the syntax file.

helprec Index of amount of help received

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	4775	38.5	38.5	38.5
	1.00	2555	20.6	20.6	59.1
	2.00	2180	17.6	17.6	76.7
	3.00	1553	12.5	12.5	89.2
	4.00	948	7.6	7.6	96.9
	5.00	340	2.7	2.7	99.6
	6.00	46	.4	.4	100.0
Total		12397	100.0	100.0	

What does this tell us? Remember that a count function, counts the number of times a respondent said yes to the questions in your index. Look at the example below, each respondent gave a yes (1) or no (2) response to the following questions:

	q110ra	q110rb	q110rc	q110rd	q110re	q110rf	Index Count
Respondent 1	1	1	2	1	1	2	4
Respondent 2	2	1	1	2	1	2	3
Respondent 3	2	2	2	2	2	2	0

Once you count each 1 across all six items, you can see that respondent 1 received help in 4 categories, respondent 2 received help in 3 categories and respondent 3 received no help. This is a rank; some people receive more help than others. Use this same principle

when you interpret the frequency distribution. 38.5% of respondents received no help, compared with .4% of respondents who received help in every category.

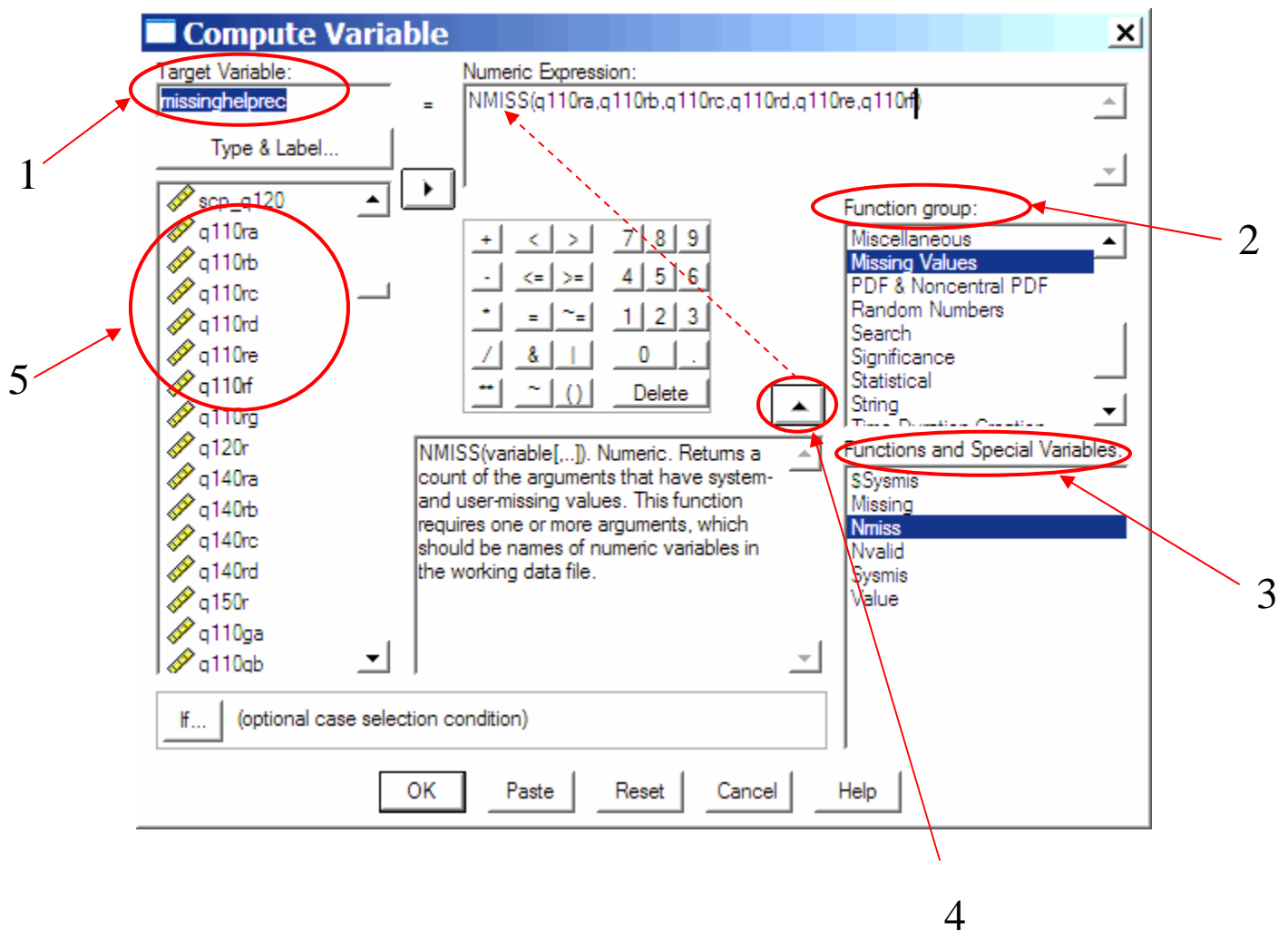
***Note:** you should run a reliability analysis to determine if the index is reliable. Go to Analyze -> Scale -> Reliability Analysis, and select all the items in your scale. A reliable scale will have a high Cronbach's Alpha; a value above .60 is usually a good indicator of a good scale. If you get a low alpha score, try removing one of the items to see if it improves. Remember that you want your items to be highly correlated with each other (which will produce a high alpha).

***Note:** Be careful when using the count function. SPSS will assign a 0 value for missing cases. So if a respondent did not answer any of the above questions, they will nonetheless receive a score of 0. But, this is not equal to answering no to all six questions. If you run a frequency distribution for the six items above, you will notice that there are only 90 missing values for each question. Use the count function with caution, if you have too many missing cases, do not use the count function unless you can make a claim that a count of 0 means the same for a respondent who answered no to all the questions as a respondent who did not answer any of the questions

***Note:** To get rid of missing cases you must filter out those individuals who gave a missing value.

Step 1: Compute a variable that will count the number of missing cases across each of the variables for each respondent. To compute a new variable go to Transform -> Compute

- 1) Write your variable name
- 2) In the "Function group", select Missing Values
- 3) In the "Functions and Special Variables" select Nmiss
- 4) Click on the triangle to bring the command up to the Numeric Expression box
- 5) On the left hand side, select all the variables you want to include in your index. Using this command, SPSS will count the number of missing cases across all six variables for each respondent.

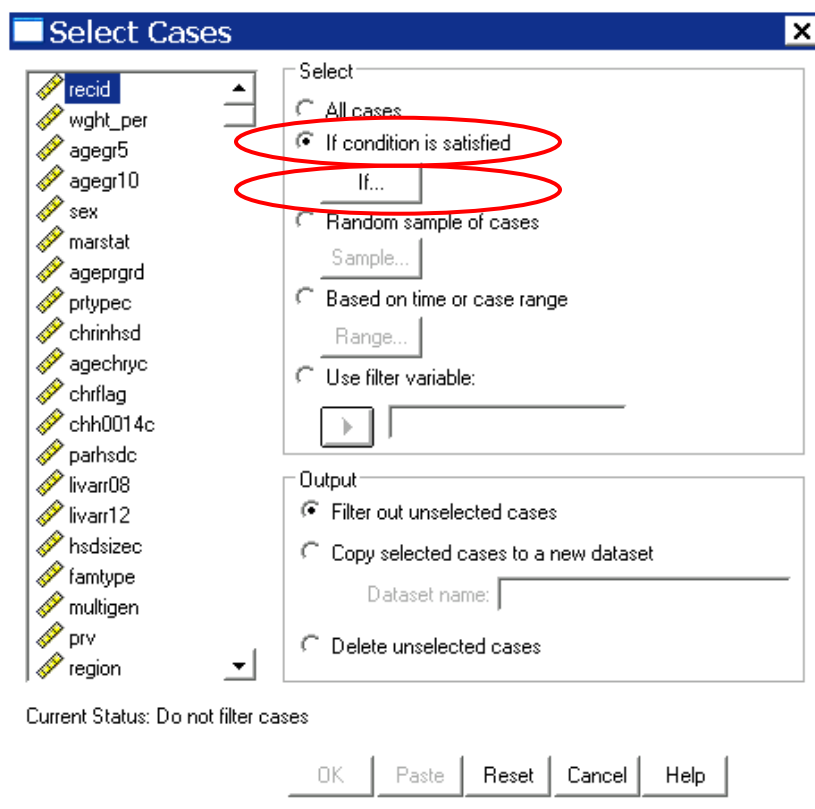


**Take a look at the frequency distribution below for the variable you just created called missinghelprec (remember you computed missing values). What this frequency distribution tells us is that 12307 respondents have 0 missing values across all six variables. And, 90 respondents have 6 missing values across the six variables.

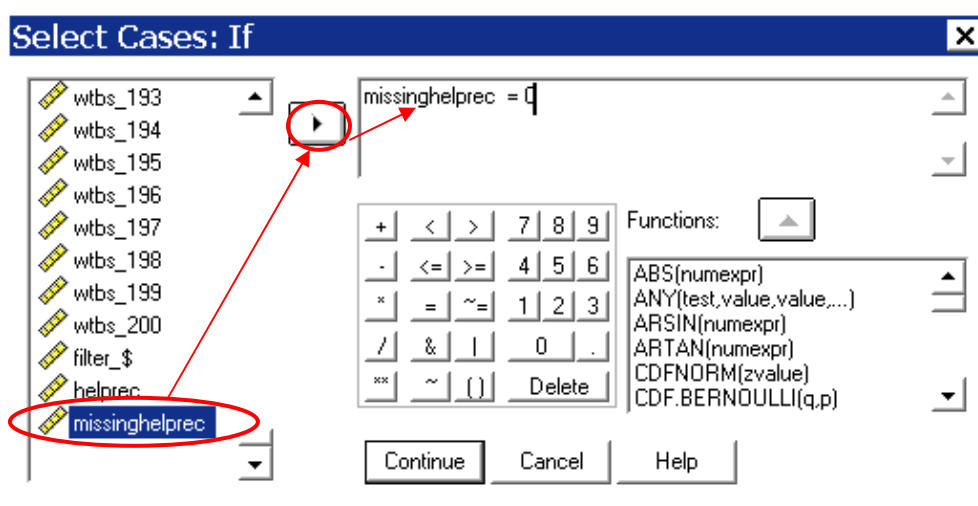
missinghelprec

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid .00	12307	99.3	99.3	99.3
6.00	90	.7	.7	100.0
Total	12397	100.0	100.0	

Step 2: Once you've computed a variable which identifies those respondents that have missing cases, you now have to filter out those respondents. To use a filter command, go to Data -> Select Cases -> then select If condition is satisfied -> then click on If...



A new dialogue box will open. Select the variable you computed “missinghelprec”. Click on the triangle to bring it into the dialogue box. Now you need to specify what it is you want to do. In our case, we want to keep all the variables that have valid cases. You will write: missinghelprec=0 (we are asking SPSS to select all the respondents who have a value of 0 for the variable missinghelprec. A value of 0 indicates that there are no missing values for all six variables in the index).



Note: If you re-run a frequency distribution for the variable helprec, you will notice that the 90 respondents with missing values have now been excluded. Your total is now 12307 instead of 12397.

helprec Index of amount of help received

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	4685	38.1	38.1	38.1
	1.00	2555	20.8	20.8	58.8
	2.00	2180	17.7	17.7	76.5
	3.00	1553	12.6	12.6	89.2
	4.00	948	7.7	7.7	96.9
	5.00	340	2.8	2.8	99.6
	6.00	46	.4	.4	100.0
	Total	12307	100.0	100.0	

3C) Compute

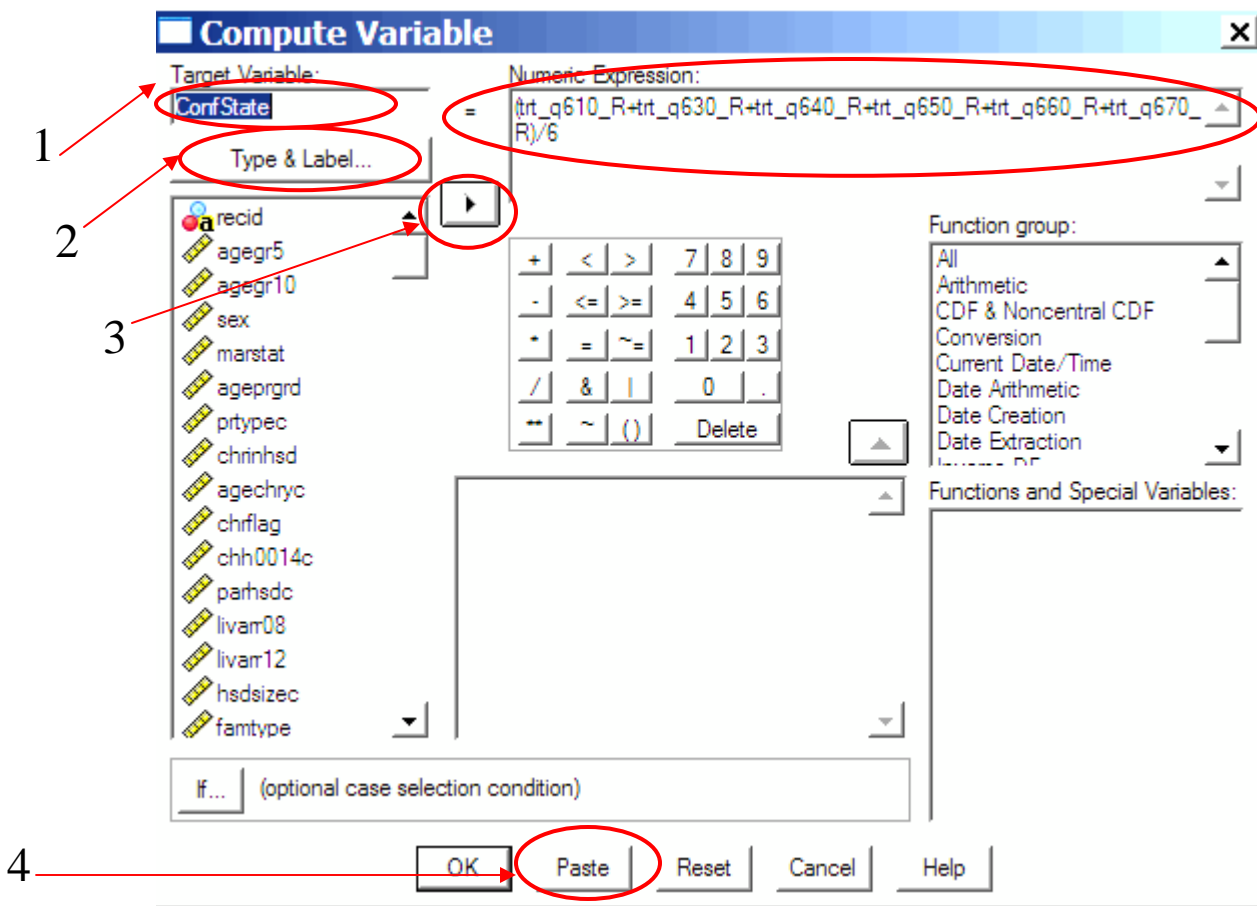
You can use the compute function to transform the values of one or more variables. For example you can compute the mean of two or more variables (if you know the total number of years the respondent attended school, and the total number of years the respondent's spouse attended school, you can calculate the average (mean) number of years of schooling between the two). You can also transform variables by taking the log or calculate the z-score. You can calculate age, if you have the date, month, and year of birth, or you can calculate percentiles, etc. Let's construct an index for level of confidence in the state.

To compute a new variable, using the add function, go to Transform - > Compute.

Let's construct an index called ConfState. First we need to recode these six items so that the values go from low to high: trt_q610, trt_q630, trt_q640, trt_q650, trt_q660, trt_q670 (I have already done this and will not show it in this example – the recode commands are in the syntax file).

Once you've recoded all your items use trt_q610_R, trt_q630_R, trt_q640_R, trt_q650_R, trt_q660_R, trt_q670_R (these are the recoded variables)

- 1) Write your target variable name.
- 2) Click on Type and Label, to enter the variable label.
- 3) On the left hand side, select the variables you wish to add and divide by the number of items. Make sure you include a plus symbol "+" in between each variable - > this is the add function.
- 6) When you are finished, click on paste. This will paste your commands into your syntax file. Run the command from the syntax file.



Note: I have used the add function “+” (it automatically excludes missing cases. Therefore, any respondent who did not give a valid response to any of the 6 variables will be excluded and will be given a system missing code).

Because the values of the ordinal items used to construct the index of “Level of Confidence in the State” range from 1 to 4 (after I recoded “1” now represents no confidence at all, 2 is “not very much confidence”, 3 is “quite a lot of confidence” and “4” represents a great deal of confidence), I first add the six items and then divide by six so that the index score will range from 1 to 4. Take a look at the table below (remember each respondent gave a response of 1 through 4 for each question):

	trt_q610_R	trt_q630_R	trt_q640_R	trt_q650_R	trt_q660_R	trt_q670_R	Index Value
Respondent 1	4	4	3	4	3	2	3.33
Respondent 2	3	2	3	3	2	2	2.5
Respondent 3	2	1	2	2	3	2	2

What does this tell us? You can rank the respondents (some will have higher levels of confidence in the state than others). Respondent 1 has an index value of 3.33 suggesting

that they have a lot of confidence in the state whereas respondent 3 (with a score of 2) has very little confidence in the state. Take a look at the frequency distribution below:

Frequency distribution for Confidence in the State

ConfState

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	20	.2	.2	.2
	1.17	16	.1	.2	.4
	1.33	30	.2	.4	.8
	1.50	60	.5	.7	1.5
	1.67	116	.9	1.4	2.9
	1.83	182	1.5	2.2	5.1
	2.00	327	2.7	3.9	9.0
	2.17	499	4.1	6.0	14.9
	2.33	686	5.6	8.2	23.1
	2.50	895	7.3	10.7	33.8
	2.67	1088	8.8	13.0	46.8
	2.83	1133	9.2	13.5	60.3
	3.00	1285	10.4	15.3	75.7
	3.17	625	5.1	7.5	83.1
	3.33	477	3.9	5.7	88.8
	3.50	295	2.4	3.5	92.3
	3.67	245	2.0	2.9	95.3
	3.83	146	1.2	1.7	97.0
	4.00	250	2.0	3.0	100.0
	Total		8375	68.1	100.0
Missing	System	3932	31.9		
Total		12307	100.0		

Look at the missing cases, when you use the + function SPSS will assign a missing value if any or all the items have a missing value.

The frequency distribution table tells us that .2% (don't forget to look at the valid percent) of respondent's have no confidence at all in the state. The most frequent category (the mode) is 3; 15.3% of respondent's have quite a lot of confidence in the state. 3% of respondent's have a great deal of confidence in the state.

***Note:** Do not use the compute function to sum nominal values. You can sum nominal values only if they have been re-coded into dummy variables.

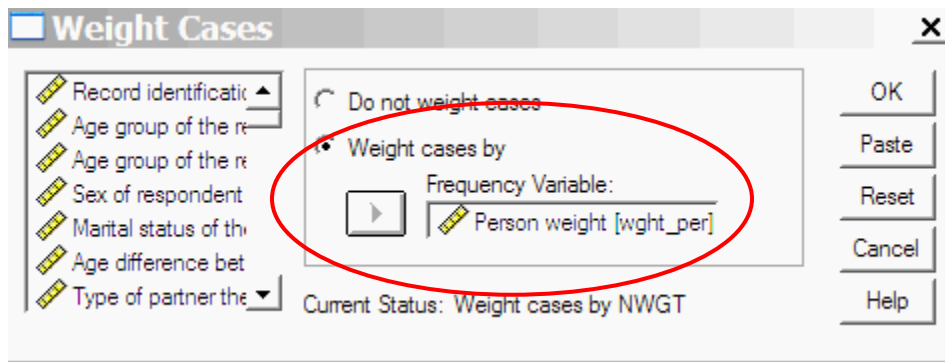
6) Weight Cases in the GSS

Researchers use weight variables to either 1) base the results on the population figure instead of the sample size, 2) and/or to adjust for sampling methods. Researchers adjust

for sampling method because cases (i.e. your unit of analysis – respondent, household, state...) don't all have an equal probability of being selected (i.e. not every case has a weight of 1). The GSS data sets will usually have a weight variable, which will allow you to base the sample on the population (i.e. the number of Canadians aged 15 or older, this will result in a count of over 20 million). However, a large N will usually produce statistically significant results. One way to compensate for this is to re-base the weight variable back to the sample size.

Procedure to weight and re-base the weight variable to the sample size:

- 1) Run a frequency distribution using a simple variable (i.e. sex). Make a note of the total number of cases, including missing cases (this will be the unweighted total N).
- 2) Weight the data using the WEIGHT BY command and the name of the weighted variable that comes with the data. In the data editor go to data - > Weight Cases - > select the variable and click the arrow. Paste the variable into the syntax file; this will keep the procedure for future reference.



- 3) Rerun a frequency on the variable sex and make note of the total number of cases including the missing cases (this will be the total weighted N).
- 4) Then include the following command at the beginning of your syntax file:
`COMPUTE NWGT = wght_per*(tnw/twt).`
`WEIGHT BY NWGT.`
 (wght_per is the name of the weight variable, tnw is the N from step 1 and twt is N from step 3)

Once this is entered, highlight the command in your syntax file and run the command.

Below is a frequency distribution table of Marital Status from the GSS cycle 17 (I am working with a sample of 50% of the GSS). I have listed the frequencies for marital status using the original data file (unweighted), the weighted data and re-based. As you can see, the N (total) is the same for the unweighted and re-based data, but the N for the weighted data is just under 13 million (this is about half the population since I am using half of the full GSS). If you look in the married column you can see that the percentage of married couples is 45.9% when the data is unweighted. However, when the data is

weighted the percentage increases to 51.3%. The percent remains the same when you re-base the data (51.3%), although the frequency (i.e. the number of cases) changes.

		(Unweighted)		(Weighted)		(Re-Based)	
Marital status of the respondent		Frequency	Percent %	Frequency	Percent %	Frequency	Percent %
Valid	Married	5691	45.9	6491377	51.3	6360	51.3
	Living common-law	1068	8.6	1250336	9.9	1225	9.9
	Widowed	1088	8.8	689433	5.4	675	5.4
	Divorced	492	4.0	328991	2.6	322	2.6
	Separated	866	7.0	579127	4.6	567	4.6
	Single (Never married)	3166	25.5	3285937	26.0	3219	26.0
	Total	12371	99.8	12625201	99.8	12370	99.8
Missing	Not stated	16	0.1	16316	0.1	16	0.1
	Don't know	10	0.1	11431	0.1	11	0.1
	Total	26	0.2	27747	0.2	27	0.2
	Total N	12397	100	12652948	100	12397	100