503/650 2016: Assignment 2 - Data management Points missed:

**Your name: Name of TA:**

Turn in this assignment sheet with your answers typed within this document. Turn in your do-file, log file, and assignment Word document file on the course LAN to your Posted folder. All Stata output in this document **must** be printed in Courier New, 9 point font, single-spaced. The assignment should be printed double sided, no line wrapping, and stapled.

Use Stata to transform and clean your data, using what you think is the best method to verify each transformation. The methods used will be used in later assignments to show that your variables are correct. You will be graded on:

1) Getting the correct answer. If the transformation is incorrect, you will get zero points. If you do not verify a transformation, it is incorrect.

2) Making a brief but convincing case that the transformation is correct. Variable and value labels are essential. Long answers are not as effective as short answers. You will lose points for answers that are unnecessarily long. For example, if *incomesq = income\*income*, do not show a multi-page table of *income* by *incomesq*. Instead, for continuous variables, use a scatterplot to show the transformation is correct.

Use the NES data (*spex cda-nes-toclean* will load it) for the following. The codebook is in the lab guide.

1. \_\_\_ of 10: Use variables *sex*, *educatio*, *didvote*, *presvote*, *partyid* and *fincome*. Create a temporary data set <username*>-nes*.dta with (a) only the six variables listed above are included; (b) only female respondents; and (c) exclude cases with missing data for any of the six variables.

Verify that you have selected the correct number of cases; that they are all females; and that missing data has been handled correctly. All subsequent transformations should use the data in <username*>-nes*.dta*.*

Your writing goes here in this font.

Stata output in this font.

1. \_\_\_ of 10: Use *educatio* to create *highschl* as a binary variable equal to 1 if *educatio* is greater than 11 years and 0 if it is 11 or less. Verify the transformation.

Your writing goes here in this font.

Stata output in this font.

1. \_\_\_ of 10: Create *presvote2* from *presvote* that has three categories: voted for Clinton, voted for Bush, and voted for someone else. Add value labels to the three categories of *presvote2.* Verify the transformation. As one of your checks (use other methods of verification as well), show that the proportions of the three categories sum to one.

Your writing goes here in this font.

Stata output in this font.

1. \_\_\_ of 10: Starting with *partyid*, create *partyid2* with these four categories: *partydem, partyindep, partyrep*, and *partyother*. Use value labels. Verify the transformation.

Your writing goes here in this font.

Stata output in this font.

1. \_\_\_ of 10: Variable *fincome* is a categorical variable classifying respondents into an income range (e.g., 3 to 5 thousand). We want a variable *income* with values equal to the midpoint for the ranges used by *fincome*. For example, if the value of *fincome* is 6 a respondent has an income between 5 and 7 thousand, so *income* would equal 6 indicating $6,000. Treat 66 as missing. Code the top open-ended interval as 125% of the start value for the interval. Code *income* in units of $1,000. Verify the transformation using the list command. Plot *fincome* against *income* and show this in your answer. Your graph must include a caption, title, and axis labels (see Workflow).

Your writing goes here in this font.

Stata output in this font.

1. \_\_\_ of 5: Create *srincome* as the square root of *income*. Plot *srincome* against *income* to verify. Your graph must include a caption, title, and axis labels (see workflow guidelines).

Your writing goes here in this font.

Stata output in this font.

1. \_\_\_ of 10: What are you asserting substantively if you use *srincome* as an independent variable rather than *income*? Giving an example can help.

Your writing goes here in this font.

Stata output in this font.

1. \_\_\_ of 10: Use codebook, compact to describe only your transformed variables. (Note: You should do this for all of the variables you use in later assignments.) For categorical variables with more than two categories, use tab1.

Your writing goes here in this font.

Stata output in this font.

1. 10: Overall effectiveness of your answer.

CDAiu2016-a02-cleaning-2016-07-26.docx