**Lab 4, Assignment #3**

**Dr. Josh Curtis**

**Due Friday October 22nd**

**Total Marks: 75**

*This lab will cover a variety of topics, including recoding, visualizing data (both frequency distributions and graphics) as well as measures of central tendency and variable dispersion. Given the breadth of topics covered you will have* ***two*** *weeks to complete this lab assignment* ***Due Friday October 22nd***

1. Using the MIDUS dataset find the variable a1sp7 which asks, “During the past year, how often have you thought your relationship might be in trouble?”. In this question you will be using this ‘raw’ variable to create **three** different versions of this variable using the recode function. Begin by running a frequency distribution for the unadjusted variable [2 points]. Next, create a dichotomous variable (i.e., one with *only* two categories). Since at least one instance of negative relationship thoughts occurring may be serious, recode this variable so that 1=never and 2=at least once. Call this variable “trouble1.” [3 points]. Label each category of the recoded variable and also label the variable overall [3 points]. Also make sure to drop all missing or inappropriate responses (inappropriate responses are responses that are missing because individuals were not in a relationship) [1 point]. Paste your syntax and a tabulation of the variable before and after recoding as your answer. [1 Marks].
2. **points total**
3. Next, create a new variable called “traouble2”, but use the categories “never”, “sometimes”, and “often”. It is up to you to decide which ‘raw’ categories should slot in to each of the three new categories listed above. You must justify you reasoning [2 marks]. Be sure to label each category of the recoded variable and also label the variable overall [2 points]. Also make sure to drop all missing or inappropriate responses (inappropriate responses are responses that are missing because individuals were not in a relationship) [1 point]. Paste your syntax and a tabulation of the variable before and after recoding as your answer. [1 Marks].

**6 points total**

1. Lastly, create a new variable called “traouble3”. Follow all steps in question 2, however this time it is up to you how you recode this variable.

**6 points total**

1. In lecture (as well as in the text) we discussed how to use graphics to analyze our data. Using the three variables you have created above (i.e., trouble1, trouble2, and trouble3) create both a) a pie chart, and b) a bar chart for each variable. Be sure to include all proper labels and titles (3 marks each \* 6 = 18 points).

**18 points total**

1. Question L3 in the **mail-in** portion of the MIDUS asks, “How many years have you lived in your current neighborhood -- or, if you live in a rural area, in your current township? (If less than one year, enter ‘0’.)” Please note that this is variable **a1sl3**. Create a new variable based on this measure that drops missing responses. Also, to prevent the influence of outliers, recode valid values above 15 to have a value of 15. Call this variable “timelive.” Also give your new variable a label that reads, “Years in neighborhood.” Paste your syntax and a tabulation from Stata for the variable both before and after coding.

**5 points total**

1. Using Stata, create a histogram for years in the neighborhood. Make sure there is a bar for each year in the data and proportion is on the Y-axis for the histogram. Also make sure to give the histogram a title and label the X and Y axes. Furthermore, make sure that the X-axis begins at the smallest response and ends at the highest response, with responses marked off on the X-axis by 5-year increments. Paste your syntax and histogram with your answer.

**5 points total**

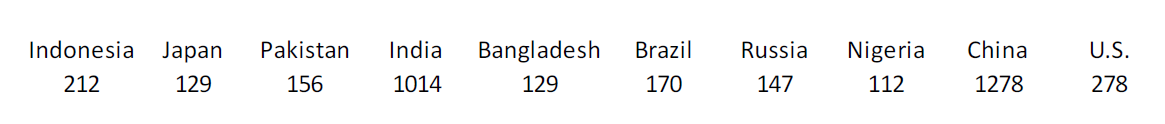
1. Question A5 in the **telephone** portion of the MIDUS asks respondents to rate their mental or emotional health by poor, fair, good, very good, or excellent. Please note that the variable for this is **a1pa5**. Recode this variable to drop missing. Also make sure to label each category of the responses, and label the responses overall. Call the new variable “selfmh” (for self-rated mental health).

**5 points total**

1. Create a bar chart for the distribution of self-rated mental health in the MIDUS data. Keep in mind that, for a bar chart, the bars should ***not*** touch. Make sure there is a bar for each response, and each bar is labeled with the response. For this figure, ***percent*** should be on the Y-axis. Also make sure to give the bar chart a title and label the X and Y axes. Furthermore, make sure that the X-axis begins at the smallest response and ends at the highest response. Paste your syntax and bar chart with your answer.

**5 points total**

1. According to United Nations figures, the populations of the world’s ten largest nations in 2000 (rounded to the nearest million inhabitants) were as follows:



**By hand,** find the values for each of the following: **range**, **variance**, and **standard deviation**

**[6 points total]**

1. In the MIDUS, respondents were asked if they are currently working for pay. Please note that this variable is **a1si14**. Using STATA, find the median, mean, range, variance, and standard deviation) [10 points]. Next, use the responses to this question to create a dichotomous variable which indicates whether the respondent is currently working for pay. Call this variable “workstatus.” Make sure to label each category of the variable and the variable itself. Paste your syntax and a tabulation of your variable before and after recoding as part of your answer. Also indicate whether the majority of the sample was working for pay or not working for pay. **[5 points]**

**15 points total**