Take Home exercise week 40

Next week (week 41) we will devote the entire class time (including take home exercises) on working with (combinations) of clustered and stratified samples. This is to get you more acquainted with the survey package in R, but also to explore some more complicated issues that you encounter in real-life settings. In order to prepare for next weeks class, there are a few things that you need to do.

1. Revisit (or catch up) with the reading from Stuart.
2. Finish the class exercises on stratified and cluster sampling.
3. Work out what is the sampling design for your "adopted" survey (see weeks 37-38). This may sound easier than it is. You may be "lucky" in the sense that your sampling design is relatively simple; even then you will need to double check whether the design is indeed as simple as you think it is. The following steps may be helpful in preparing you for class in week 41.
   a. Read the survey documentation on the website of your adopted survey to figure out whether stratification and clustering was used. I usually make a drawing to make sure I understand what is going on. If you adopted a longitudinal survey, focus on the first wave of the study. If you adopted a cross-national survey, focus on one country (of your choosing).
   b. Download the datafile. Sometimes, the datafile can be directly downloaded from the website of the study. More often, you will need to download it from a "data archive", where studies deposit the data. Data archives ensure that the study is well documented, and going into the future, that the data are saved in a format that is readable by statistical software. In some cases, the data archive will want you to shortly write why you need the data.
   c. Load the data into R. Datasets are seldomly released in the "R" format and most often are released as a ".sav" file (SPSS), .dta (STATA) or .csv (general spreadsheets). You will need to install a package (e.g. "foreign") to read .sav and .dta files in.
   d. Inspect the data to make sure it loaded correctly. Identify the variables in your dataset that indicate clustering and/or stratification. In some cases such information may not be directly available, but be summarized in a "design weight" variable. If so, that is fine as well, we will discuss this next week (in that case, for now it will suffice to plot the design weights (e.g. a histogram).
   e. Using both the survey documentation (a.) and the datafile, can you work out the sampling design? It helps here again to draw a figure and add numbers if you are dealing with strata or clusters.
   f. Keep a list of issues you are encountering; for example where the documentation is unclear, or where the dataset is inconsistent with the documentation. Add your findings/conclusions and questions to the document you started for your adopted survey in week37/38. We will continue working on this document in the next weeks, and use that as the basis for assignment 2.

We will start week 41 with a short discussion of issues you encountered, so bring your document to class.