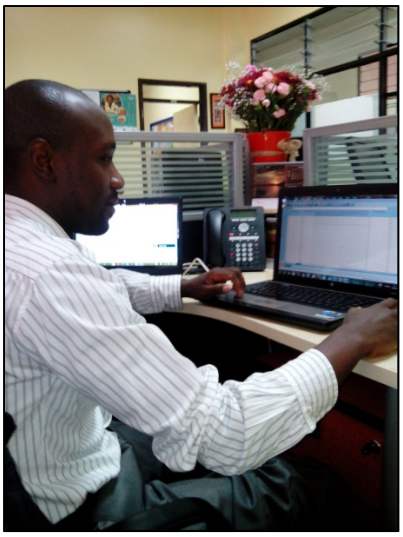
Foreign Affairs, Trade and Development Canada

Helen Keller INTERNATIONAL

STATISTICAL ANALYSIS OF POST EVENT COVERAGE SURVEY (PECS) DATA USING SPSS



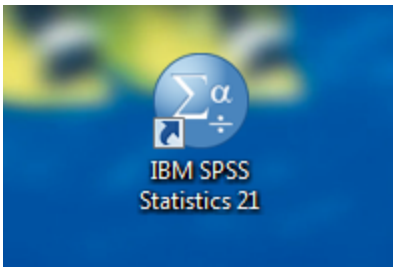
March 2014

PECS Survey Data Analysis using SPSS V2

Opening Data Files in SPSS

How to open SPSS

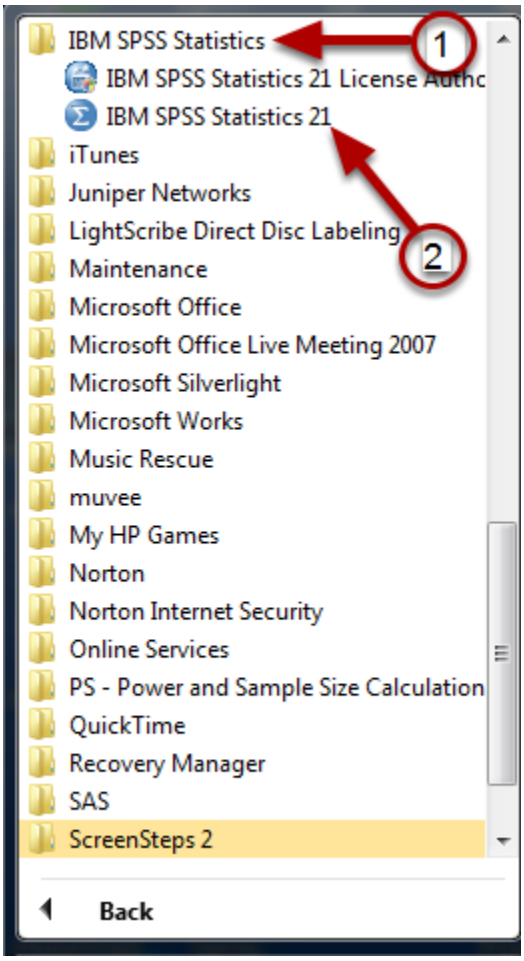
If you have a SPSS icon on your desktop



Double click the IBM SPSS Statistics Icon to open SPSS.

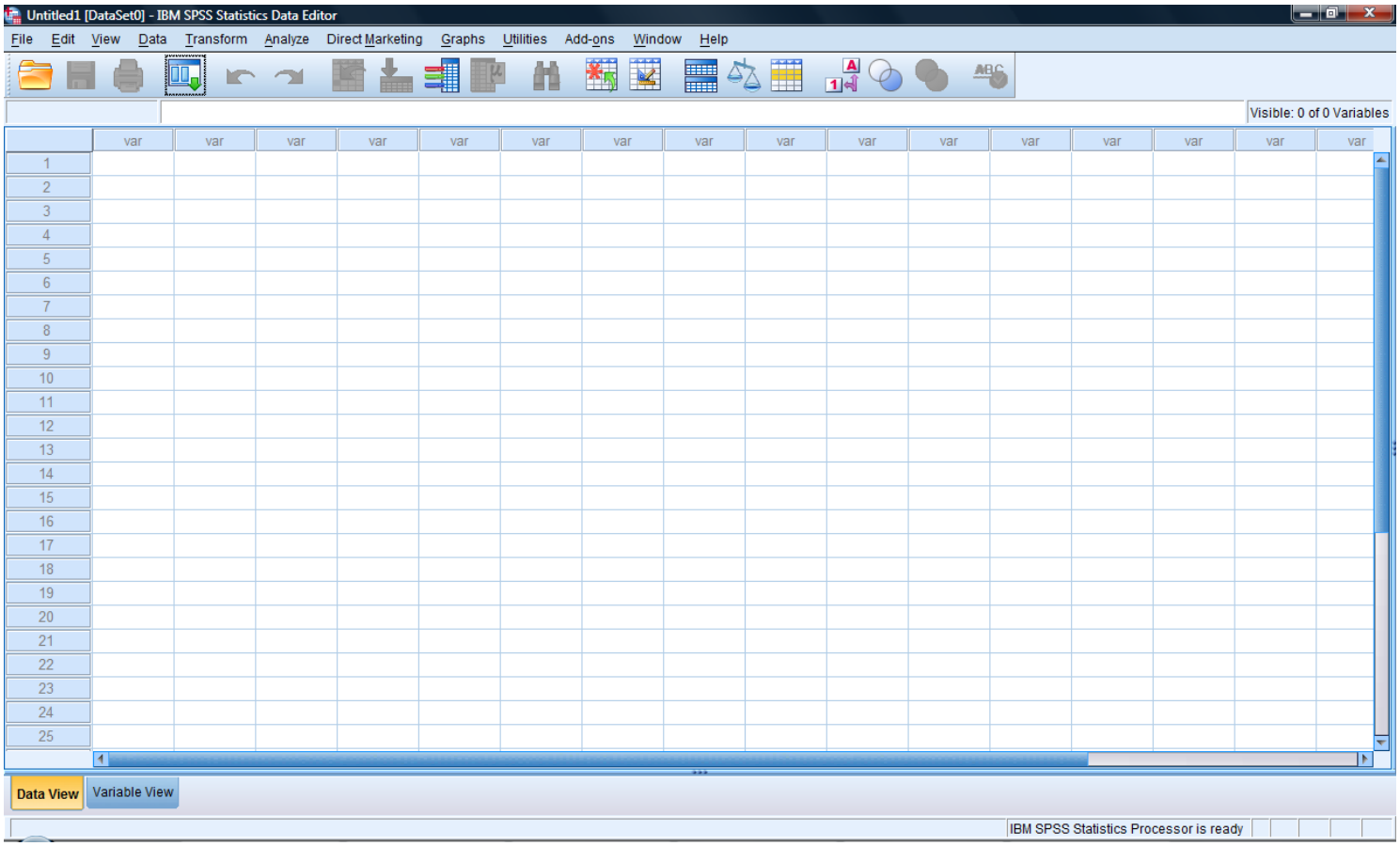
If you do not have an SPSS icon on your desktop

Click on the Start menu as shown in #1. Then left click "All Programs" (#2) to display the programs on your computer.



Click the IBM Statistics folder (#1) to display its contents and then left click "IBM SPSS Statistics" (#2). This will open the SPSS program.

Once open, the program will look something like this.

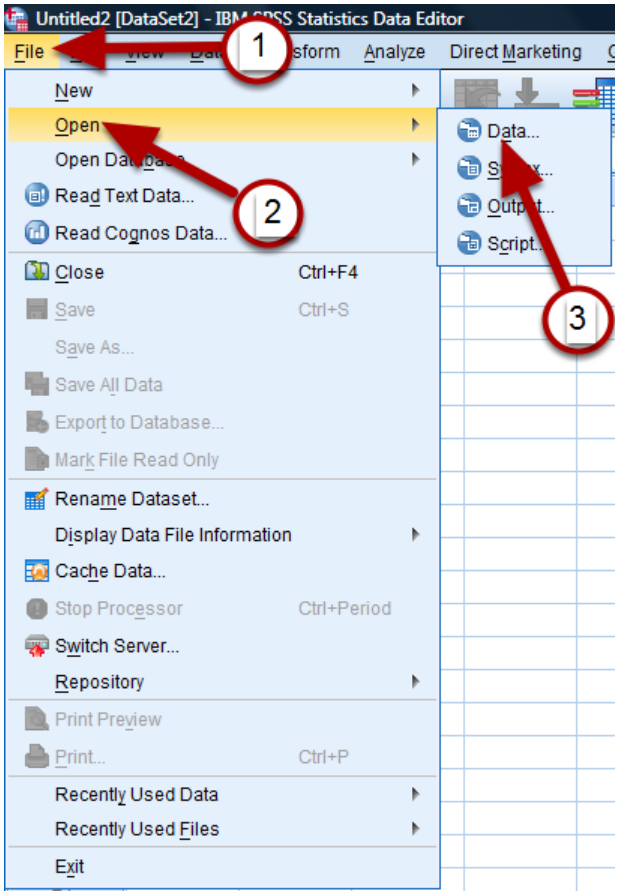


You are now ready to import your data and begin the survey analysis!

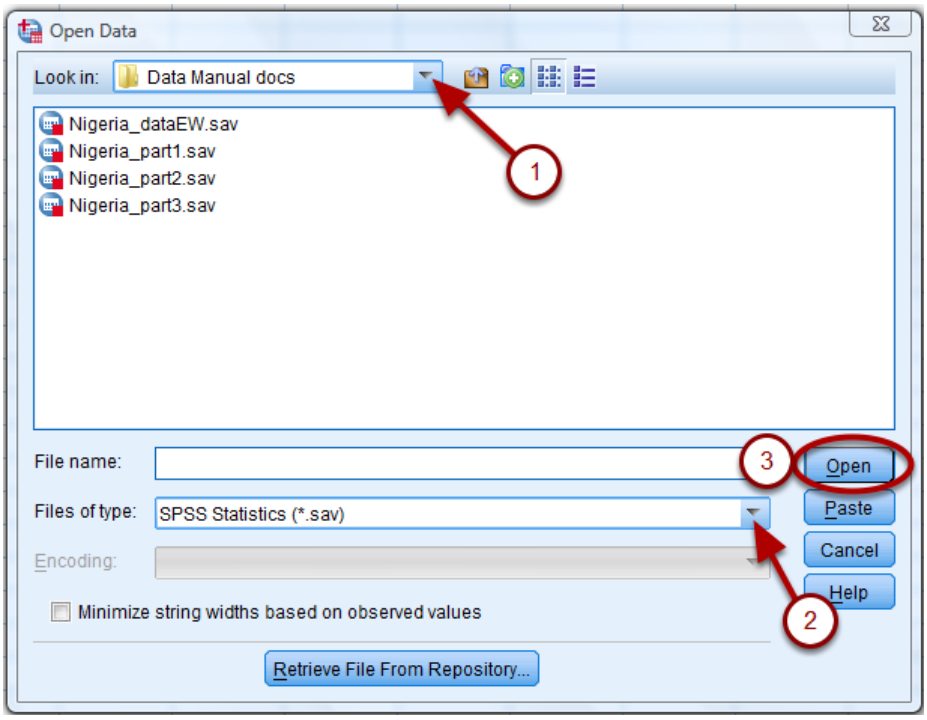
Importing Data from SPSS, Excel or Stata Files

Data can be imported from a SPSS (*.sav, *.sys, *.por), Excel (*.xlm, *.xlsx, *.xlsm), dbase (*.dbase), or STATA(*.dta) file. **NOTE: These procedures will not work if your data file is open in another window.**

Select the data file to be opened.

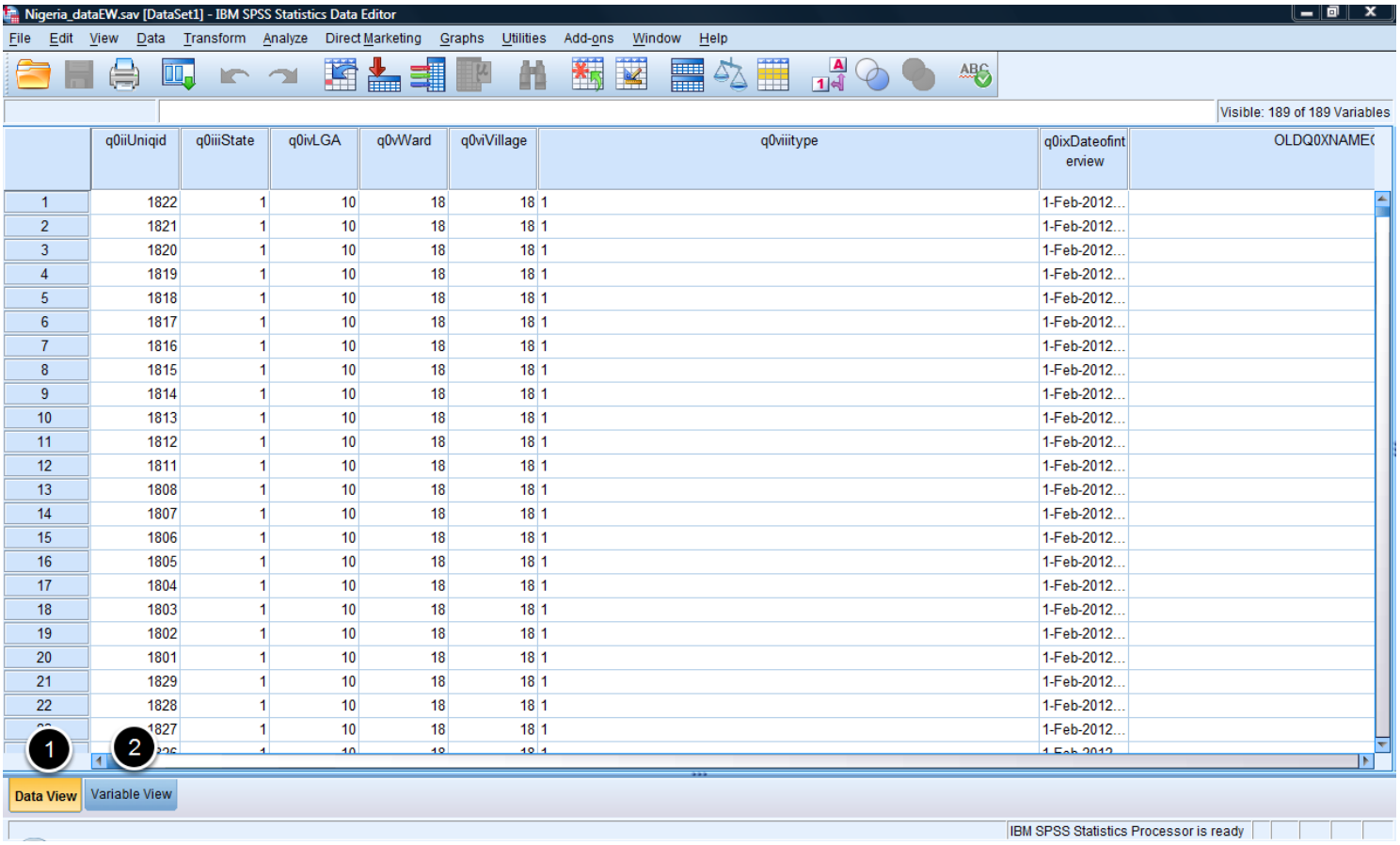


First left click "File" (#1). From the drop down menu select "Open" (#2). Then left click on "Data" (#3).



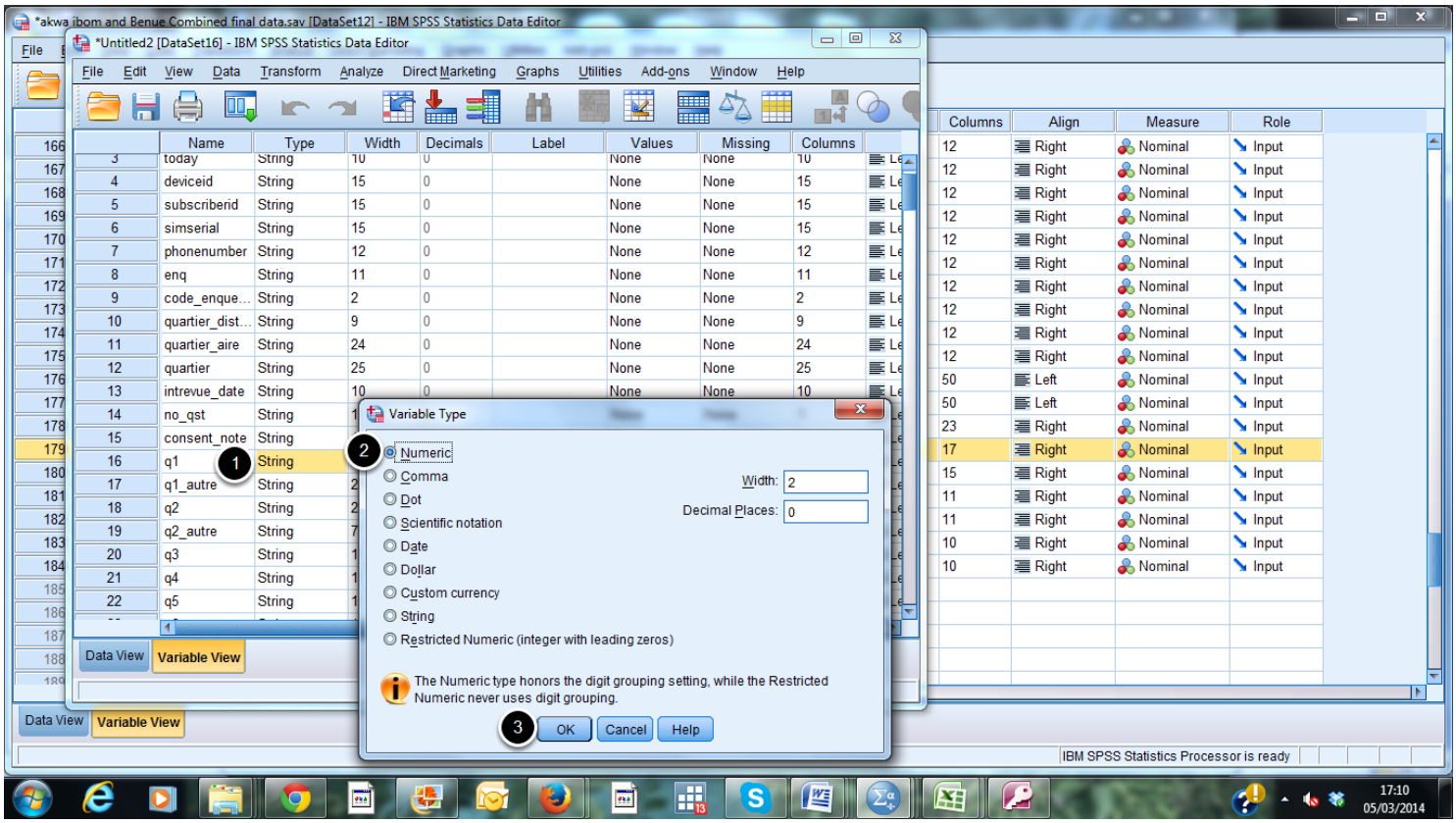
In the "Look in:" drop down menu (#1), go to the location where your data set is saved. If your data is not a SPSS file with a *.sav extension, click the "Files of type" drop down menu and select the correct file type (#2) . If you are not sure of the file type, select "All files (*.*) from the "Files of type" menu. Once located, single left click the desired data set and then click "Open" (#3).

Your data will be displayed in DataView



As the name suggests, Data View (#1) provides a window for viewing data, whereas Variable View (#2) shows the variable names and characteristics.

Reclassify Categorical Data as Numeric



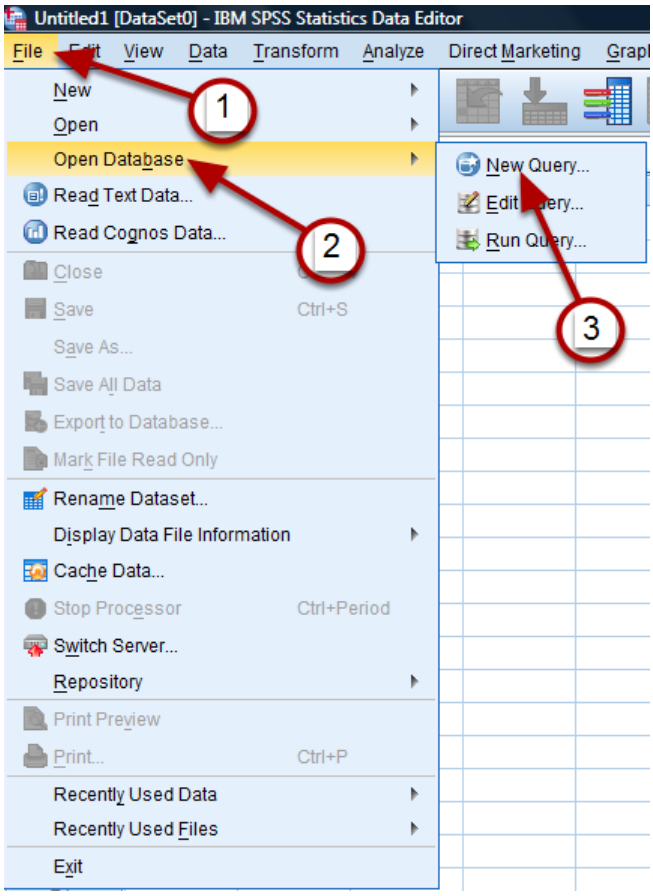
For all variables with categorical data, ensure that the variable type is listed as Numeric and not String in the Variable View. To change the Type of a variable, click on the cell where "String" is given and change the Variable Type to Numeric. Then click on "OK"

Importing Data from dBase or Access Files

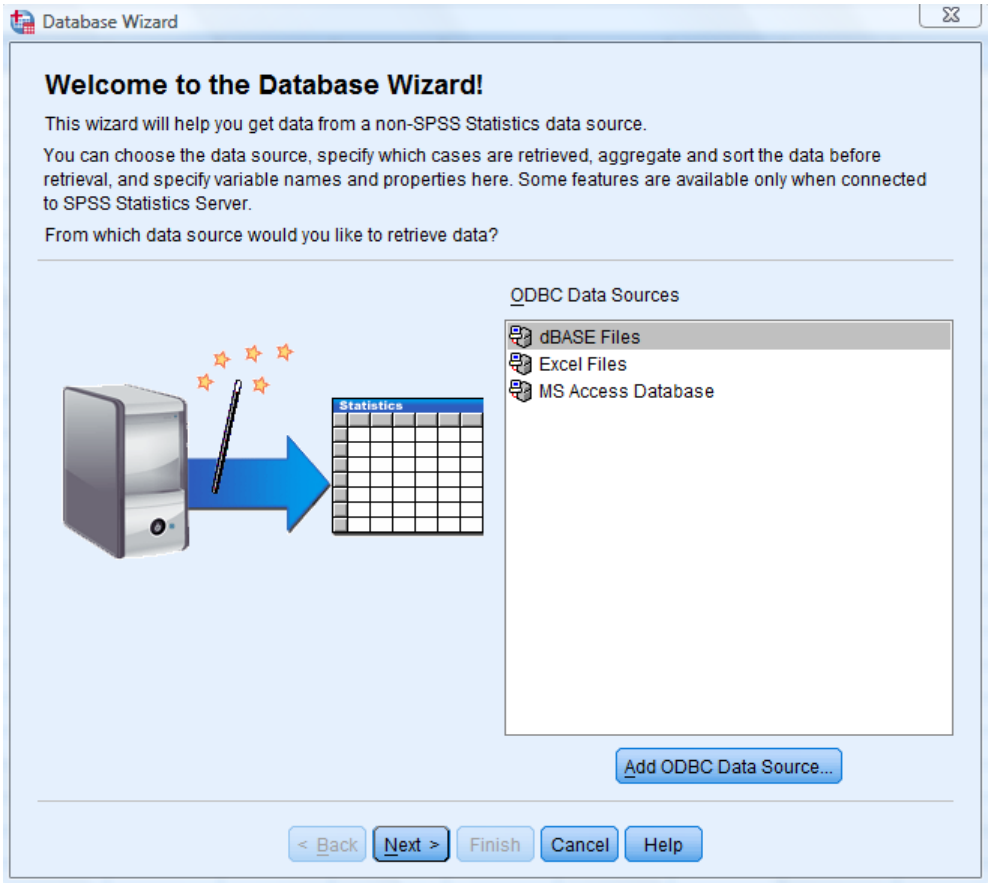
Data from dBase files (*.dbase) or Microsoft Access files (*.mdb) file can be imported into SPSS using the following steps:

NOTE: These procedures will not work if your data file is open in another window.

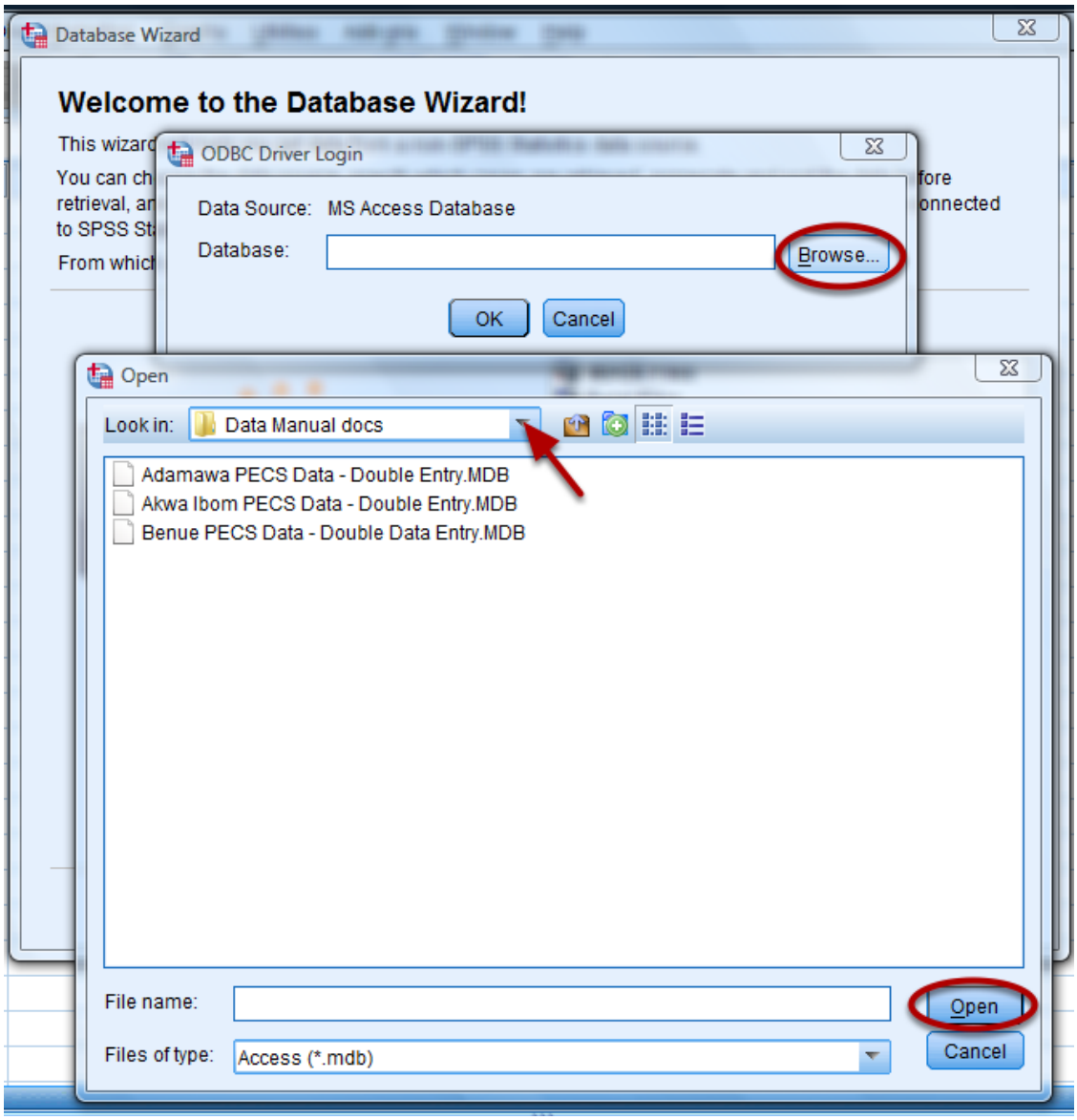
Select the data file to be opened



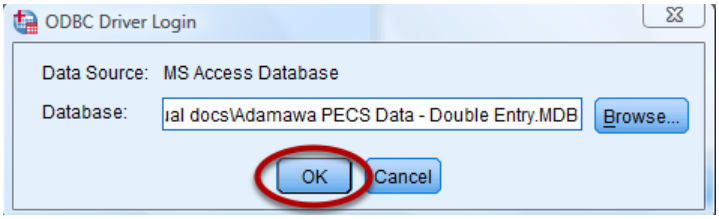
To open data from a dBase or Access file, first single left click “File” (#1), then “Open Database” (#2), and finally “New Query” (#3).



Select the appropriate data source from the list (dBASE File, Excel File or MS Access Database) and click "Next."

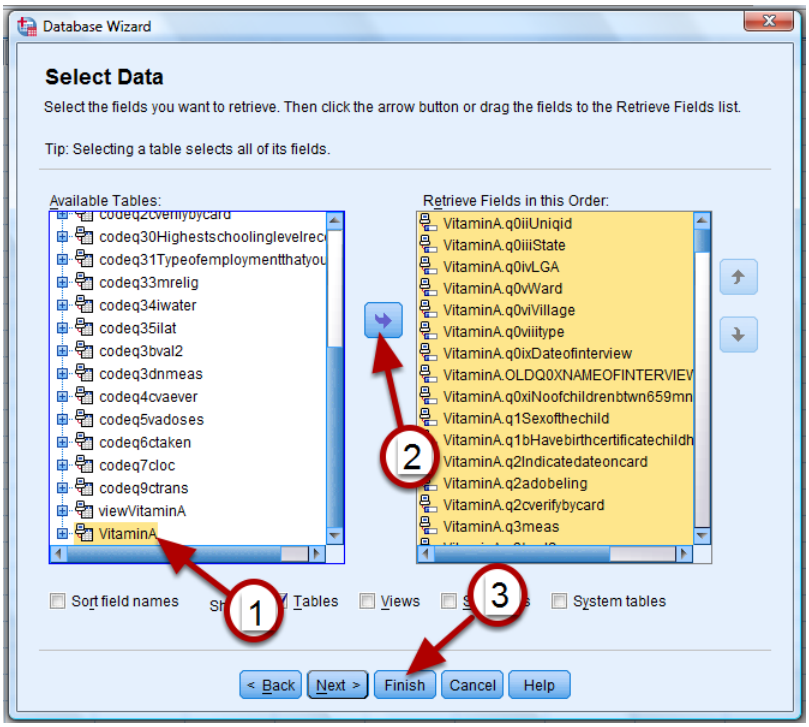


To locate your data file, click Browse and follow the path to the file using the "Look in:" drop down menu. Once you have located the correct folder, click the desired data file and single left click "Open".



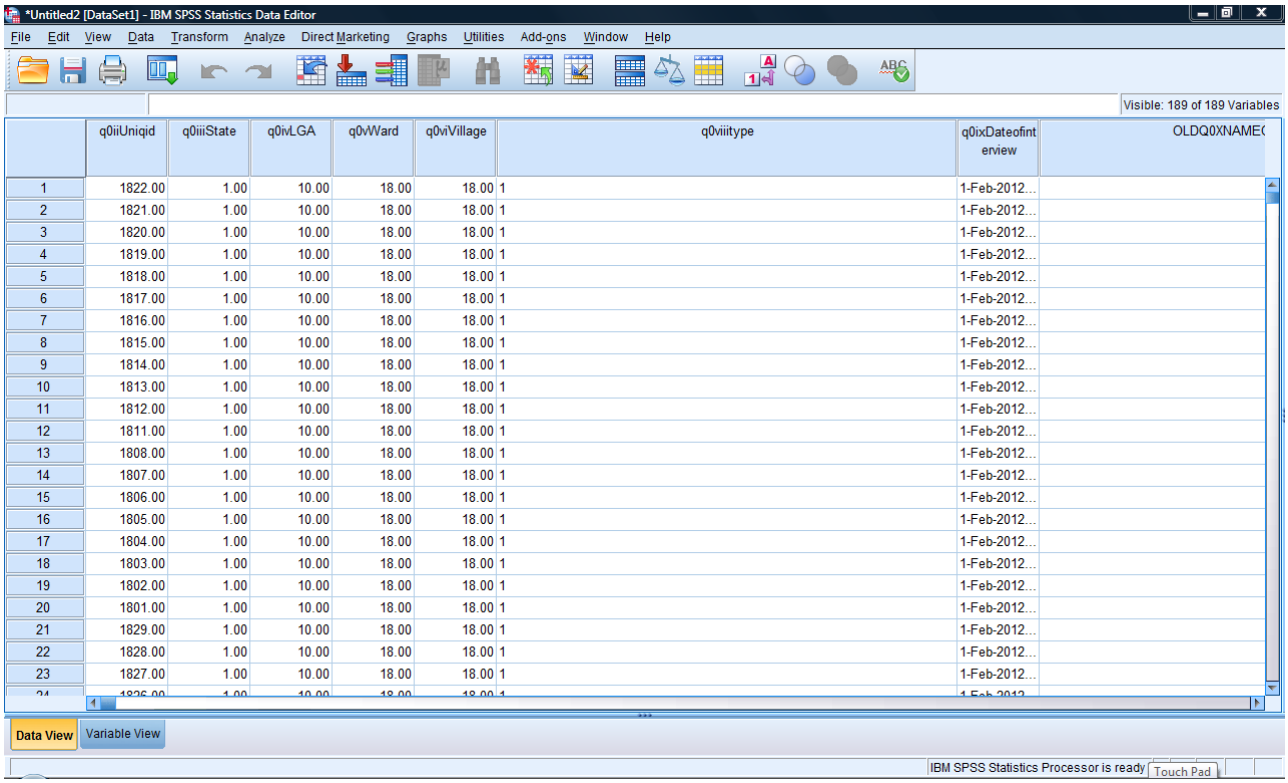
Click "OK" to continue the database wizard.

Import data from the selected database



Select the table that contains the data you will be analyzing from the "Available Tables" column on the left (#1) and click the blue arrow (#2) to move its variables to the "Retrieve Fields in this Order" column on the right. All of the variables in the database should now be listed in the "Retrieve Fields in this Order" column. Then single left click the "Finish" button (#3).

Your data will be displayed in the Data View window



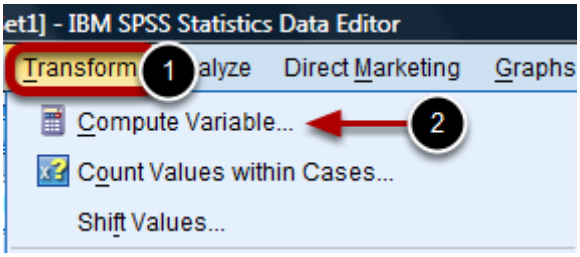
	q0iiUniqid	q0iiiState	q0ivLGA	q0vWard	q0viVillage	q0viitype	q0ixDateofinterview	OLDQ0XNAME(
1	1822.00	1.00	10.00	18.00	18.00	1	1-Feb-2012...	
2	1821.00	1.00	10.00	18.00	18.00	1	1-Feb-2012...	
3	1820.00	1.00	10.00	18.00	18.00	1	1-Feb-2012...	
4	1819.00	1.00	10.00	18.00	18.00	1	1-Feb-2012...	
5	1818.00	1.00	10.00	18.00	18.00	1	1-Feb-2012...	
6	1817.00	1.00	10.00	18.00	18.00	1	1-Feb-2012...	
7	1816.00	1.00	10.00	18.00	18.00	1	1-Feb-2012...	
8	1815.00	1.00	10.00	18.00	18.00	1	1-Feb-2012...	
9	1814.00	1.00	10.00	18.00	18.00	1	1-Feb-2012...	
10	1813.00	1.00	10.00	18.00	18.00	1	1-Feb-2012...	
11	1812.00	1.00	10.00	18.00	18.00	1	1-Feb-2012...	
12	1811.00	1.00	10.00	18.00	18.00	1	1-Feb-2012...	
13	1808.00	1.00	10.00	18.00	18.00	1	1-Feb-2012...	
14	1807.00	1.00	10.00	18.00	18.00	1	1-Feb-2012...	
15	1806.00	1.00	10.00	18.00	18.00	1	1-Feb-2012...	
16	1805.00	1.00	10.00	18.00	18.00	1	1-Feb-2012...	
17	1804.00	1.00	10.00	18.00	18.00	1	1-Feb-2012...	
18	1803.00	1.00	10.00	18.00	18.00	1	1-Feb-2012...	
19	1802.00	1.00	10.00	18.00	18.00	1	1-Feb-2012...	
20	1801.00	1.00	10.00	18.00	18.00	1	1-Feb-2012...	
21	1829.00	1.00	10.00	18.00	18.00	1	1-Feb-2012...	
22	1828.00	1.00	10.00	18.00	18.00	1	1-Feb-2012...	
23	1827.00	1.00	10.00	18.00	18.00	1	1-Feb-2012...	

Data Cleaning

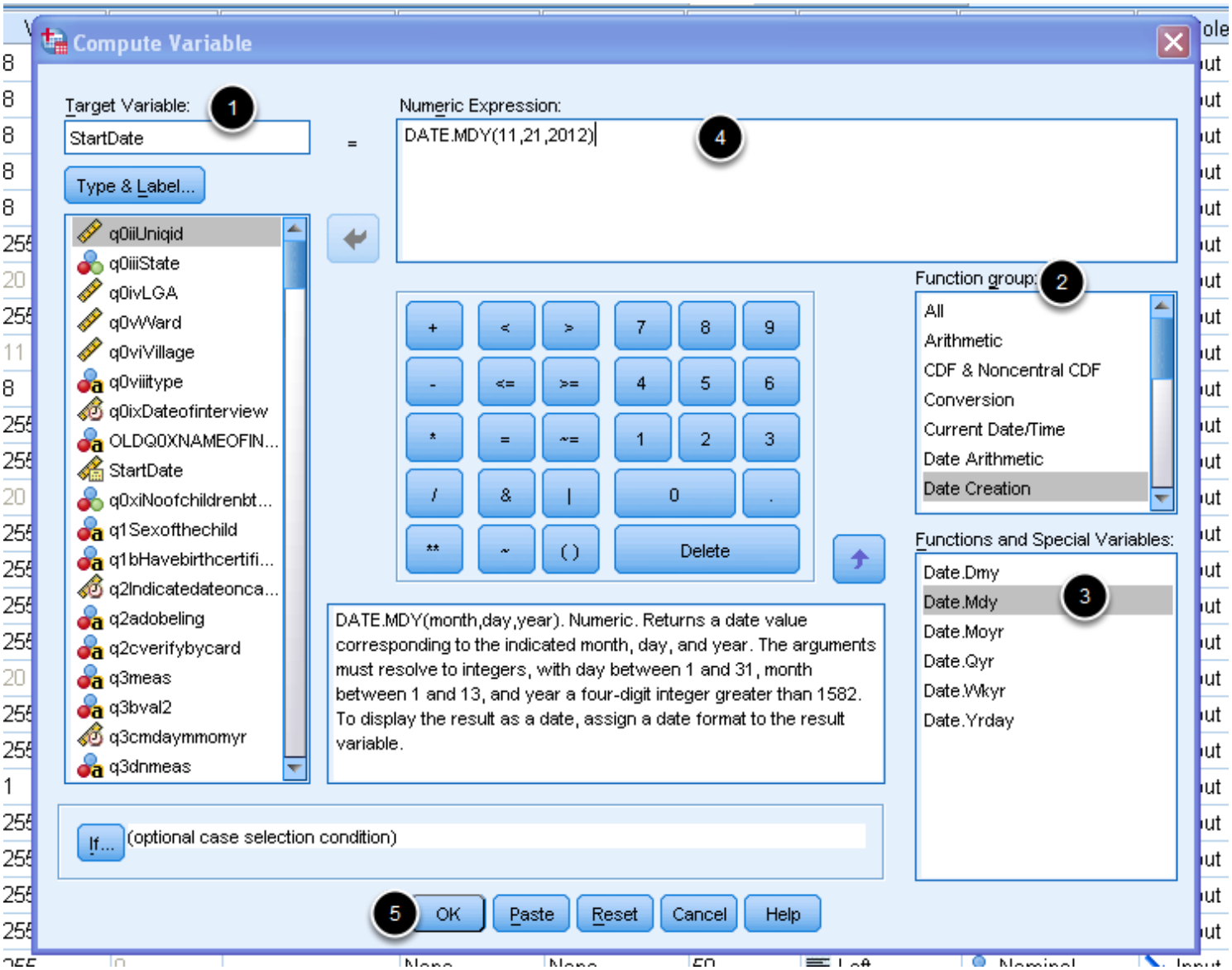
Calculating Age from Date of Birth and Confirming Eligibility

Before analyzing data, it is important to ensure all respondents meet the eligibility requirements. For vitamin A supplementation post event coverage surveys (PECS), children must have been between 6 and 59 months of age at the time of the supplementation event. During the data cleaning step, all children outside of the 6-59 month age range should be identified and excluded from the analysis.

Create a variable containing the last date of the vitamin A supplementation event



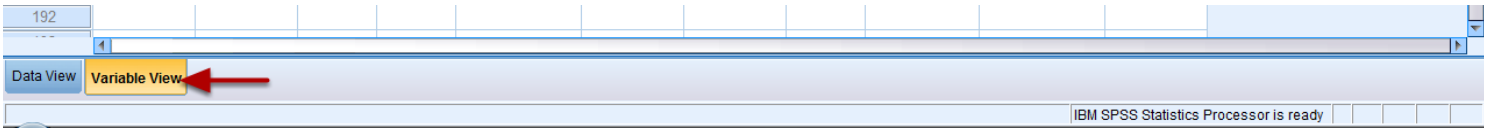
To enter the date of the vitamin A supplementation event, on the top menu, click "Transform" (#1) and select the Compute Variable option (#2).



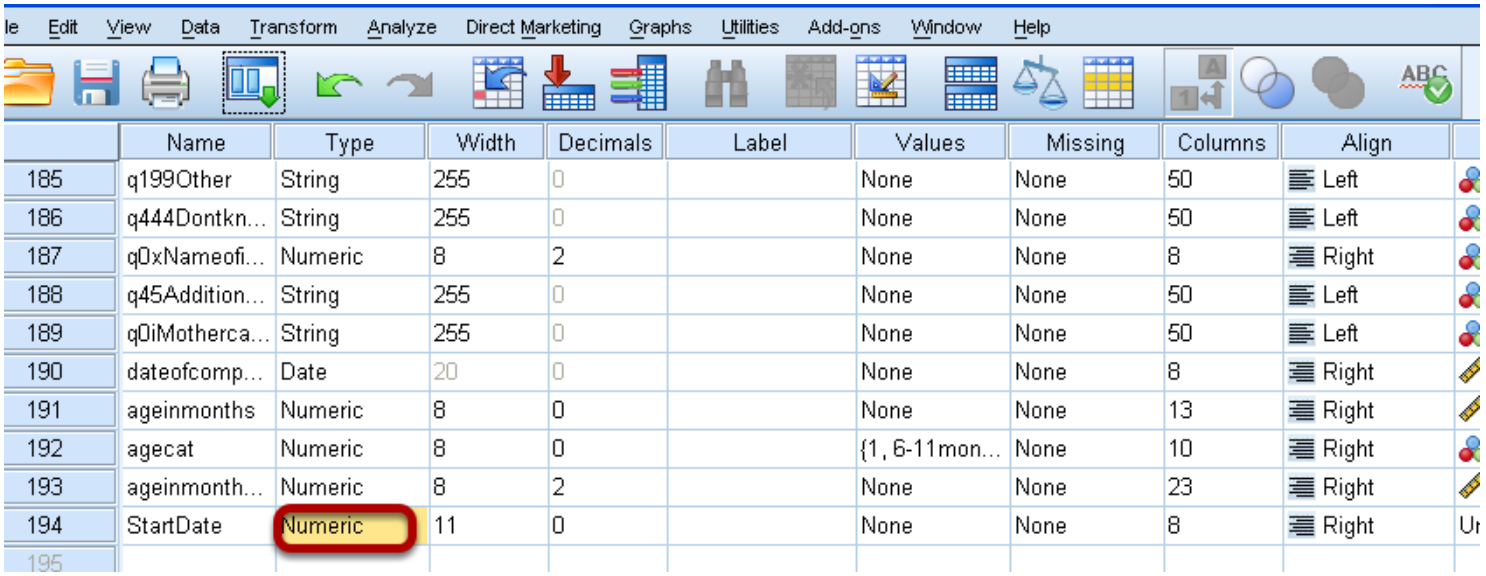
When the compute variable window opens, perform the following steps to add the last date of the supplementation event to each row:

1. Enter the name of the variable you want to create for the date of the supplementation event in the Target Variable box (#1).
2. In the function group list, left click "Date Creation" (#2).
3. Double click the format you wish to use to enter your date (#3). In this example the mm/dd/yyyy (Mdy) format was selected.
4. DATE.MDY(?,?,?) automatically appears in the Numeric Expression window. Inside the parentheses, change the "?" to the month, day, and year of the last date of the supplementation event (#4). Take care not to remove the commas or parentheses.
5. Click "OK" (#5).

Change the format of the new variable to date format

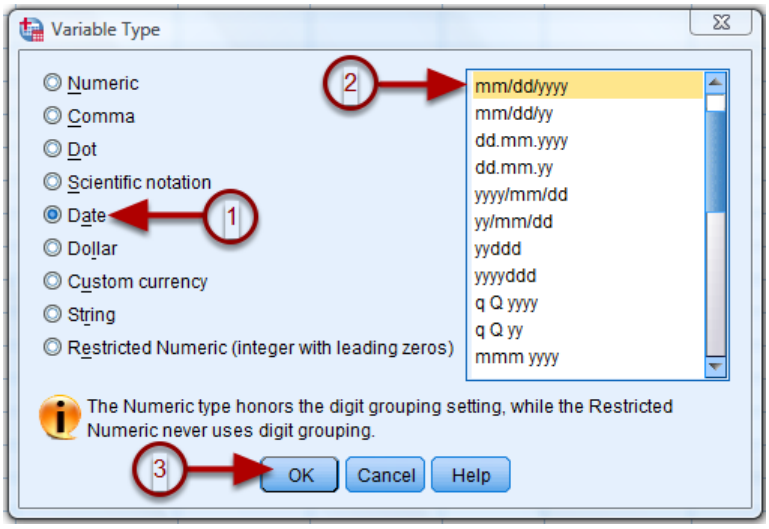


Click the "Variable View" tab at the bottom of the data table. This will display all of the variables in your database.



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align
185	q199Other	String	255	0		None	None	50	Left
186	q444Dontkn...	String	255	0		None	None	50	Left
187	q0xNameofi...	Numeric	8	2		None	None	8	Right
188	q45Addition...	String	255	0		None	None	50	Left
189	q0iMotherca...	String	255	0		None	None	50	Left
190	dateofcomp...	Date	20	0		None	None	8	Right
191	ageinmonths	Numeric	8	0		None	None	13	Right
192	agecat	Numeric	8	0		{1, 6-11mon...	None	10	Right
193	ageinmonth...	Numeric	8	2		None	None	23	Right
194	StartDate	Numeric	11	0		None	None	8	Right
195									

Scroll down to the last variable to find the new variable name that you entered. Then click on the cell in the "Type" column.



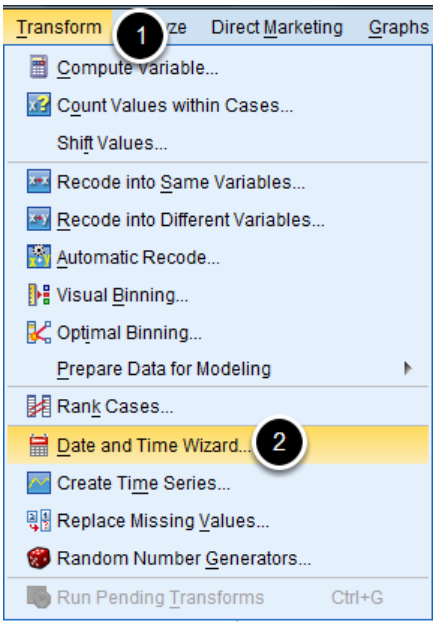
When the Variable Type window opens, first select "Date" (#1) and then select the date format that you prefer (#2). For example, in the example above, the "mm/dd/yyyy" format is selected. Once this is complete click "OK" (#3). Return to the Data View window and confirm that the new variable

has been entered correctly. The new variable will appear as the last column in the database.

Events with varying dates of distribution

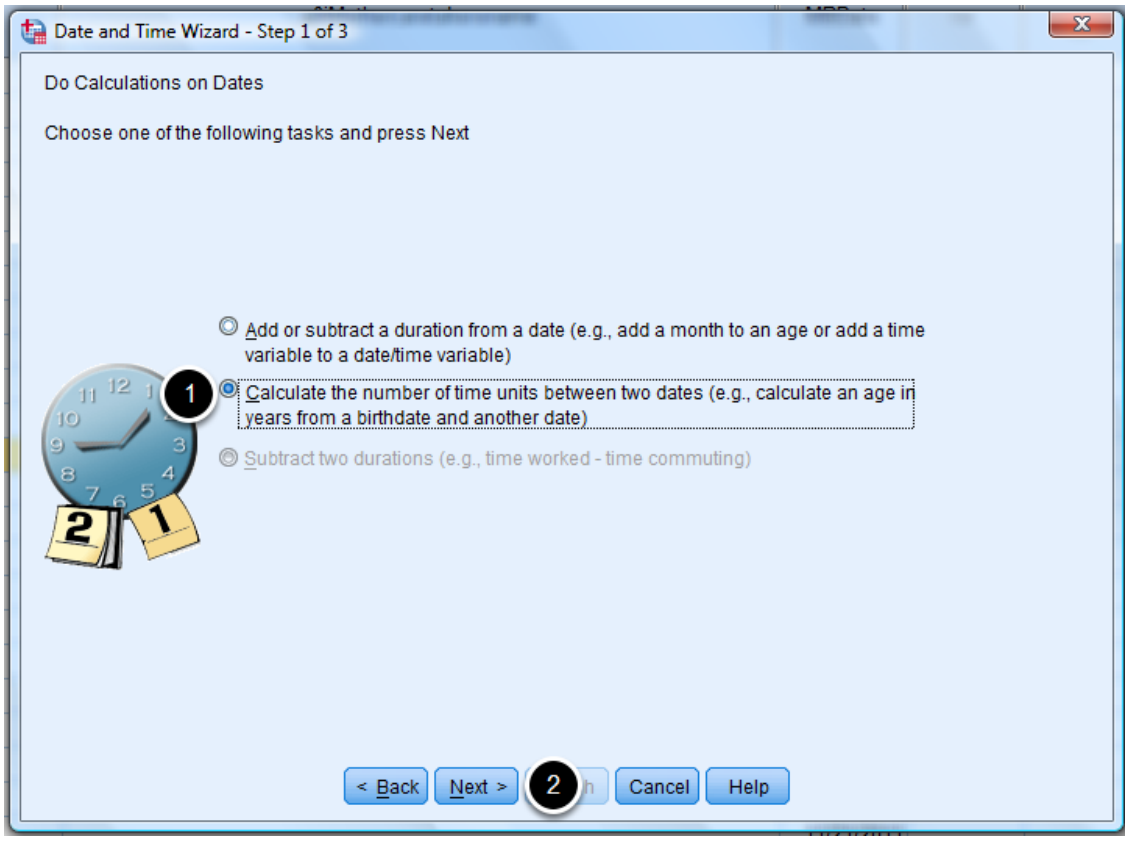
If different survey locations had different dates of distribution, then use the "If" condition in the Compute Variable window (#1). The "Compute variable: if case" window appears.
Select "Include case if satisfies condition" (#2).
From the list of variables in the first column, select the one that identifies the various regions. In this case, dates of distribution differed by state hence state was selected (#3). Click the blue arrow (#4) to send it to the text box.
Specify the condition to be met. e.g. state = 1. Then click continue (#5).
Enter the date of distribution for that location and select OK (#6).
Repeat this procedure for each region until the dates for all regions have been entered.

Calculate the age of each child during the supplementation event

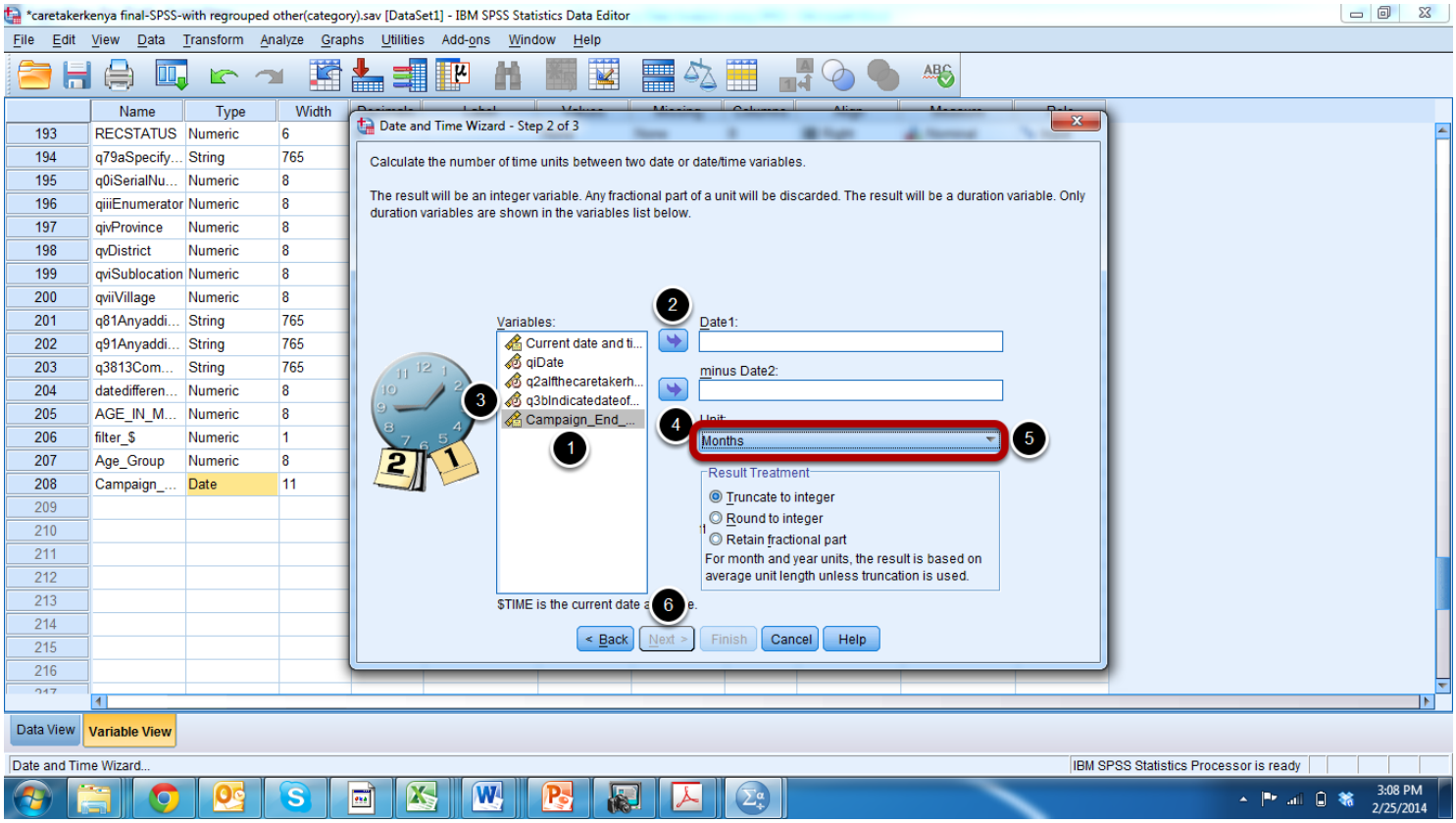


Calculate the age of each child at the last vitamin Asupplementation event for children with known dates of birth using the Date and Time Wizard. First click "Transform" (#1) and then select "Date and Time Wizard" (#2) from the drop down menu.

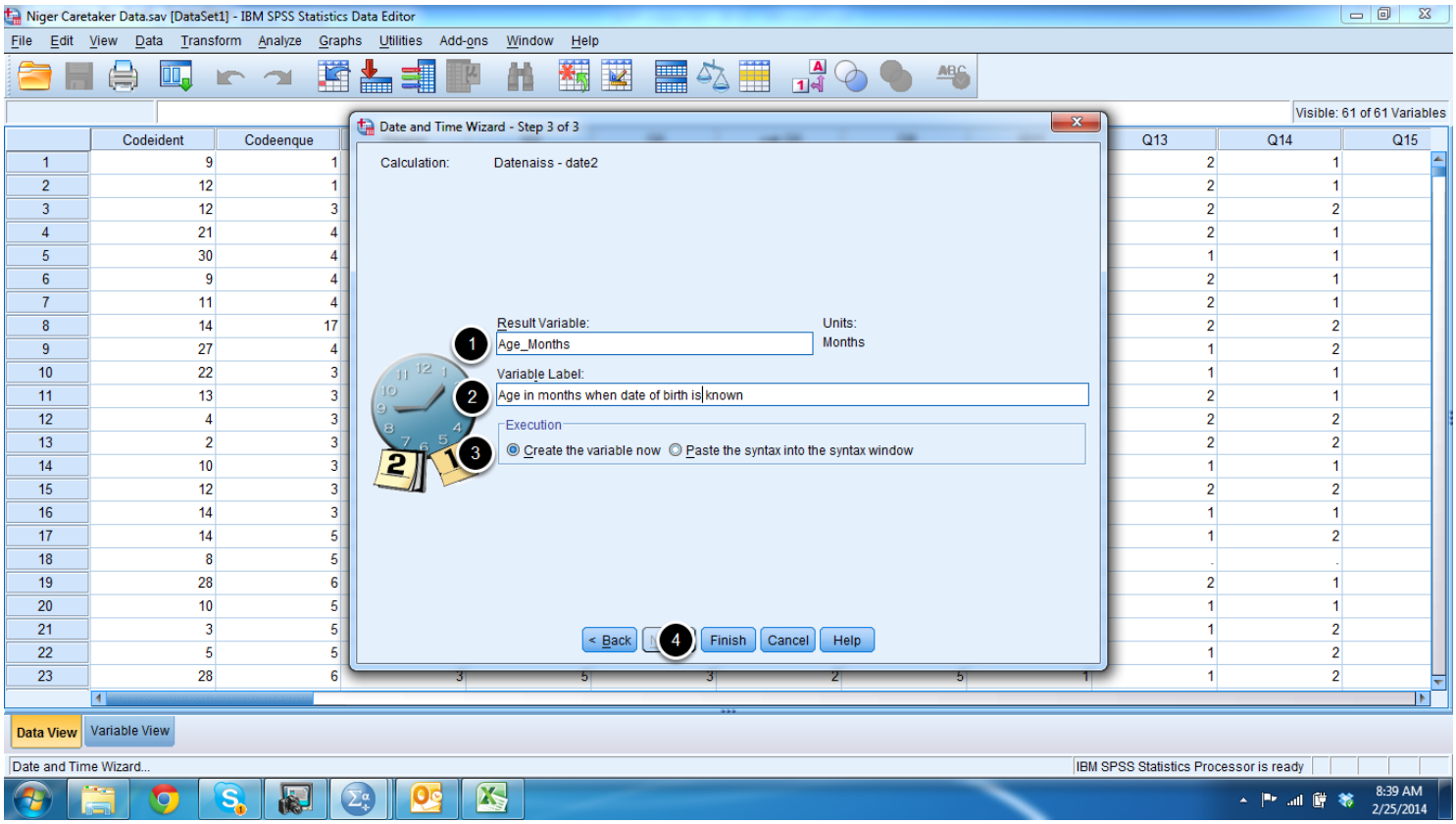
When the wizard starts, select "Calculate with dates and times" (#1) and left click "Next" (#2).



For step 1 of 3, choose "Calculate the number of time units between two dates" (#1) and left click "Next" (#2).

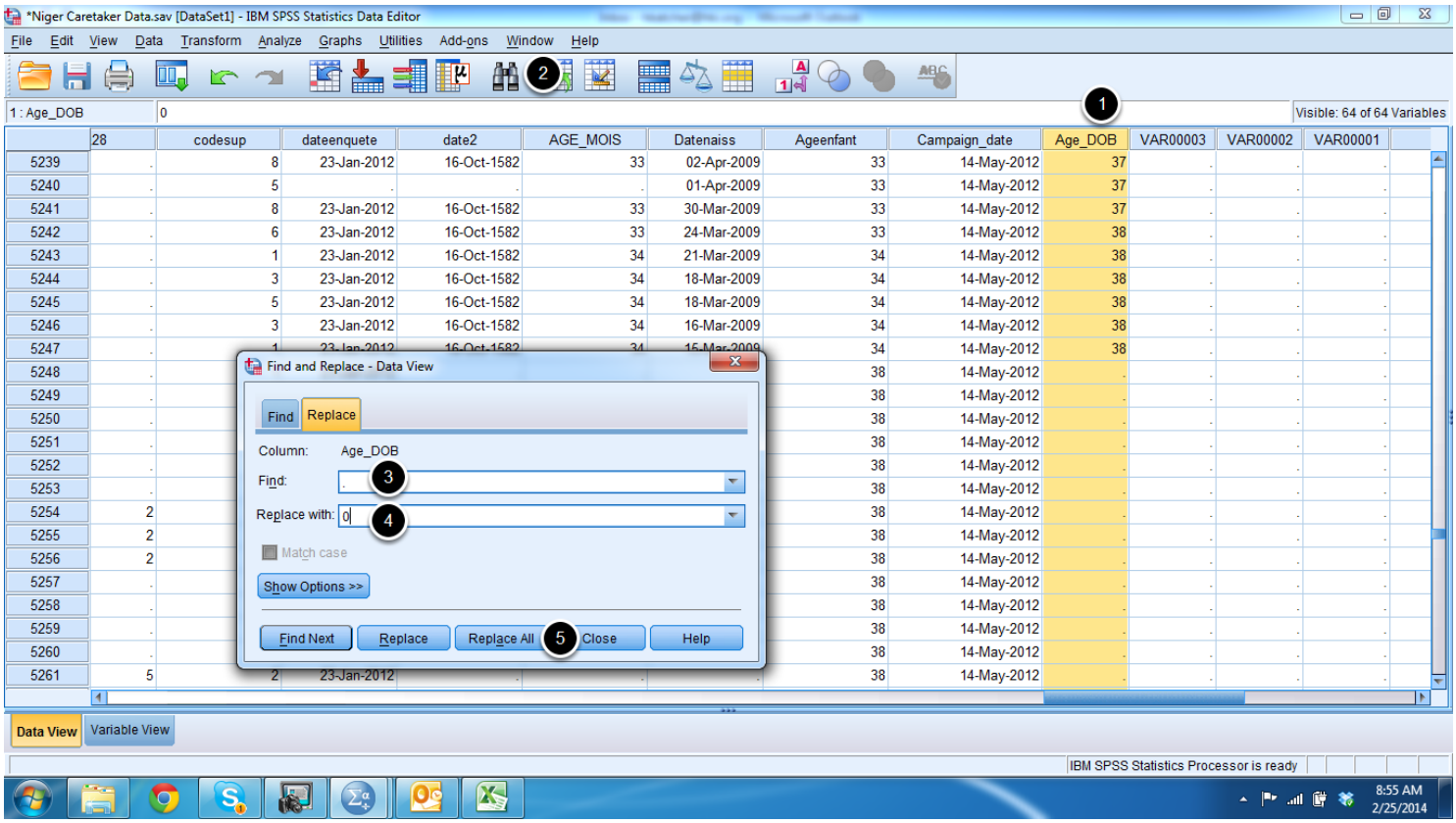


Step 2 of 3 in the Date and Time Wizard is where you will select the two dates from column #1 that you will use to calculate each child's' age. First select the date of the vitamin Asupplementation event that you created in the Variables box (#1) and click on the right arrow (#2). Then select the variable name for the child's date of birth (#3) and click on the lower right arrow (#4). To have the child's age calculated in months, select "Months" from the Unit drop down menu (#5). Finally choose "Truncate to integer" and then left click "Next" (#6).



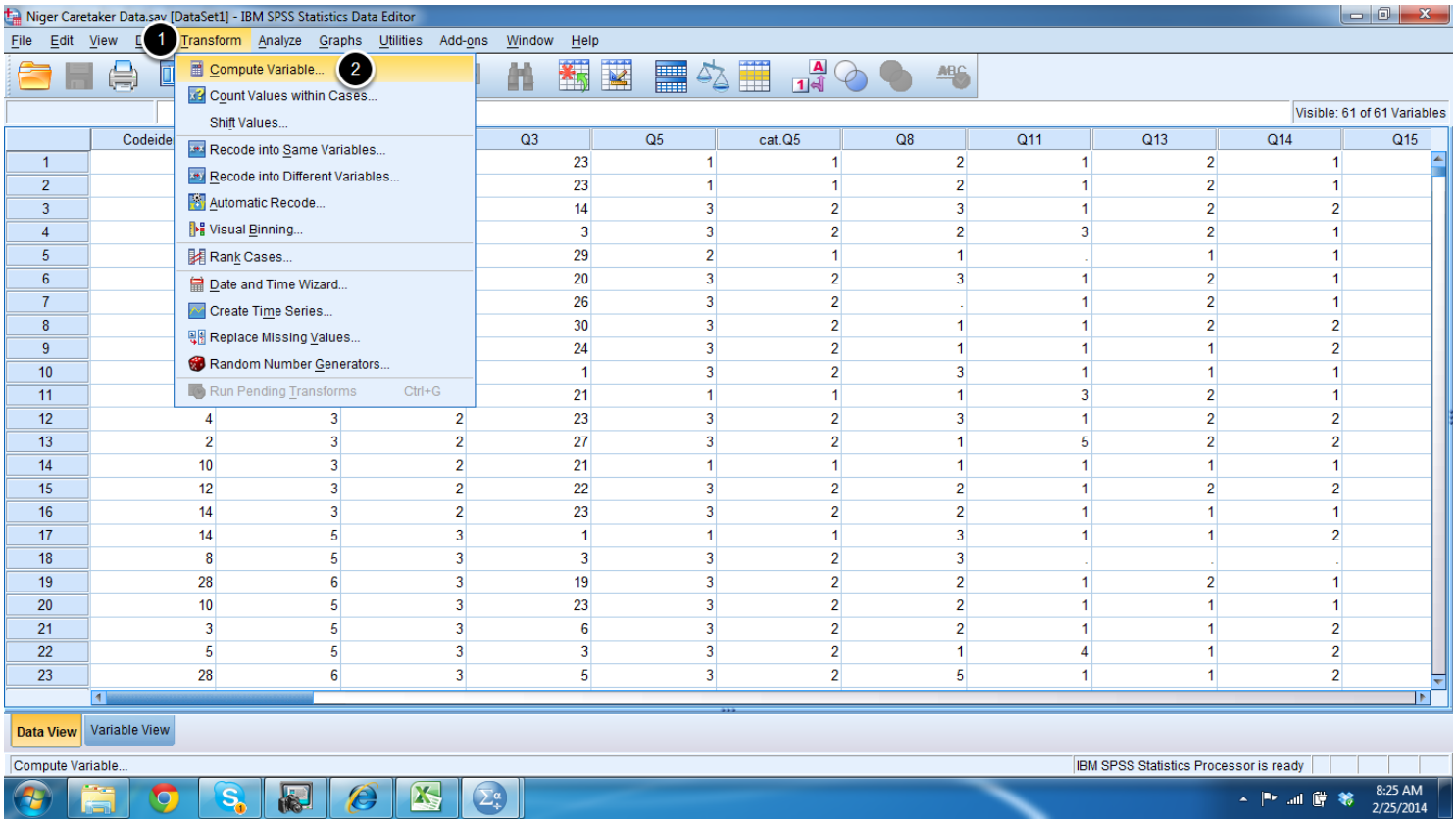
In the final step, Step 3 of 3, assign a new variable name for the age of the child at the time of the supplementation event (#1). Then add a description of the variable in the "Variable Label" field (#2). Ensure that the "Create the variable now" option is selected in the Execution box (#3). Then click on "Finish" (#4).

Recode missing data as 0

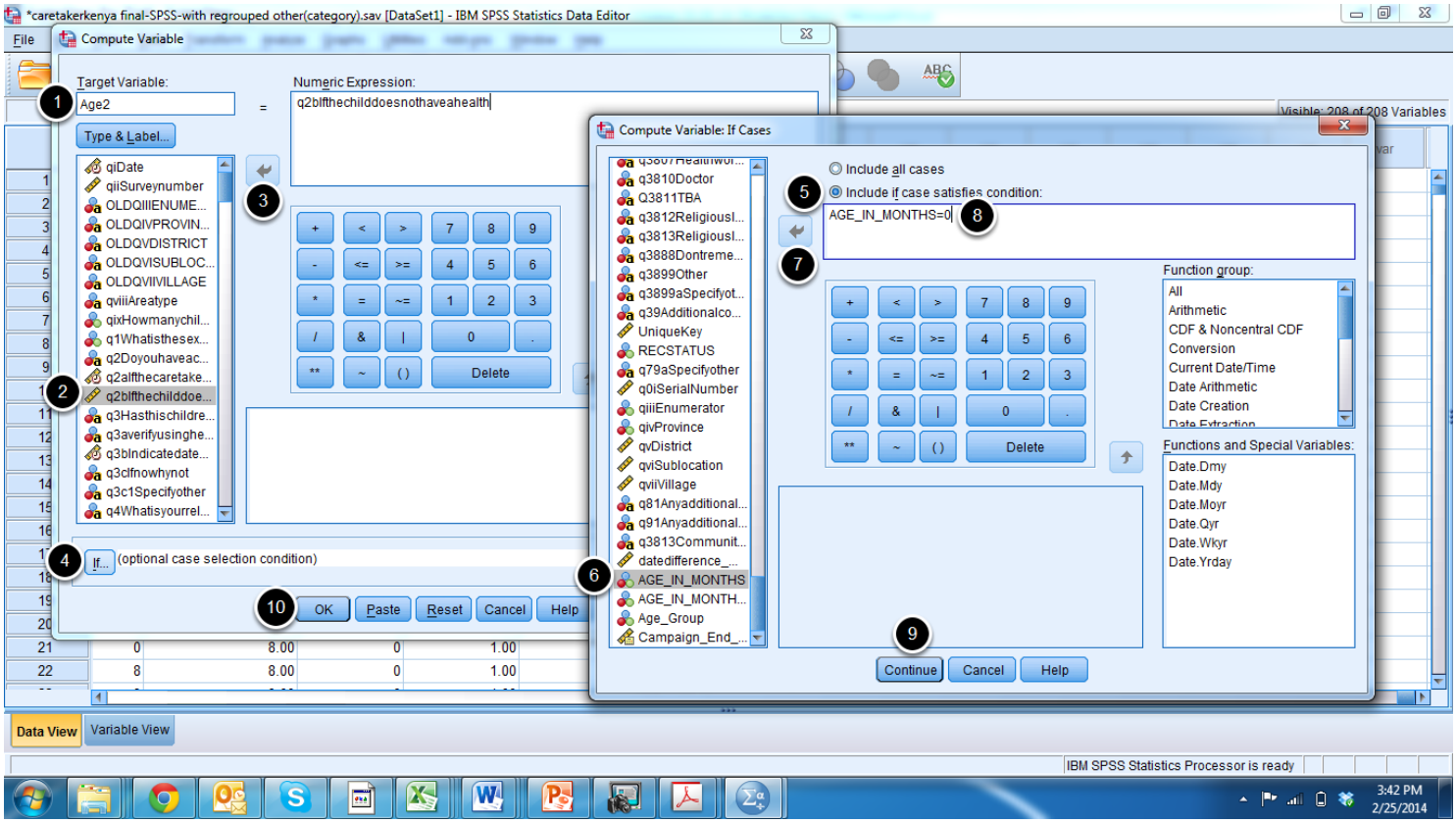


Select the column in which age was calculated from date of birth (#1). Click on the Find icon (#2) to open the Find and Replace window. In the find section (#3) type . (dot) as shown in the photo above. In the "Replace With" section (#4), type 0 (zero) and click "Replace All" (#5). Using this procedure, all cases where the child's date of birth is not known will be replaced a with 0 (zero).

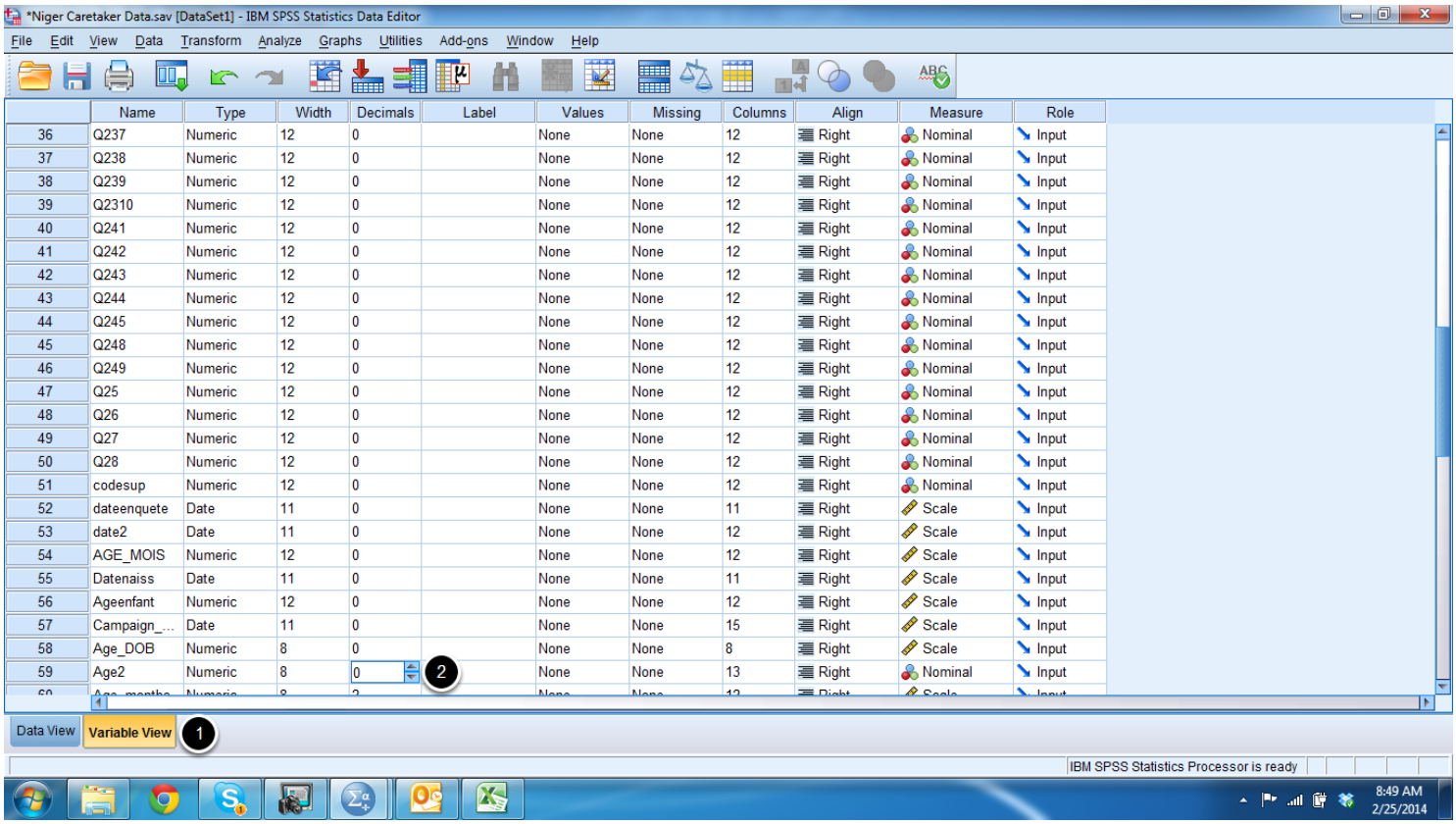
Create a new variable containing the age of children whose date of birth is not known



Select Transform (#1) and then Compute Variable (#2).

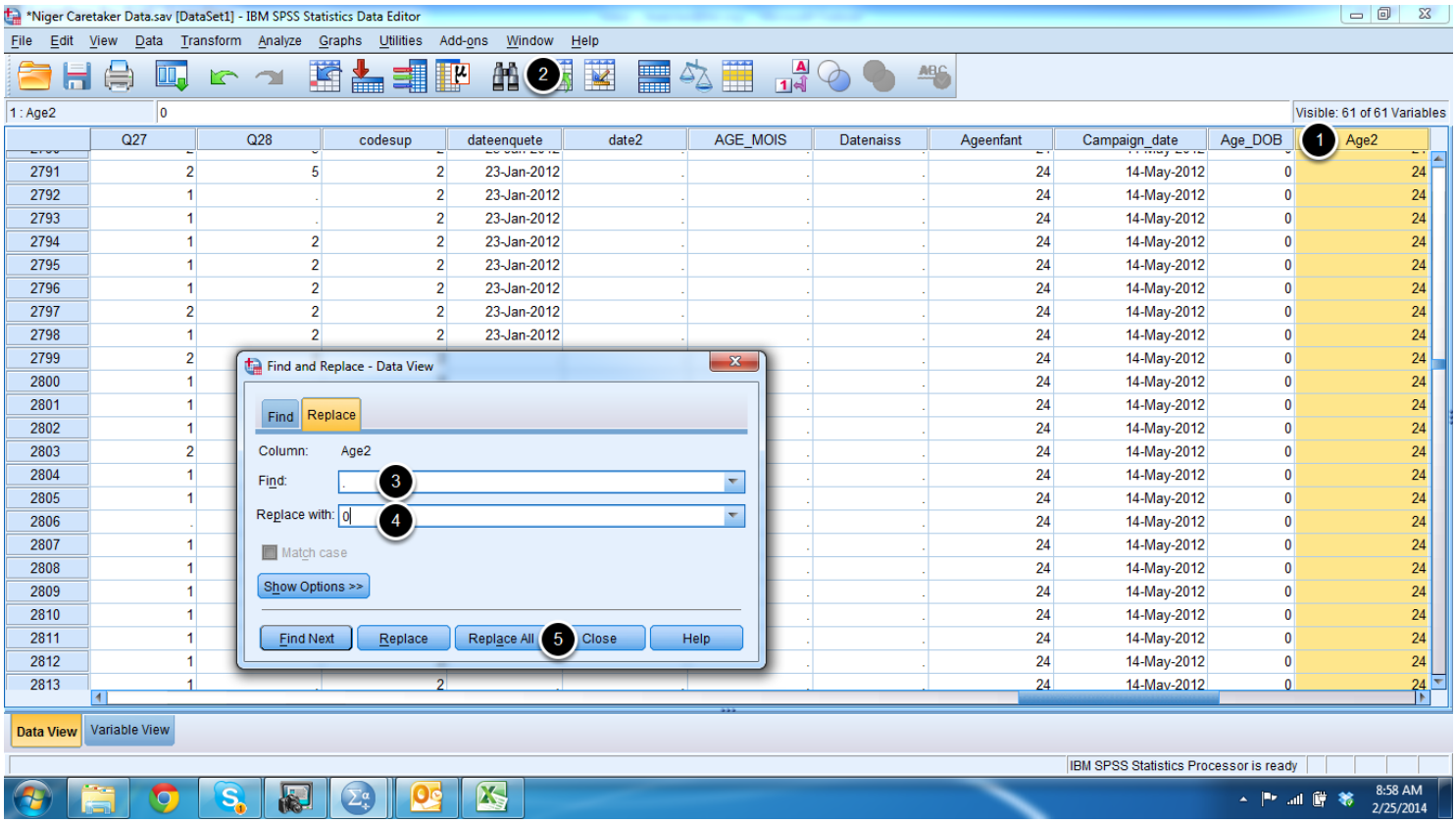


In the Target Variable box (#1), enter the name you want to give the new variable. In the Type and Label column, select the variable for age in months (#2) given for children whose date of birth is not known. Click the arrow (#3) so that the variable moves to the Numeric Expression box. Click on "If" (#4) and select "Include if case satifies condition" (#5). Select the variable that you created for age in months calculated from date of birth (#6) and click on the arrow (#7). Then type "=0" in the text box after the variable name (#8) and press continue (#9) and then OK (#10).



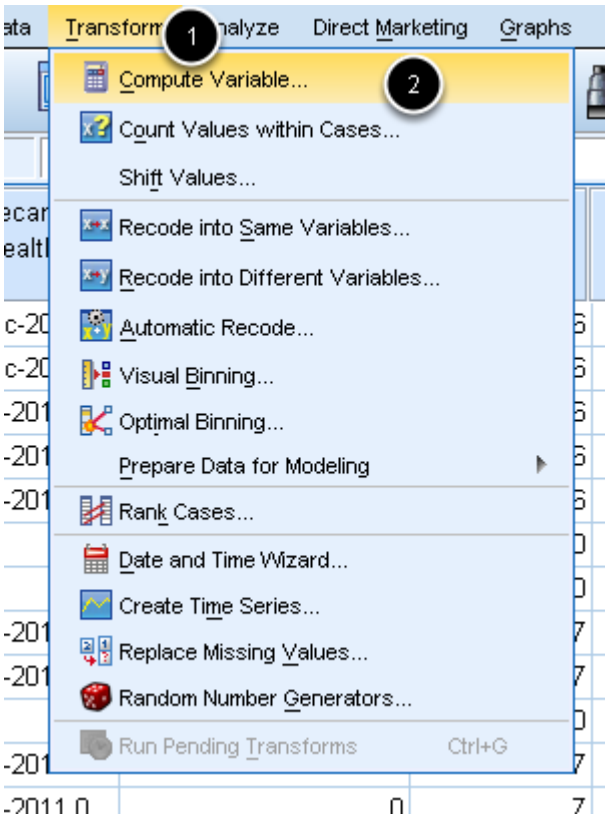
Click on the variable view tab and scroll to the bottom to view the new variable that was created. Enter 0 in the Decimals column for the new variable.

Recode missing data as 0

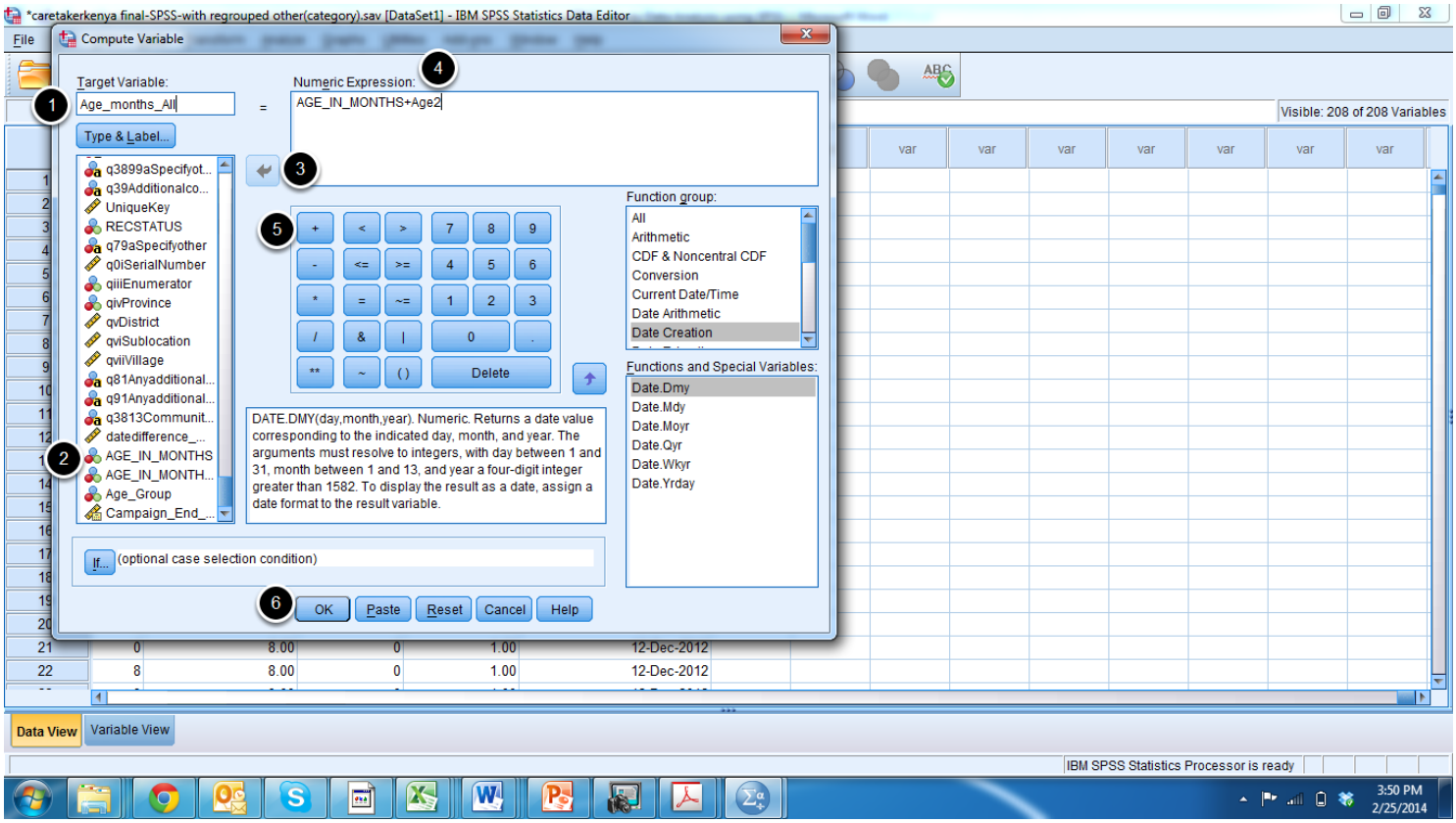


Select the column for the new variable for age in months (#1). Click on the Find icon (#2) to open the Find and Replace window. In the find section (#3) type . (dot) as shown in the photo above. In the Replace With section (#4), type 0 (zero) and click Replace All (#5). Using this procedure, all cases where the child's date of birth is known will be replaced a with 0 (zero).

Compute a single age variable

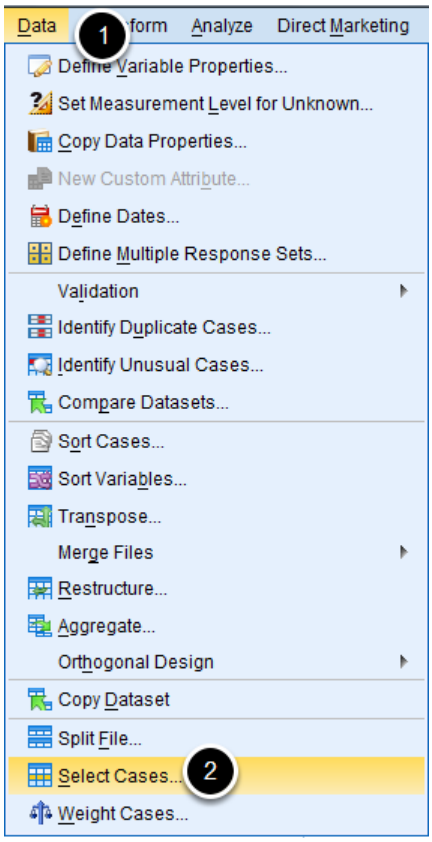


In order to analyze coverage by age, a single variable needs to be created that contains the age of the child in months. Click on Transform (#1) and then select Compute Variable (#2).

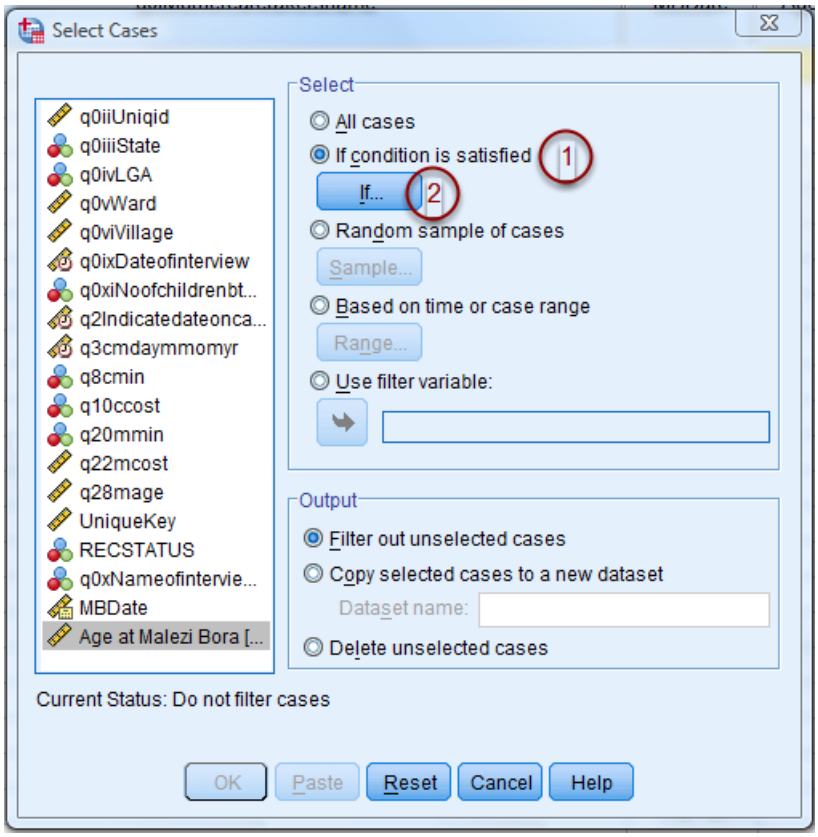


In the Target Variable box (#1), enter the name of the final variable you want to create for age in months. Select the first variable for age in months that you created (for children with a known date of birth) from the Type and Label column (#2) and move it using the arrow (#3) to the Numeric Expression section (#4). Click on the addition sign (#5), below the Numeric Expression box and go back to the Type & Label column to select the second variable for age in months that you created (for children with an unknown date of birth) and send it to #4. The numeric expression box should now contain an equation for the sum of two ages that were calculated. Click OK (#6) to create a variable for age in months for all children.

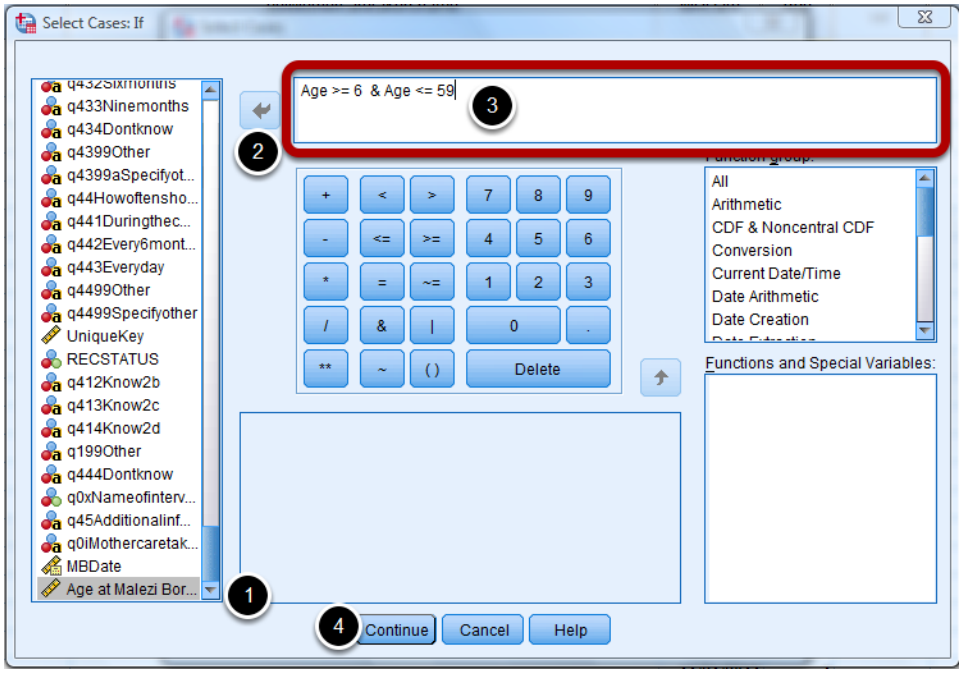
Exclude surveys of children who were not 6-59 months at the time of the supplement event



First, go to the Data menu (#1) and click "Select Cases" (#2).



When the Select Cases window opens, select the option "If condition is satisfied" (#1) and left click the blue "If..." button below (#2).

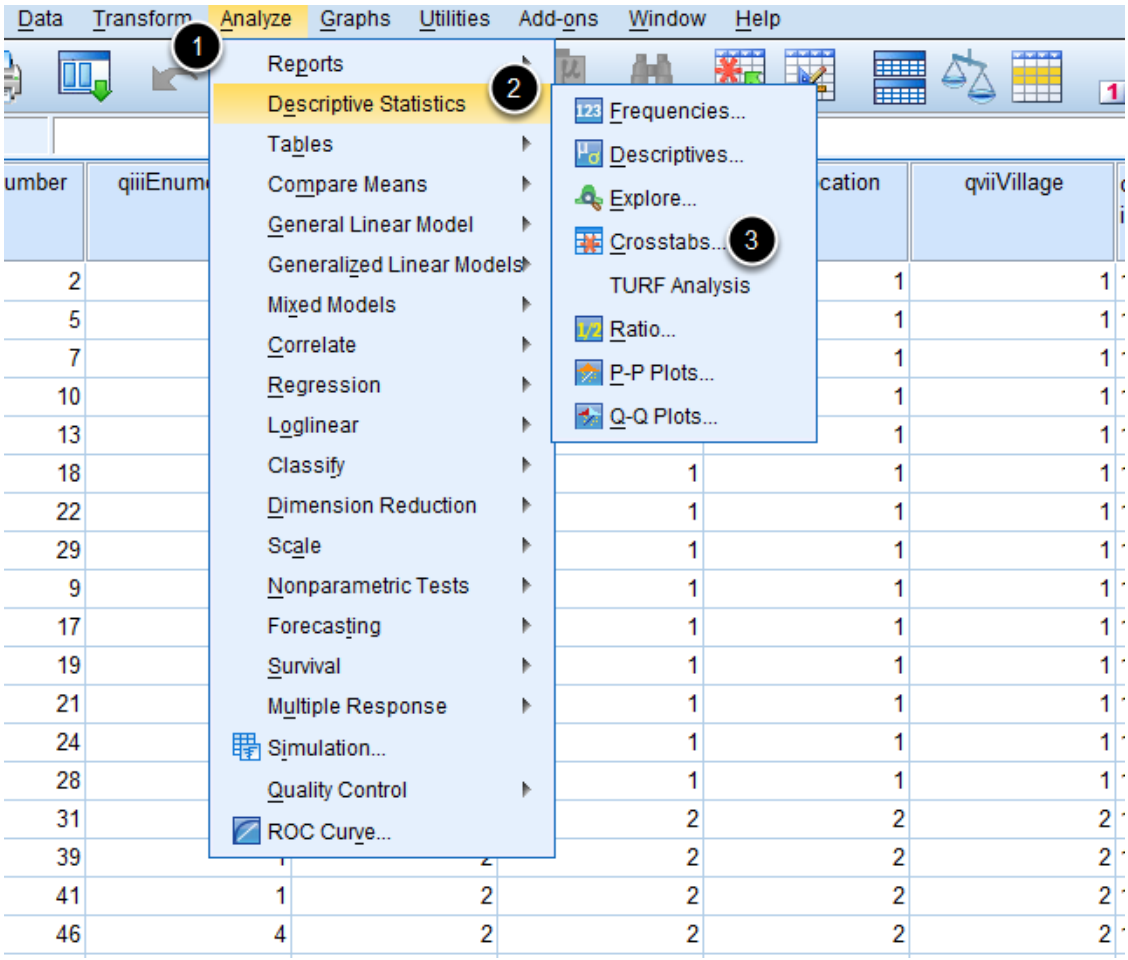


The Select Cases If window will appear where you can specify what cases you want to select. In this case, we want to only include children who were between 6 and 59 months of age at the time of the supplementation event. First select the variable that corresponds to the child's age at the

time of the supplementation event (#1). This is the variable we created in previous step. Then click the right arrow (#2) . Enter the conditions for eligibility as shown above, for example, if age >=6 and age <=59.

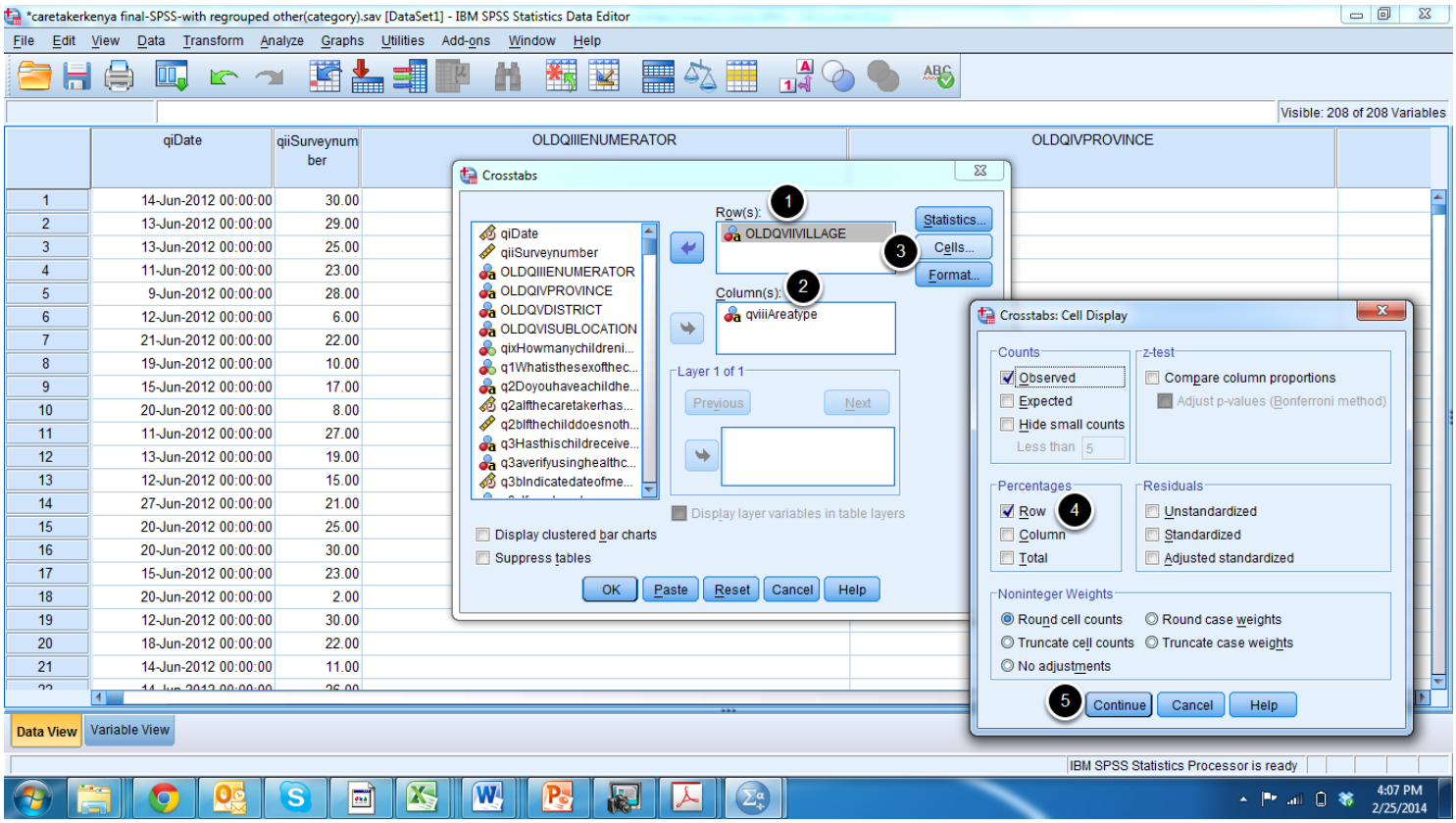
**Note: Using the survey end date to calculate age at the time of the supplementation event ensures that all children who were six months at the time of the campaign will be included in the analysis. However, it is possible that children at the upper end of the age range could be excluded. It is important to check that children listed as having 60 months were not mistakenly excluded.

Data Validation - Cross Variables



It is important to check that your data makes sense before embarking on analysis. To begin with, it is advisable to check how your sample size is distributed among the smallest units (in this case clusters). Go to Analyze (#1), then click on Descriptive Statistics (#2) and finally Crosstabs (#3).

Crosstab Variables



Crosstabs allows you to view the responses to two questions side by side to make sure the responses make sense. For example, we can view the Area Type selected by each respondent broken down by cluster. In most cases, a rural village should not have been marked as urban and vice versa. In cases where both area types are selected for a cluster, the reason for the discrepancy should be investigated. To do the crosstabs calculation, select one variable (e.g. cluster) and move it to the Rows box (#1), then select the other variable (e.g. Area type) move it to the Columns box (#2). Click on Cells (#3) to open the Cell Display window. Here click on "Row" to display the percent who gave each possible response. Click on Continue (#5) and finally OK (#6) to run the analysis.

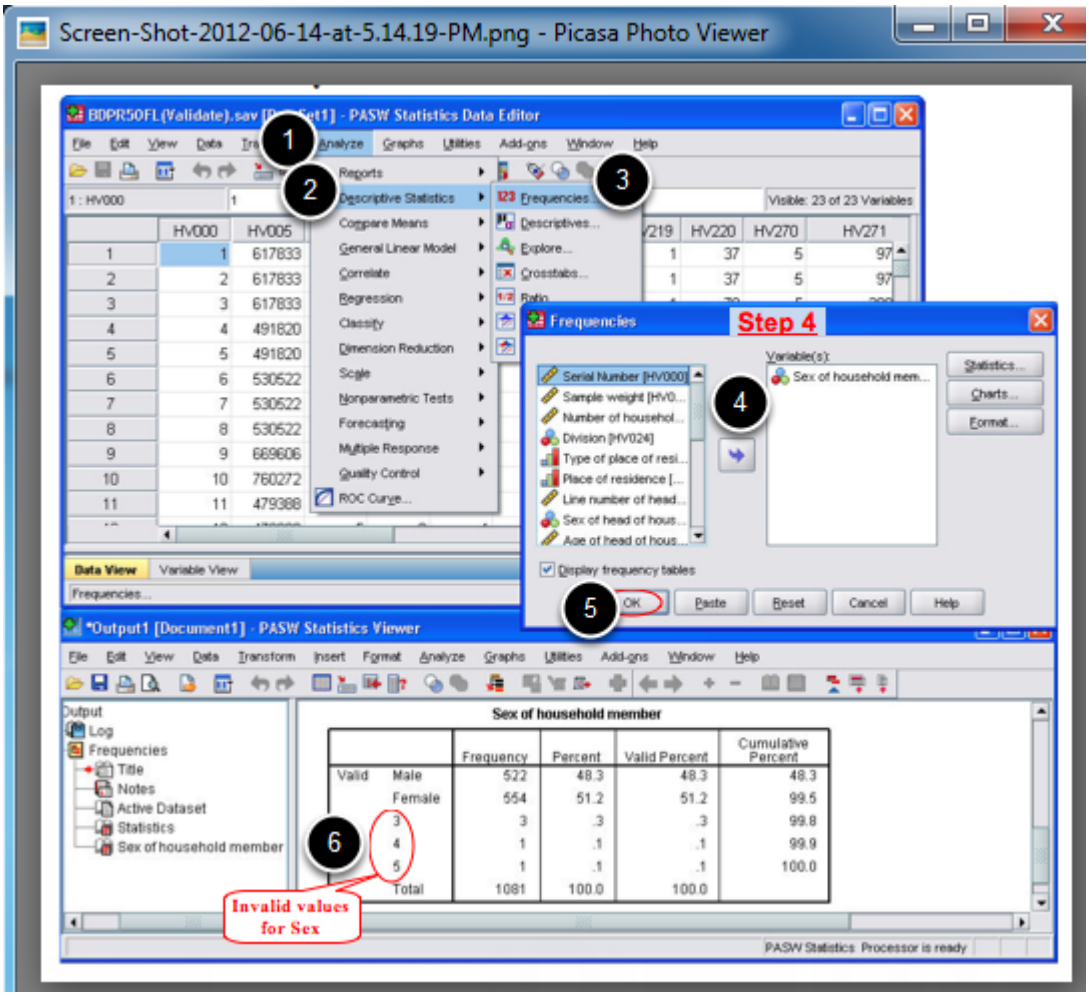
Output

qviiVillage * qviiiAreatype Crosstabulation

			qviiiAreatype		Total
			1	2	
6	Count	1	27	0	28
	% within qviiiAreatype	100.0%	3.8%	0.0%	3.1%
7	Count	0	2	0	2
	% within qviiiAreatype	0.0%	0.3%	0.0%	0.2%
15	Count	0	0	30	30
	% within qviiiAreatype	0.0%	0.0%	16.4%	3.3%
16	Count	0	1	30	31
	% within qviiiAreatype	0.0%	0.1%	16.4%	3.4%
17	Count	0	29	0	29
	% within qviiiAreatype	0.0%	4.1%	0.0%	3.2%
18	Count	0	27	3	30
	% within qviiiAreatype	0.0%	3.8%	1.6%	3.3%
23	1	715	183	899	
	100.0%	100.0%	100.0%	100.0%	
Total					

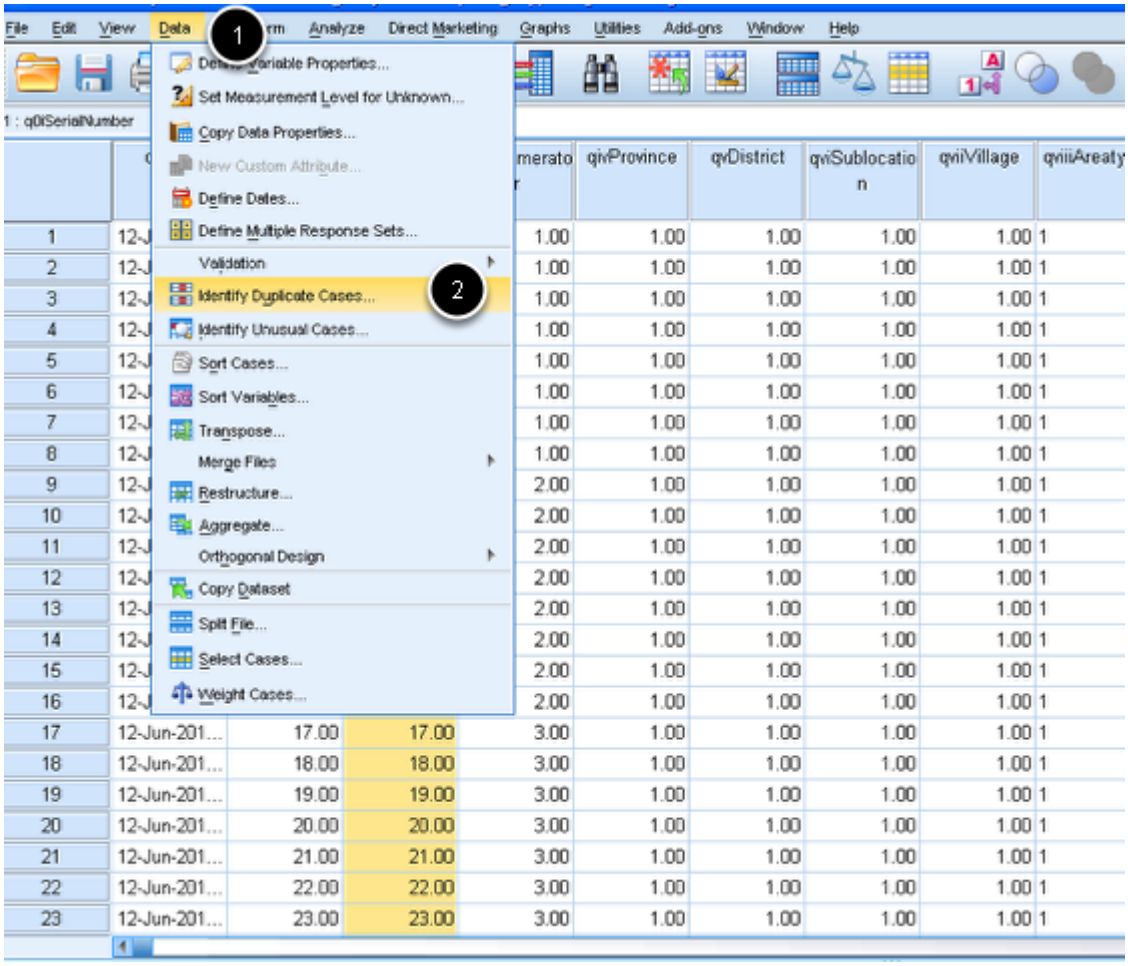
In this output you will identify a number of mistakes. Column #1 shows missing information on area type from cluster 6 (#2). #3 and #4 indicate clusters in which area type responses are inconsistent. These entries should be given a second look to see if revisions are needed.

Validation Check - Incorrect Responses

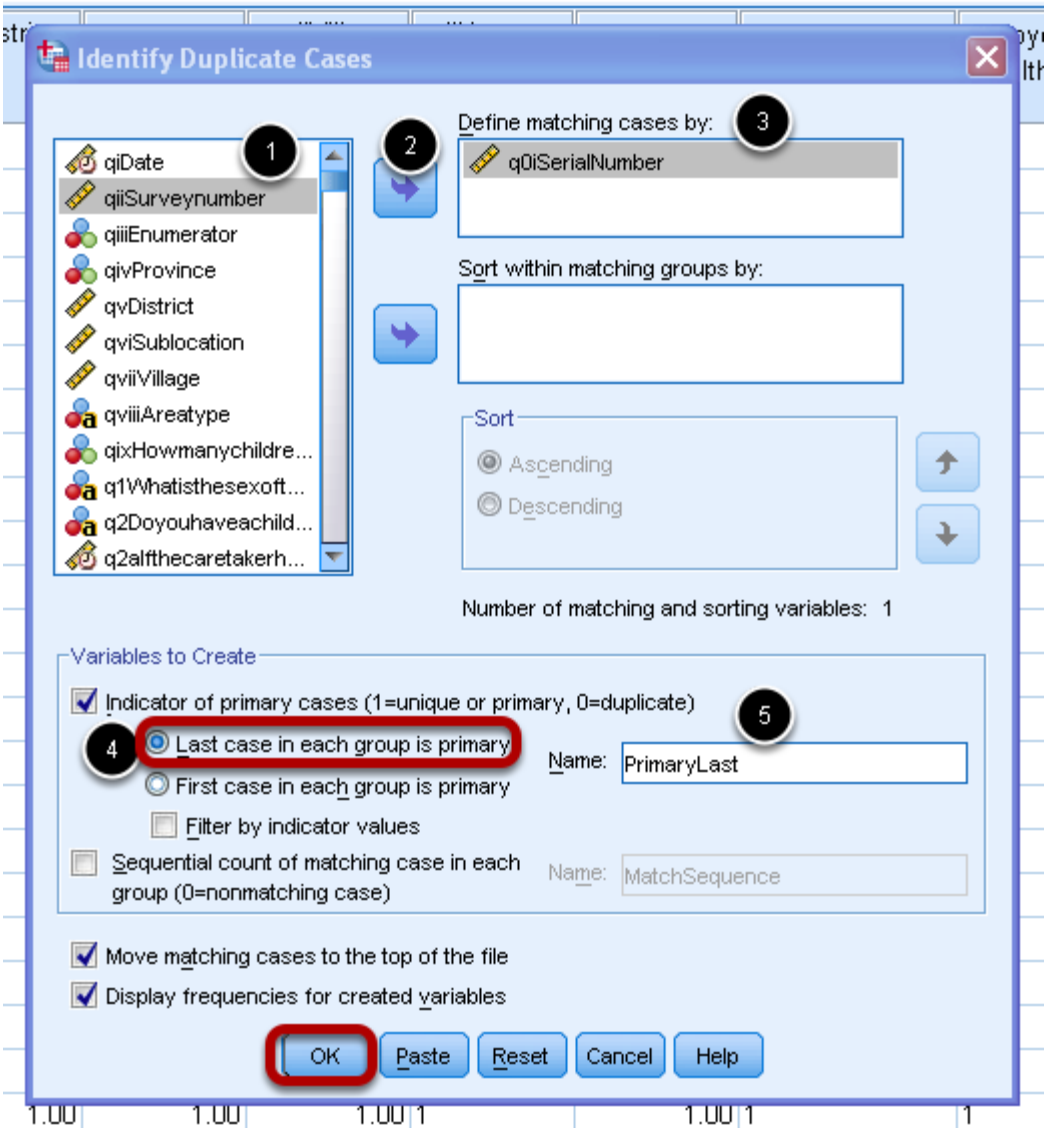


In some cases, incorrect entries can be determined by running a simple frequency analysis. If you want to investigate if out of range responses appeared in any question, it is recommended to run descriptive statistics. To do this click on Analyze (#1) and then Descriptive Statistics (#2). Select Frequencies (#3) and a Frequency window appears. Select all the variables to investigate and click on the arrow to move them to the Variable(s) box (#4). Click OK (#5) and go to output window to see the frequency of responses to each variable (#6). In the example above, the question asks the gender of the child, however there are five different options given whereas only two options (male and female) were expected. In this case you should go back to the original entry and try to determine the correct answer.

Data Cleaning - Identifying Duplicate Surveys



Each survey should have a unique identifier (survey number) in order to identify duplicates. If there are two surveys with the same number, then the reason for this needs to be addressed. Perhaps the same data was entered twice, or there was a data entry error and the wrong survey number was entered. To search for duplicates, go to Data (#1) and select "Identify Duplicate Cases" (#2).



Awindow for identifying duplicate cases will appear. In the first column, select the variable name for the unique identifier (#1), and click the blue arrow (#2) to move the variable to the "Define matching cases by" box (#3). By default, in the Variables to create box, "Last case in each groups is primary" (#4) will be selected. This means that if a duplicate is found, the first entry will be treated as a duplicate and the second/last entry will be treated as a duplicate. #5 indicates that a new variable will be created and the variable name will be PimaryLast. Click OK to view if there are any duplicates.

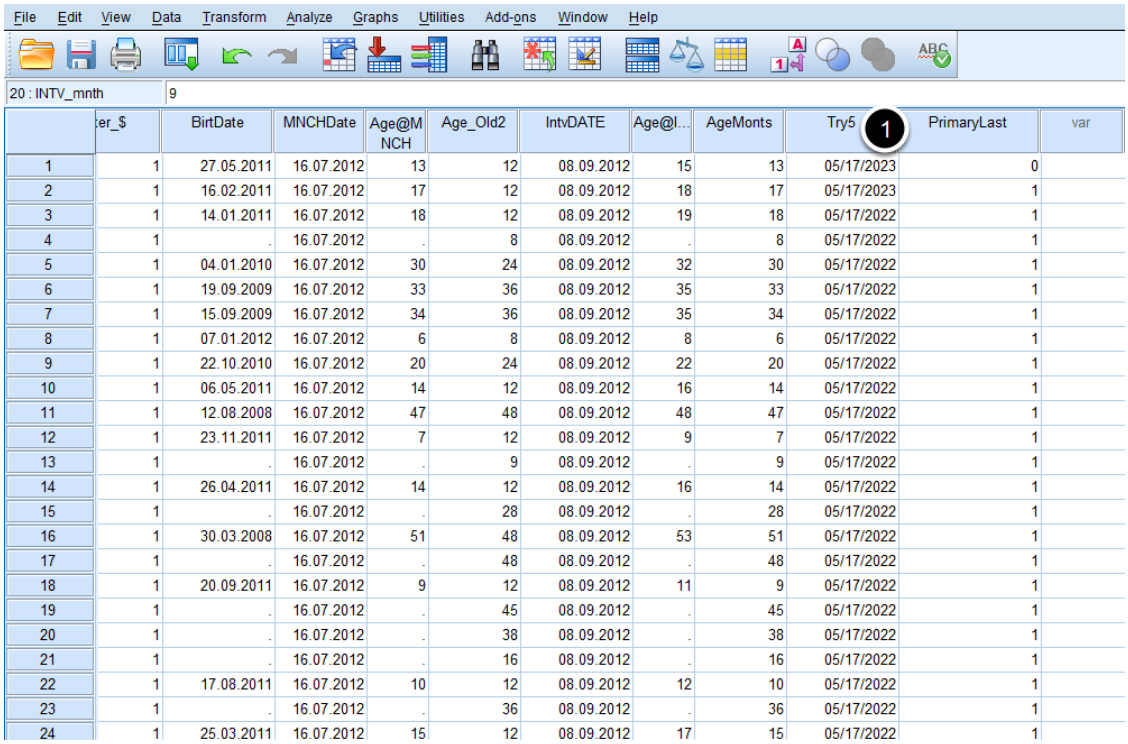
Output window - Duplicates

Indicator of each last matching case as Primary

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Duplicate Case	1	.1	.1	.1
	Primary Case	899	99.9	99.9	100.0
	Total	900	100.0	100.0	

In the output window, you will have a table showing the number of duplicates identified (#1). In this case, there is one duplicate (#2).

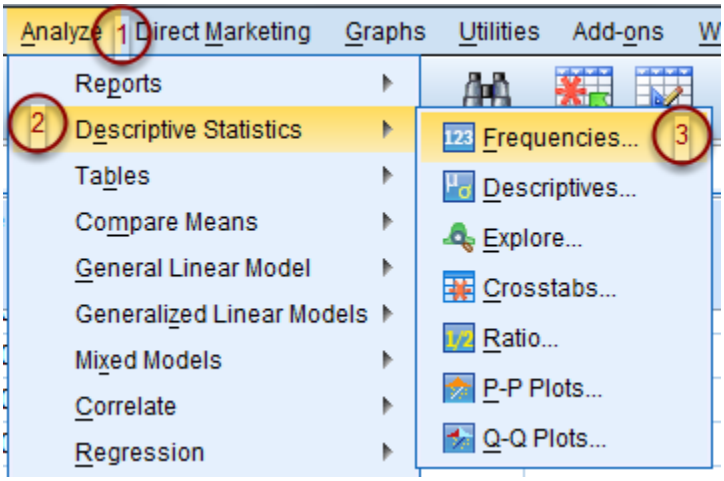
Deleting the duplicates



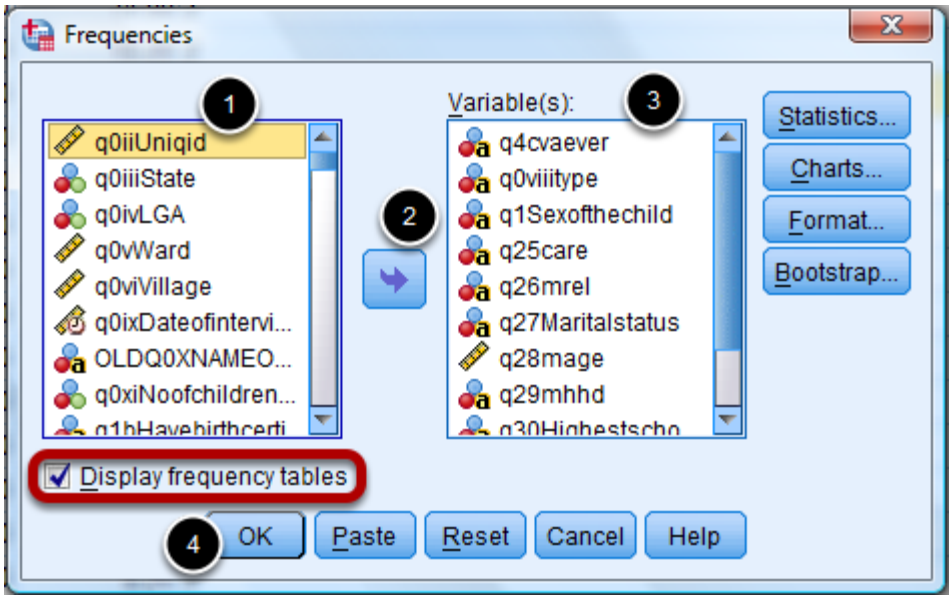
	er_$	BirtDate	MNCHDate	Age@MNCH	Age_Old2	IntvDATE	Age@I...	AgeMonts	Try5	PrimaryLast	var
1	1	27.05.2011	16.07.2012	13	12	08.09.2012	15	13	05/17/2023	0	
2	1	16.02.2011	16.07.2012	17	12	08.09.2012	18	17	05/17/2023	1	
3	1	14.01.2011	16.07.2012	18	12	08.09.2012	19	18	05/17/2022	1	
4	1	.	16.07.2012	.	8	08.09.2012	.	8	05/17/2022	1	
5	1	04.01.2010	16.07.2012	30	24	08.09.2012	32	30	05/17/2022	1	
6	1	19.09.2009	16.07.2012	33	36	08.09.2012	35	33	05/17/2022	1	
7	1	15.09.2009	16.07.2012	34	36	08.09.2012	35	34	05/17/2022	1	
8	1	07.01.2012	16.07.2012	6	8	08.09.2012	8	6	05/17/2022	1	
9	1	22.10.2010	16.07.2012	20	24	08.09.2012	22	20	05/17/2022	1	
10	1	06.05.2011	16.07.2012	14	12	08.09.2012	16	14	05/17/2022	1	
11	1	12.08.2008	16.07.2012	47	48	08.09.2012	48	47	05/17/2022	1	
12	1	23.11.2011	16.07.2012	7	12	08.09.2012	9	7	05/17/2022	1	
13	1	.	16.07.2012	.	9	08.09.2012	.	9	05/17/2022	1	
14	1	26.04.2011	16.07.2012	14	12	08.09.2012	16	14	05/17/2022	1	
15	1	.	16.07.2012	.	28	08.09.2012	.	28	05/17/2022	1	
16	1	30.03.2008	16.07.2012	51	48	08.09.2012	53	51	05/17/2022	1	
17	1	.	16.07.2012	.	48	08.09.2012	.	48	05/17/2022	1	
18	1	20.09.2011	16.07.2012	9	12	08.09.2012	11	9	05/17/2022	1	
19	1	.	16.07.2012	.	45	08.09.2012	.	45	05/17/2022	1	
20	1	.	16.07.2012	.	38	08.09.2012	.	38	05/17/2022	1	
21	1	.	16.07.2012	.	16	08.09.2012	.	16	05/17/2022	1	
22	1	17.08.2011	16.07.2012	10	12	08.09.2012	12	10	05/17/2022	1	
23	1	.	16.07.2012	.	36	08.09.2012	.	36	05/17/2022	1	
24	1	25.03.2011	16.07.2012	15	12	08.09.2012	17	15	05/17/2022	1	

Once you know how many duplicates you have, it is easy to locate them using the new variable that was created (PrimaryLast) (#1). By default, the duplicates will be listed first. It is recommended to go back to the surveys to confirm if two surveys were given the same number by mistake or if one record was entered twice. Delete one record if it is a duplicate.

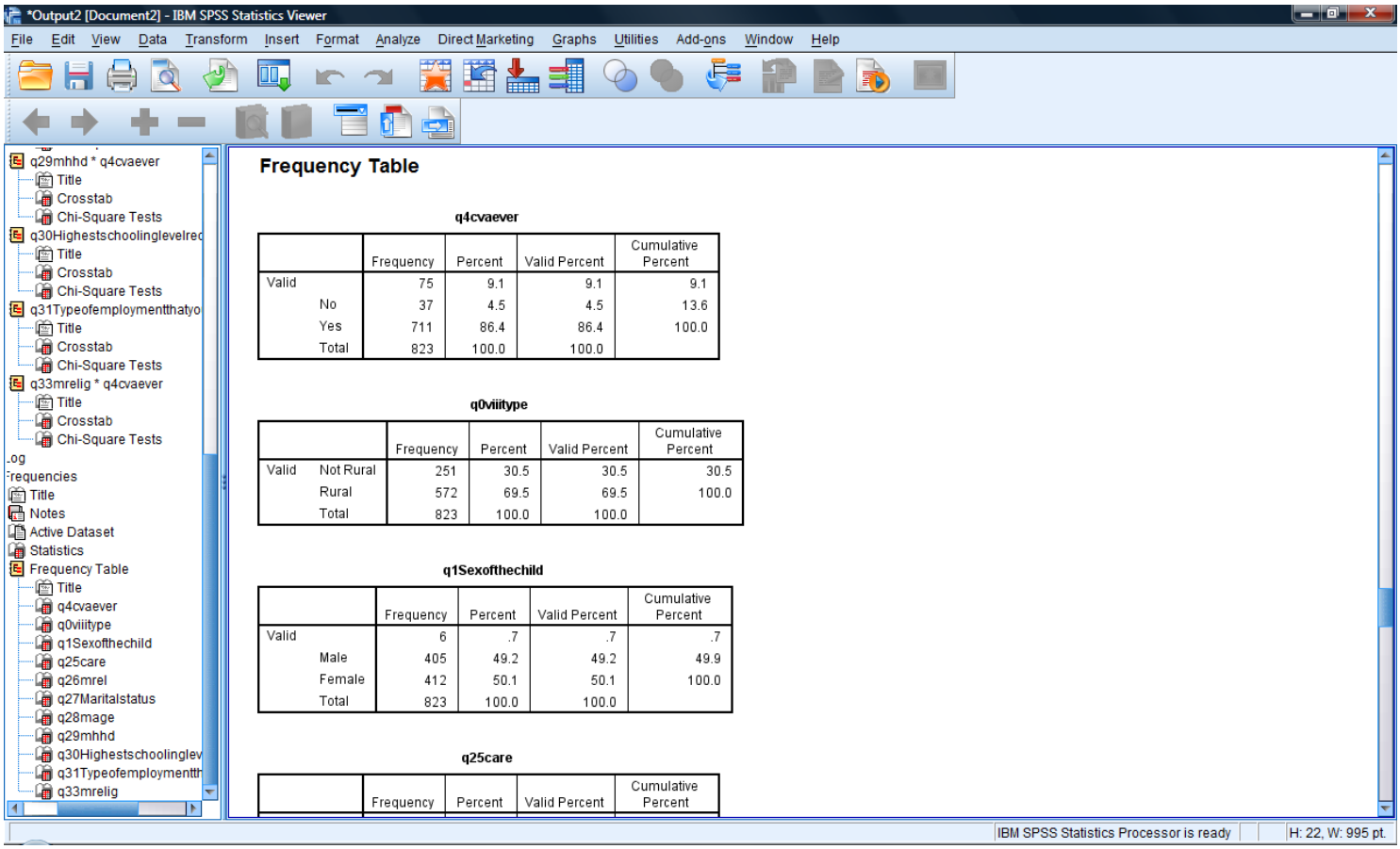
Data Analysis



To view frequencies and percentages for each variable, on the menu bar select "Analyze"(#1). From the drop down menu select "Descriptive Statistics" (#2) then "Frequencies" (#3).



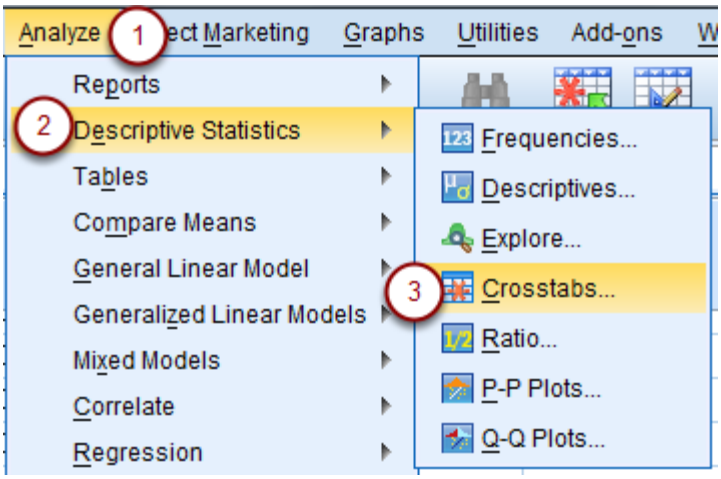
When the Frequencies window opens, select all of the variables you would like to calculate frequencies for from the list on the left (#1). To select multiple variables at one time press and hold "CTRL" while selecting each. Once selected, click the blue arrow between the two lists (#2). All of the variables selected will then appear in the list on the right (#3). Ensure the "Display frequency tables" box is checked and click "OK" (#4).



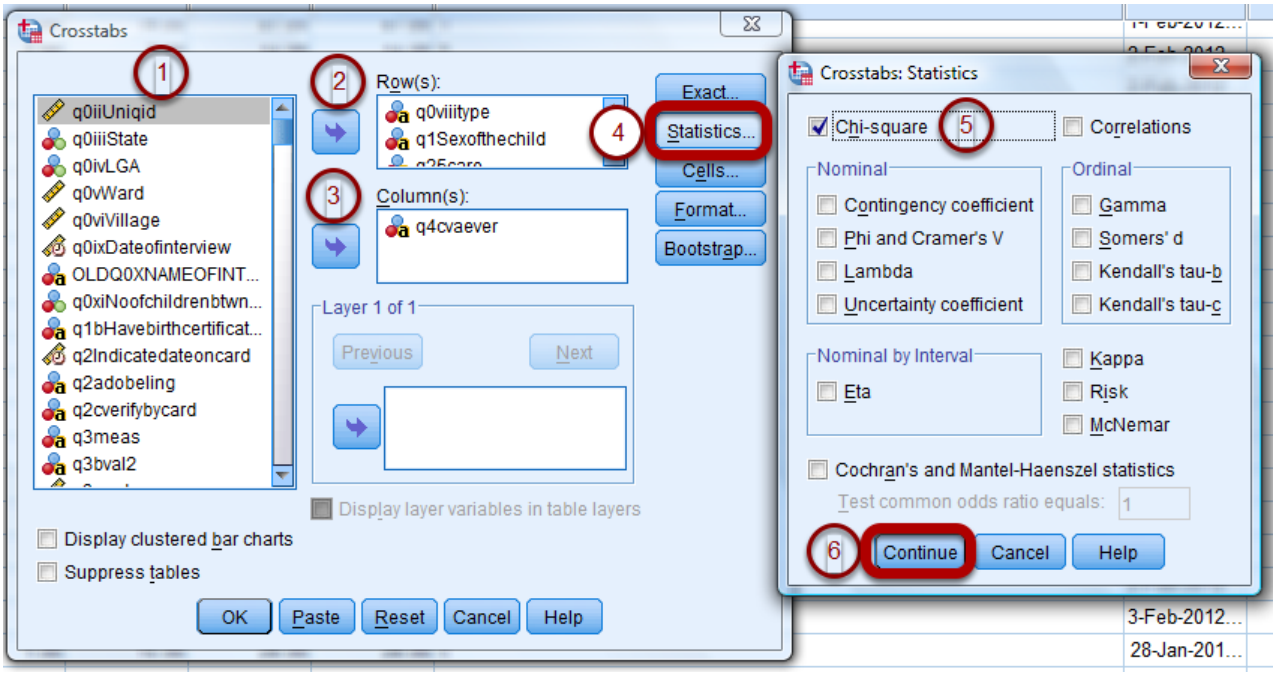
After selecting "OK" the frequency tables can be viewed in the Output window.

Data Analysis: Chi Square Calculation

In this section we will examine differences between participants whose child received vitamin A and those who did not. The chi-square test and respective p-value will determine whether there are significant differences between Vitamin Asupplemented and non vitamin Asupplemented individuals by given variables.

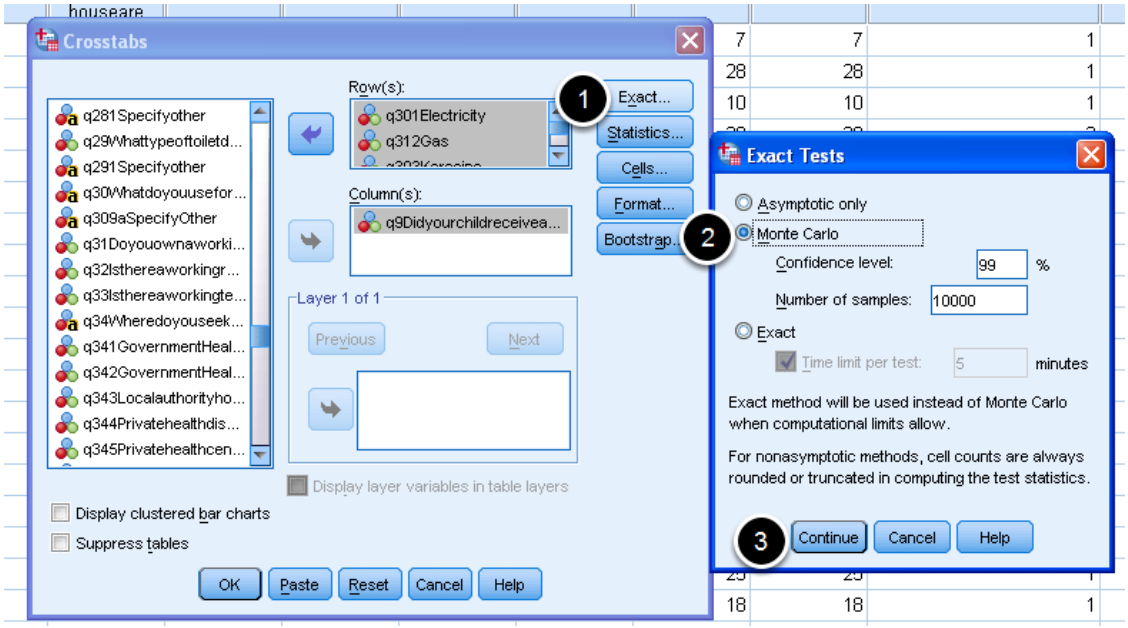


Left click "Analyze" (#1) at the top of the screen. From the drop down menu select "Descriptive Statistics" (#2) then select "Crosstabs" (#3)

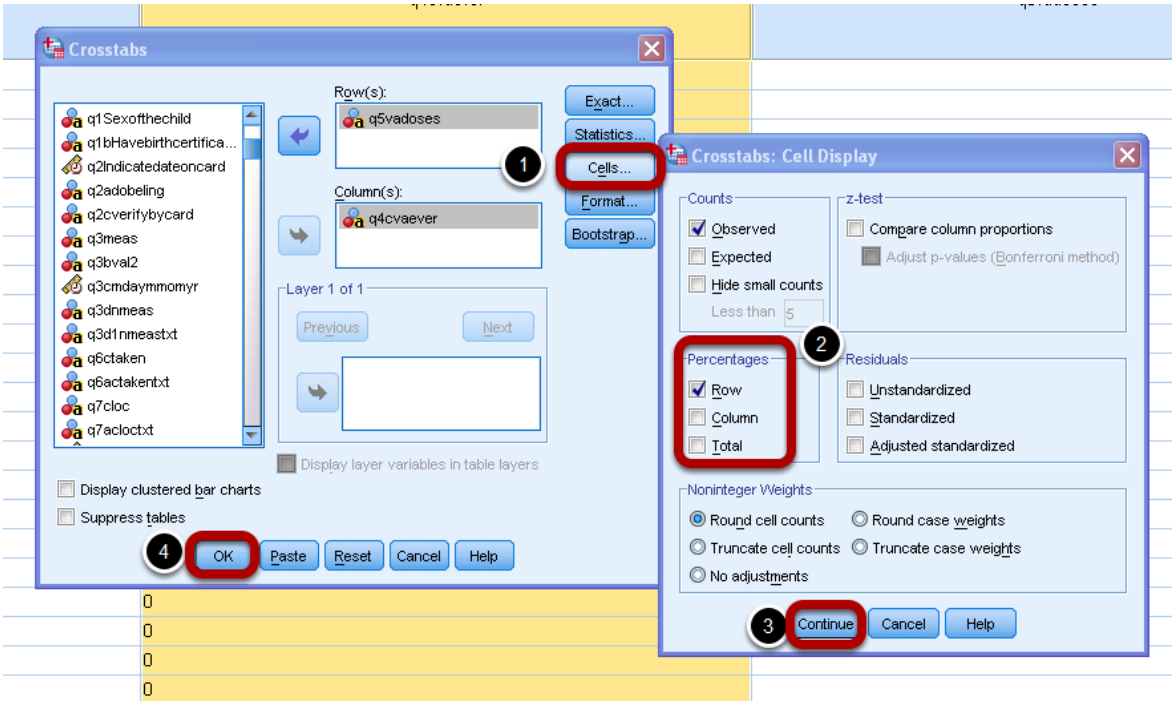


When the Crosstabs window appears, select from the list on the left (#1) all of the variables you would like to examine differences for between supplemented and non-supplemented children. Once the variables are selected, click the blue arrow (#2) to add them to the Row(s) box. For the box labeled "Column(s), select the variable that corresponds to whether vitamin Awas recieved

and click the blue arrow (#3) to the right of the box. Then select the "Statistics" button (#4). In the Crosstabs:Statistics window select Chi-square by clicking the check box to its left (#5) and then press "Continue" (#6).

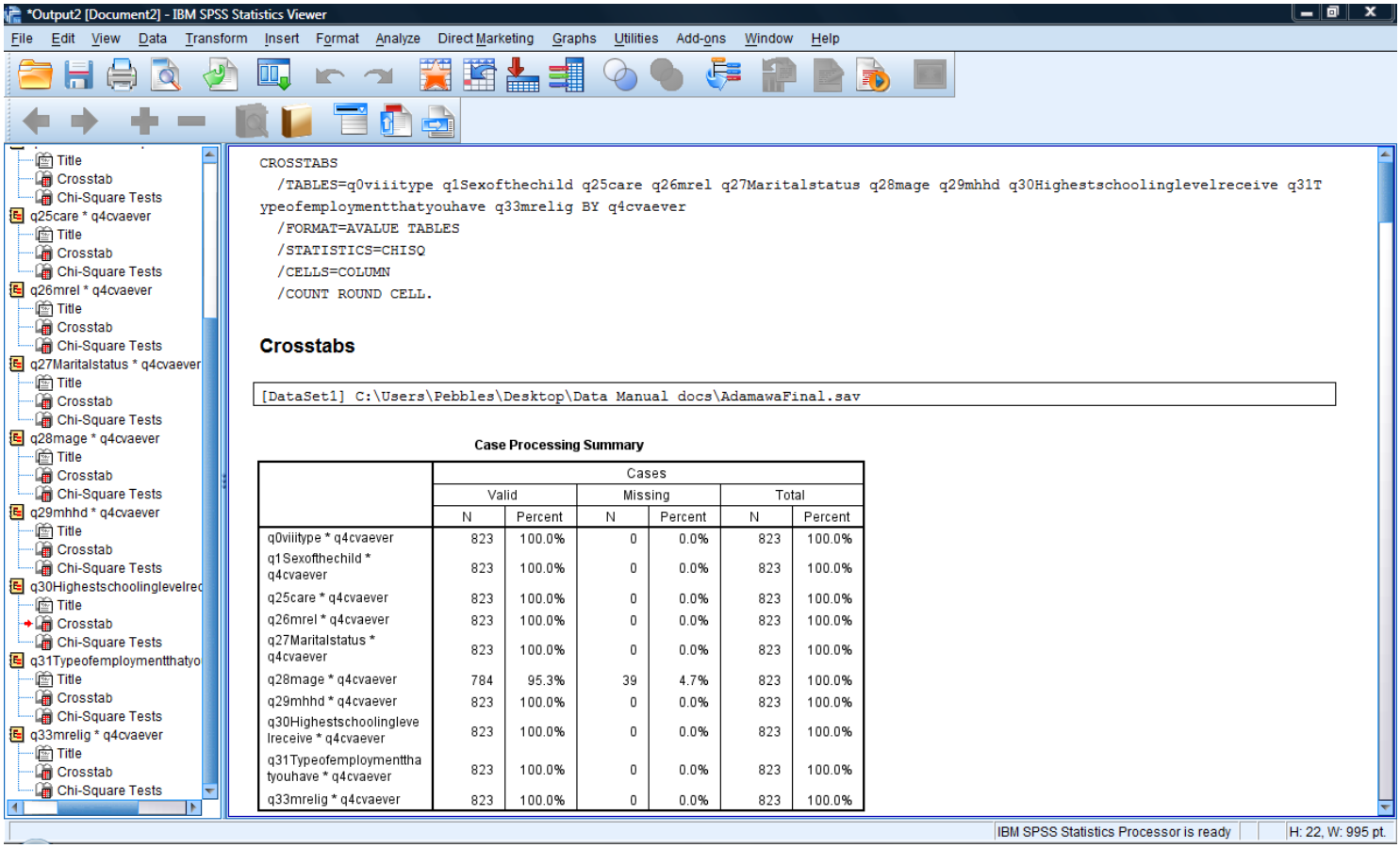


Left click the "Exact" button (#1) and then select Monte Carlo (#2) and press "Continue" (#3). This will give the results of the Fisher's exact test in the output.



To compare the various responses it is important to have their percentages and frequencies. To view the frequencies and percentages for each response, click on "Cells" (#1) and then in the Crosstabs: Cell Display window, select "Row" (#2). Click continue (#3). Once this has been

completed, select "OK" (#4) to run the analysis.



When the crosstabs procedure is complete, the results will be available to view in the output window. A p-value less than or equal to 0.05 can be interpreted as a significant difference while p-values higher than 0.05 can be considered not significant.

Chi-Square Output

1

qviiiAreatype * q9Didyourchildreceiveavitaminacapsule Crosstabulation

			q9Didyourchildreceiveavitaminacapsule		
			0	1	Total
qviiiAreatype	1	Count	465	226	691
		% within qviiiAreatype	67.3%	32.7%	100.0%
	2	Count	126	53	179
		% within qviiiAreatype	70.4%	29.6%	100.0%
Total		Count	591	279	870
		% within qviiiAreatype	67.9%	32.1%	100.0%

2

Chi-Square Tests[c]

4

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.626[a]	1	.429	.473	.243
Continuity Correction[b]	.492	1	.483		
Likelihood Ratio	.633	1	.426	.473	.243
Fisher's Exact Test				.473	.243
N of Valid Cases	870				

3

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 57.40.

b. Computed only for a 2x2 table

c. For 2x2 crosstabulation, exact results are provided instead of Monte Carlo results.

Two tables will appear in the output window. Table 1 shows how the responses were distributed between those who were supplemented and those who were not. Table 2 shows the association between vitamin Asupplementation and each variable (In this case, area type). Below table two, you will see some notes about the data. If the first point (#3) indicates that there are 0 cells (0.0%) which have expected counts less than 5, then it is ok to use Pearson Chi-Square to determine the significance of the two variables. The significance level is indicated in the column titled Asymp.sig. In this example, since the Asymp. Sig value is more than 0.05, so there is no significant difference between groups.

Fishers Exact Test

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Monte Carlo Sig. (2-sided) Sig.	99% Confidence Interval Lower Bound	99% Confidence Interval Upper Bound
Pearson Chi-Square	5.961[a]	2	.051	.032[b]	.028	.037
Likelihood Ratio	6.094	2	.047	.035[b]	.030	.039
Fisher's Exact Test	5.793			.032[b]	.028	.037
N of Valid Cases	870					

a. 2 cells (33.3%) have expected count less than 5. The minimum expected count is .32.

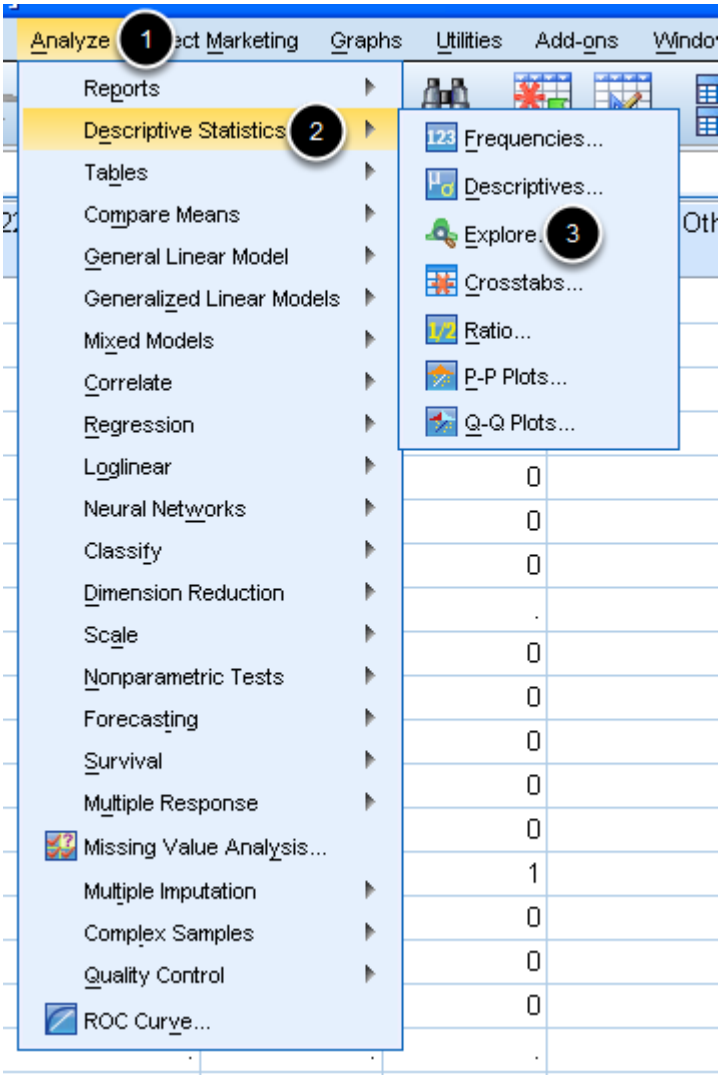
b. Based on 10000 sampled tables with starting seed 1487459085.

If the notes below the chi-square table indicate that some cells have expected count of less than five (#1), then Fisher's exact (#2) should be used as the test statistic for these variables.

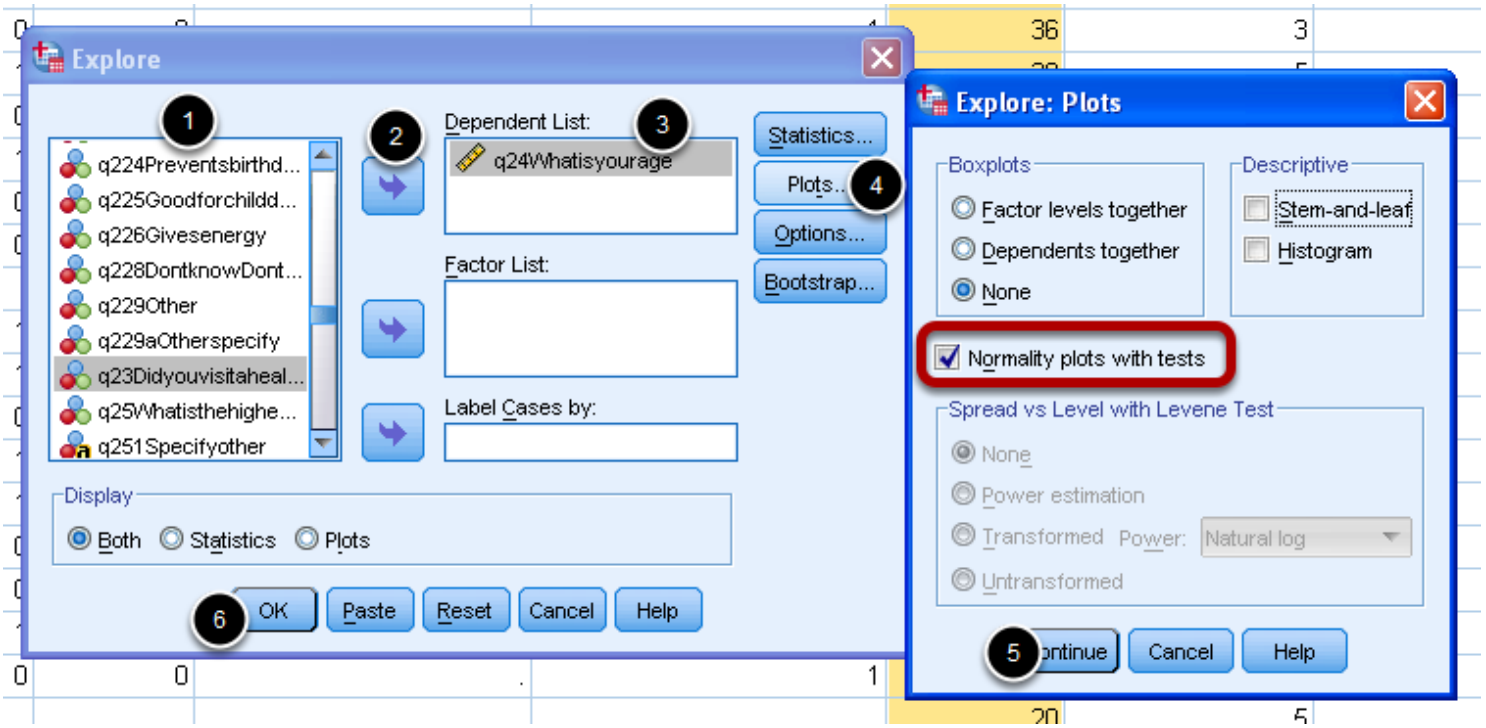
Data Analysis: Comparison of Means

Acomparison of means can be done to determine if there are differences in numeric values, such as cost to attend the event, or distance to travel to the event. In order to determine which test to use, it first needs to be determined if the data are normally distributed.

Testing if data are normally distributed



Click on Analyze (#1), then select Descriptive Statistics (#2). Finally select Explore (#3).



In the Explore window, select the continuous variable you want to test for normality from first column on the left (#1). Click the blue arrow (#2) to move it to the Dependent List box (#3). Select Plots (#4), to go to the plots window. Ensure that "Normality plots with tests" is selected. Click "Continue" (#5) and finally click "OK" (#6).

Output window

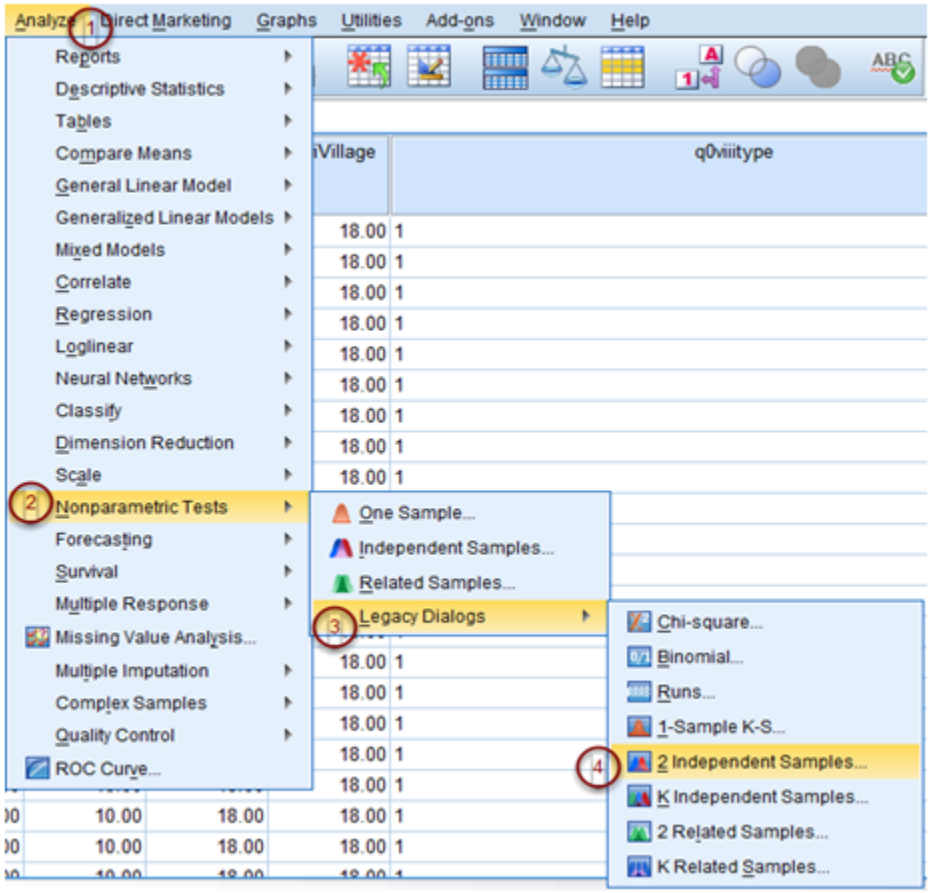
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
q24Whatisyourage	.148	861	.000	.864	861	.000

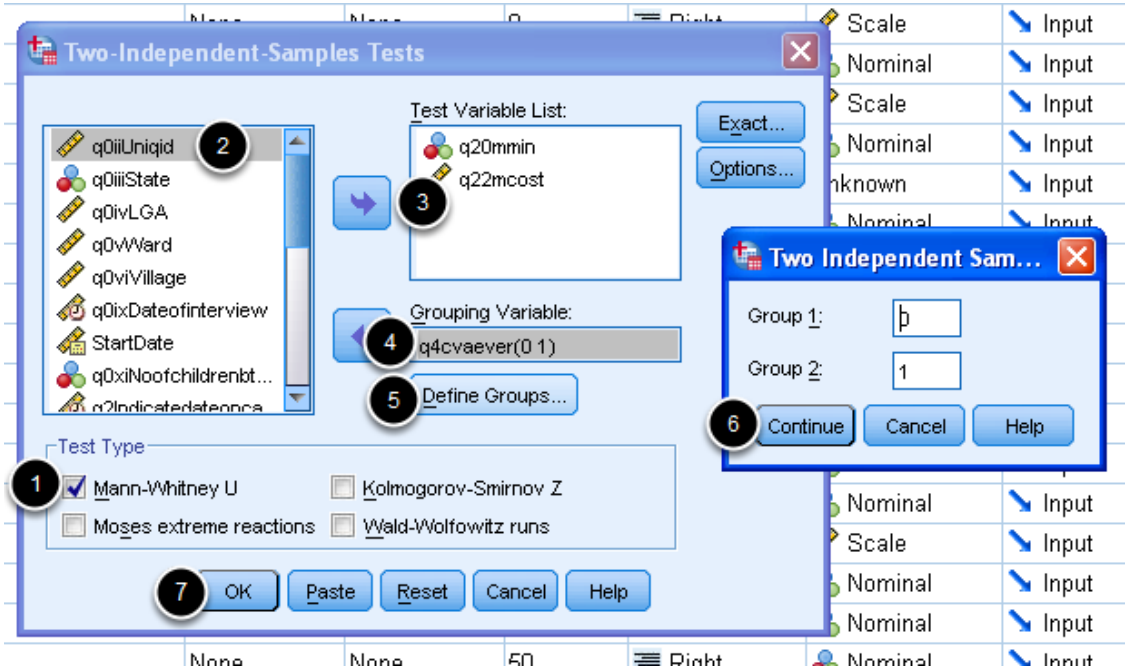
a. Lilliefors Significance Correction

The Shapiro-Wilk test should be used to determine if the data are normally distributed. If the Sig. value (p-value) (#1) is < 0.05 the data are not normally distributed, while if the Sig. value is > 0.05 the data are normally distributed. Since we will be comparing two groups (supplemented vs. not supplemented) the two-sample t-test should be used for if the data are normally distributed (Sig. ≥ 0.05) and the Mann -Whitney test should be used if data are not normally distributed (Sig. < 0.05).

Wilcoxon-Man-Whitney Test



To perform the Wilcoxon-Mann-Whitney test, first click "Analyze" (#1) and select "Nonparametric Test" (#2) from the drop down menu. From the available nonparametric test options, select "Legacy Dialogs" (#3) and choose "2 Independent Samples..." (#4) as the final option.



When the Two Independent Samples Test window appears, ensure the Mann-Whitney U box in the Test Type area is checked (#1). Select from the list on the left (#2) the test variable(s) you want to analyze (example: cost of transport) and use the blue arrow next to the test variable list to add them to the list (#3) . In the grouping variable box, add the variable that corresponds to your dependent variable (e.g. whether or not the child received vitamin A) (#4). Next select "Define Groups..." (#5). In the Two Independent Samples box, specify the values for the two groups you would like to compare (for example 0 and 1 for those not supplemented and those who were supplemented) (#6). Please note that when defining groups, only two groups are allowed and the values should be numeric. Once this is done select Continue and then "OK" (#7)

Mann-Whitney Test Output

Mann-Whitney Test

Ranks

	q4cvaever	N	Mean Rank	Sum of Ranks
q20mmin	0	163	85.29	13903.00
	1	7	90.29	632.00
	Total	170		
q22mcost	0	146	76.79	11211.50
	1	7	81.36	569.50
	Total	153		

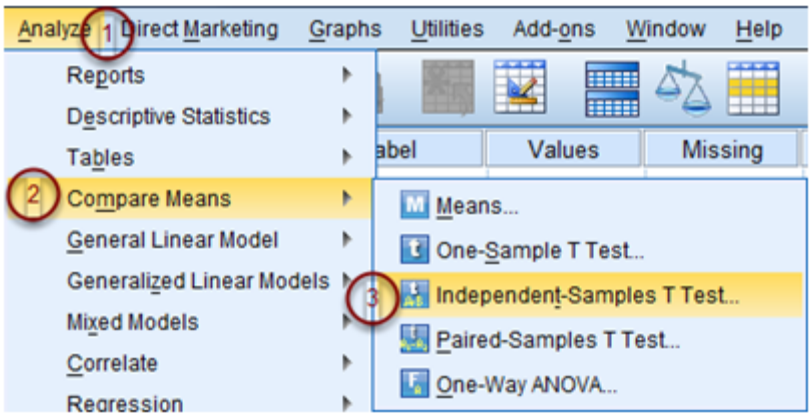
Test Statistics[a]

	q20mmin	q22mcost
Mann-Whitney U	537.000	480.500
Wilcoxon W	13903.000	11211.500
Z	-.268	-.269
Asymp. Sig. (2-tailed)	.789	.788

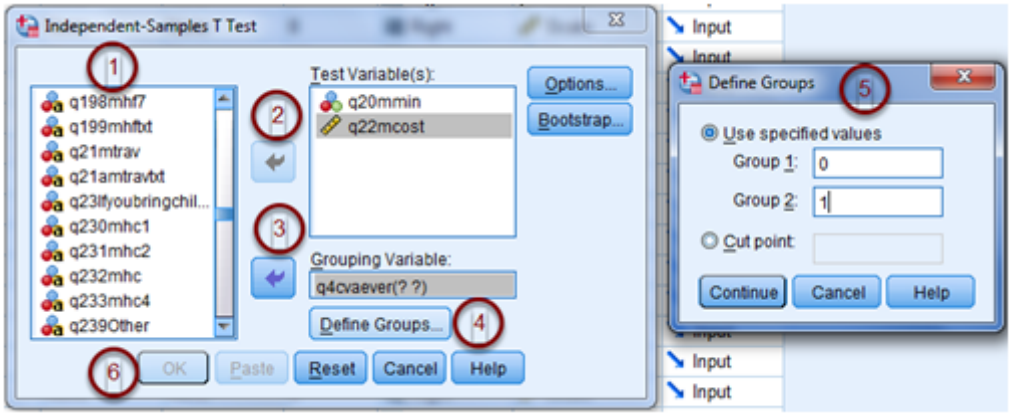
a. Grouping Variable: q4cvaever

The final data will be displayed in the output window. To determine if there is a significant difference between groups, take note of each Asymp Sig. (p-value).

Two-sample t-test (For normally distributed data)



To conduct the two-sample t-test, click "Analyze" on the drop down menu (#1) and select "Compare Means" (#2). From the extended menu options, select "Independent-Samples T Test..." (#3).



When the menu opens, select from the list on the left the test variables you want to analyze (#1) and use the blue arrow next to the test variable list to add them to the list (#2). In the grouping variable box, add the variable that corresponds to whether or not the child received vitamin A(#3). Next select "Define Groups..." (#4) and specify the two groups you would like to compare (example 0 and 1 for those not supplemented and those who were supplemented) (#5). Once this is done select "Continue" and finally "OK" (#6).

Output

Group Statistics

	q4cvaever	N	Mean	Std. Deviation	Std. Error Mean
q20mmin	0	163	20.00	19.738	1.546
	1	7	20.00	15.275	5.774
q22mcost	0	146	132.95	209.654	17.351
	1	7	100.00	58.595	22.147

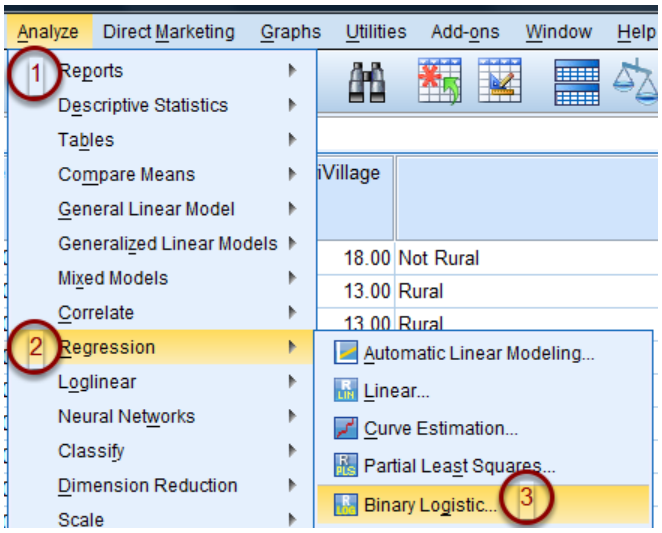
Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
q20mmin	Equal variances assumed	.046	.831 (1)	.000	168	1.000	.000	7.564	-14.933	14.93
	Equal variances not assumed			.000	6.890	1.000	.000	5.977	-14.179	14.17
q22mcost	Equal variances assumed	.711	.401	.414	151	.680	32.945	79.619	-124.367	190.25
	Equal variances not assumed			1.171	15.386	.259	32.945	28.134	-26.891	92.78

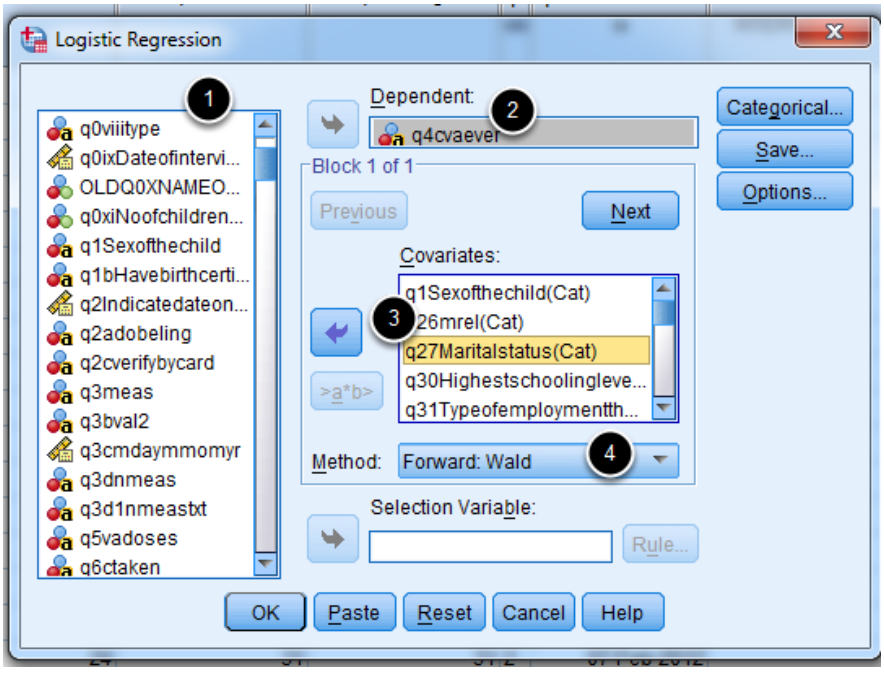
Once the t-test test is run, the final results will appear in the output window. Look at the "Equal Variances Assumed" row to determine if there are differences between groups. A significant value can be determined by looking in the Sig. column and observing if the Sig. value is < 0.05 (#1). In the example, there is no significant difference in time to get to a health facility between caretakers who attended the event and caretakers who didn't attend the event since the Sig. value is ≥ 0.05.

Data Analysis: Logistic Regression

Bivariate logistic regression is performed to identify significant relationships between chosen variables and a dependent variable (e.g. vitamin Asupplementation). In this analysis we will look at odds ratios and p-values to identify characteristics of participants who had significantly higher odds of receiving vitamin Asupplementation.



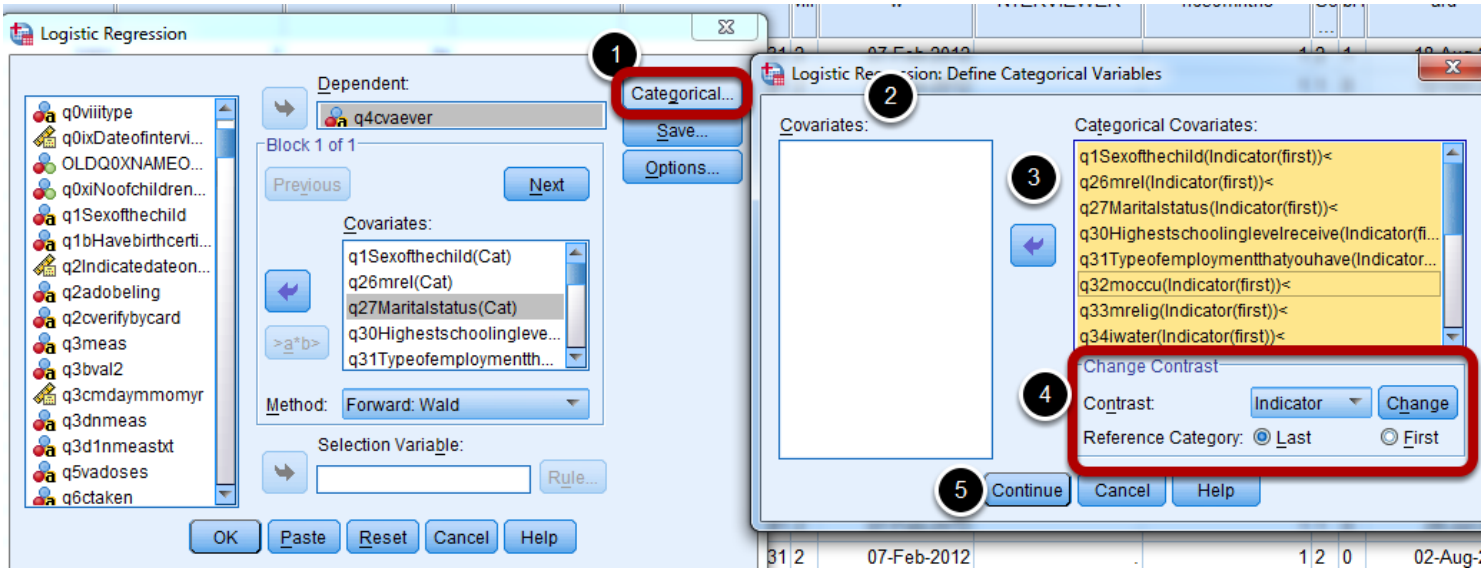
First click "Analyze" (#1) from the menu on the top of the screen. Select "Regression" (#2) from the drop down menu and finally "Binary Logisitc..." (#3).



When the Logistic Regression menu appears, select the variable that corresponds to whether or not vitamin Awas received as the dependent variable. This is done by selecting the variable from the list on the left (#1) and clicking on the blue arrow button adjacent to the box labeled

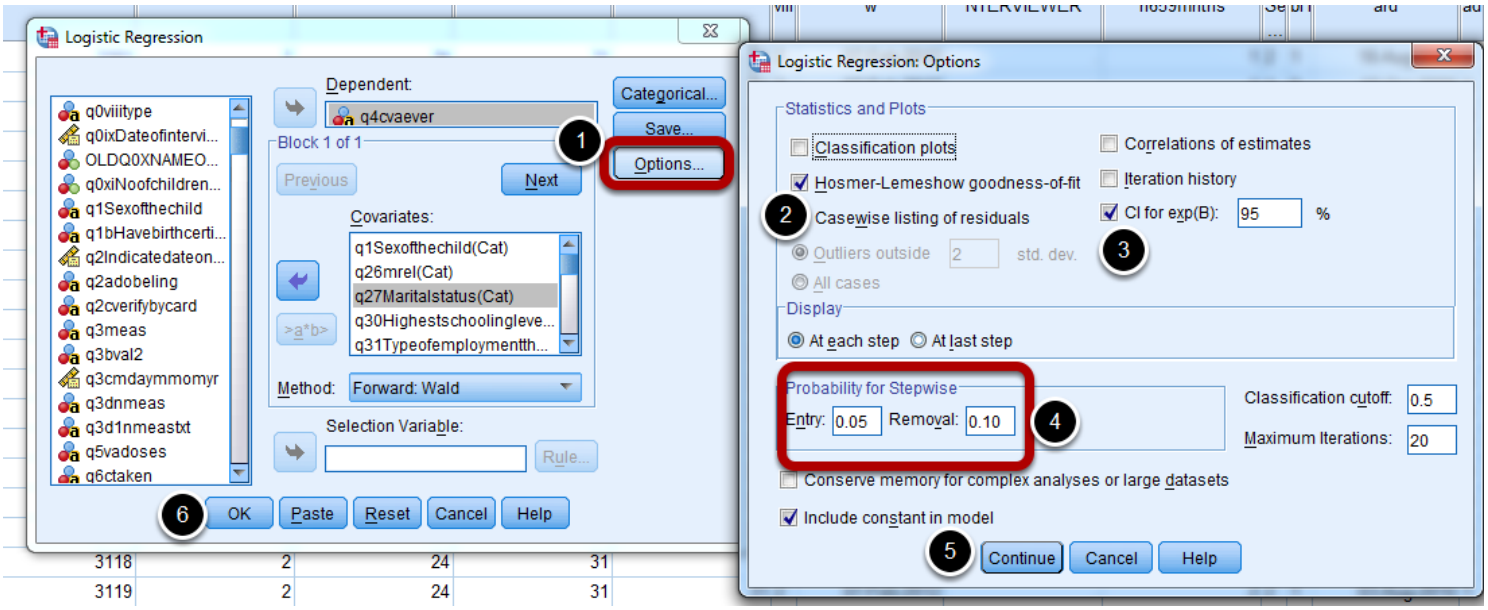
"Dependent:" (#2). In the larger Covariates box select and add all of the variables that you want to conduct odds ratios for (#3). Change the method to Forward: Wald (#4). Forward Wald is a method that enables you to look at only the significant variables and their odds.

Reference groups



In the Logistic Regression window, click "Categorical" (#1). Anew window called Define Categorical Variables appears. There are two columns in this case, one being Covariates (#2) and the other being Categorical Covariates. Ensure that all categorical variables except for continuous variables (e.g. age, distance) are in the covariates box on the right side. Variables can be moved to the Categorical Covariates box using the blue arrow (#3). Below the Categorical Covariates box, there is a section called Change Contrast (#4). Change the reference category to First and press “Change.” By doing this, the odds ratios will be calculated in reference to the first response to each question. (Last is not used for PEC surveys since in many cases, the last response is “Other”, which makes it difficult to interpret the results.). Click "Continue" (#5) to go back to logistic regression window. *Note: Ensure that the categorical variables that you select have their Variable Type listed as numeric in the Variable View tab, as described in Section 1.2.

Calculate 95% Confidence Intervals



In the Logistic Regression window, click "Options" (#1). In the first section Statistics and Plots, check Hosmer-Lemeshow goodness of fit (#2) and CI for exp(B) (#3). In the third section called Probability for Stepwise (#4) you will define your required entry and removal criteria for independent variables. By default, the entry level is 0.05 while the removal level is 0.10. This means any independent variable whose significance level is more than 0.10 will be excluded from the model. Click "Continue" (#5) to go back to the Logistic Regression window. Finally select "OK" (#6) and the logistic regression analysis results will appear in the output window.

Interpreting the output

Variables in the Equation

		B	S.E.	Sig. (2)	Exp(B) (3)	95% C.I.for EXP(B)	
						Lower	Upper
Step 1[a]	q35Doesyourchildattendthe ECDEarly(1)	-.542	.179	.002	.581	.409	.826
	Constant	-.624	.082	.000	.536		
Step 2[b]	q346Districthospital(1)	-.518	.180	.004	.596	.419	.848
	q35Doesyourchildattendthe ECDEarly(1)	-.543	.180	.003	.581	.409	.826
	Constant	-.673	.113	.000	.510		
(1) Step 6[f]	q346Districthospital(1)	-.545	.182	.003	.580	.406	.828
	q35Doesyourchildattendthe ECDEarly(1)	-.447	.184	.015	.640	.446	.917
	q361Routinevisitscheckupsgrowth(1)	.367	.157	.019	1.443	1.062	1.962
	q3702Postersomewhereotherthanhealth(1)	-1.293	.626	.039	.274	.080	.936
	q3713Religiousleaders(1)	-.688	.341	.043	.503	.258	.980
	q3714Communityleaders(1)	.459	.207	.027	1.583	1.055	2.374
	Constant	-.649	.114	.000	.523		

As described above, using Forward Wald ensures that only variables that are significantly associated with the dependent variable are included in the model. It starts by inserting one variable into the model (Step 1) and determining if it is significantly associated with the dependent variable. If so, a second variable is added to the model and the cumulative effect is looked at. There will be as many steps as there are significant associations. The last step listed (#1) comprises of all of the predictors (variables significantly associated with the dependent variable) in the final model. These are the odds ratios and p-values that should be reported. To determine a variable's significance level, look at the significance level (#2, Sig. column). Then, proceed to look at the Exp(B) column (#3) which in this case gives the odds of a child being supplemented when compared to the reference group (not being supplemented).