**Assignments: Day 1**

**Group Assignment (2:00-3:30pm):**

1. Open the data titled “ihds.dta” in the dropbox. This is data from women across all of India and includes measures of their marriage and decision-making.
2. Explore this data. What types of variables do you have? Note down how many dichotomous, categorical, ordinal, and continuous variables are in the data.
3. We want to know the first age at which each woman in the dataset was married. This is stored in the variable married\_first\_age, but the data is missing for a lot of women. Why is this? Create a new measure named married\_first\_age\_complete that corrects for this.
4. Plot a density plot of your variable for age of first marriage.
5. Plot a histogram of the variable for caste (id\_caste\_category).
6. We want the education variable to be better formed into education categories. Recode the education variable so that it is a categorical variable with five categories: No Education, 1st-5th Class, 6th-9th Class, Secondary – High Secondary, and More than High Secondary.
7. Run a cross-tabulation using your new education variable and id\_caste\_category.
8. Using all of the variables on decision-making (beginning with decision\_) construct an index as the arithmetic sum of all measures.
9. Construct another index using the same variables by taking the mean (in this case proportion of 1’s) across all of the variables.
10. Using these variables, run a principal components analysis. Extract the first two principal components as additional indices.
11. Look at the correlation across these four indices.
12. Calculate the mean, standard deviation, min, and max for each of the following variables: id\_caste\_category, id\_marital\_status, your new education variable, married\_first\_age\_complete, married\_distance\_fam, and your four created indices.
13. Upload your R script to the dropbox by 9pm as: “Assignment Submissions/Group Assignments/Day 1/”.

**Individual Assignment (4:00-6:00pm):** The final assignment for this course is the creation of a 3-4 page research/policy brief addressing a specific research question.Go through the provided survey data in the class dropbox.

1. Identify a research question that could be answered using one of these data sets (or a different survey data set of your choosing, but please get approval from Davi).
2. Extract the variables that you will need to answer your research question and save this new dataset to the class dropbox folder with your name as a dta file with the filename lastname\_firstname\_rawdata.dta (“Assignment Submissions/Individual Assignments/ /Day 1/lastname\_firstname\_rawdata.dta”).
3. Get to know your data. Create cross-tabs with any relevant categorical variables. Plot the distributions of your relevant variables. Calculate the means and standard deviations and create a summary table for your variables.
4. Determine whether you need to recode or reformat any of your data to better suit your hypotheses/analysis (such as by creating binary indicators, clustering of response options, or indices to combine multiple measure). If so, create new variables for all of these. Note that it is the rare exception that a project would not need to create at least one new variable.
5. Save your edited data to the dropbox as: “Assignment Submissions/Individual Assignments/Day 1/lastname\_firstname\_editdata.dta”.
6. Save your R script to the dropbox as: “Assignment Submissions/Individual Assignments/Day 1/lastname\_firstname\_editscript.R”.
7. Write a short description (~1 page) of your research question, hypotheses, and data. In this description, make sure to include a section on variables, describing the independent and dependent variables from your dataset that you will need to measure in your survey. Also include the summary table and any interesting/useful cross-tabs or distribution plots for your data along with a corresponding description.
8. Upload your write-up to the dropbox by 9pm as: “Assignment Submissions/Individual Assignments/Day 1/lastname\_firstname\_day1writeup.docx”.