Survey design

Assignments: Day 3

**Group Assignment (3:00-5:00pm):**

In your group, walk through the provided Stata script/do file on sampling. After doing so, complete the below assignment in Sata using the provided data. Upload your stata do file to the “Assignment Submissions/Group Assignments/Day 3” folder in the class dropbox by 9pm.

The following questions are to be completed with your group. The majority of the questions involve coding in Stata, for which you will find the shared scripts on sampling useful. Even for the non-coding questions, please record all answers in a Stata do file, just commenting out responses that are not code.

1. Sample Frames

Below is a list of populations that we would like to sample for a survey. For each, come up with a sampling frame for this population. Describe the pros and cons of your chosen sampling frame, paying particular attention to coverage, listing availability, and representativeness.

1. All households in Brazil.
2. All migrant workers living and working in Nepal.
3. All members of Danish biker gangs.
4. All candidates for national office in the United States 2018 election.
5. All women living in rural India.
6. All politicians in Uruguay who have accepted a bribe.
7. All international aid organizations working in Sub-Saharan Africa.
8. All households in favelas in Brazil.

2. Choosing a Population

1. Select a dataset from the Survey Examples in the course dropbox (or request approval for a dataset that your group has found and would prefer). Load this data into Stata. This data is your population. Do not worry about how the data was collected; for the purposes of this class, treat this data as the total population (and as your sample frame). Select at least one variable for which you would like to “estimate” the population mean.
2. Calculate the true population mean and standard deviation. Note that you will not usually have this information.

3. Simple Random Sample

1. Take a simple random subsample from your “population” in Stata of size 1000 (or less if your population size is less than that).
2. Use this subsample to estimate the sample mean for your chosen variable of interest.
3. Compute the standard error for this estimate.

***Optional:***

4. *Stratified Random Sample*

1. *Take a stratified random subsample from your “population” in Stata using proportionate to population sampling, selecting at least one variable to use as your strata. Comment in your code as to why you chose this variable(s).*
2. *Use this subsample to estimate the sample mean for your chosen variable of interest*
3. *Compute the standard error for this estimate.*
4. *Calculate the design effect for your stratified random sample. Did you improve your precision?*

*5. Clustered Random Sample*

1. *Take a cluster random subsample from your “population”, selecting at least one variable to use as your cluster variable. Comment in your code as to why you chose this variable(s).*
   1. *First sample m clusters at random (with m being a cluster sample size that seems reasonable).*
   2. *Then sample 1000 elements with the same number of units sampled per cluster.*
2. *Use this subsample to estimate the sample mean for your chosen variable of interest*
3. *Compute the standard error for this estimate.*
4. *Calculate the design effect for your clustered random sample. Did you improve your precision from the simple random sample?*

**Individual Assignment (5:00-6:00pm):**

Building on the research question you described yesterday, first decide on a target population and then develop a sample frame and sampling protocol for this study. This sample frame must be realistic and feasible. Write up a description of this protocol and upload this to the “Assignment Submissions/Individual Assignments/Day 2” folder in the class dropbox by 9pm.