

Survey Data Analysis

Take Home exercise week 4 – Stratification and clustering in sampling.

Peter Lugtig – p.lugtig@uu.nl

Next week 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 week's class, there are a few things that you need to do. This exercise will also directly help you to answer the questions in assignment 2 (due in November)

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 1,3). You should have done this based on the documentation of your survey, but notw the idea is that you get the data for your survey, and match what you found in the documentation to variables in the dataset. I.e. to find the stratification and cluster variables in your data. The following steps may be helpful:
 - a. 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. You can then just mention that you need the data for a class in survey data analysis. If you encounter problems here: contact Peter via p.lugtig@uu.nl
 - b. 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. Use your best friend ‘google’ to find out how to do this.
 - c. 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 is summarized in a “design weight” or “analysis weight” variable. If so, that is fine as well, we will discuss this next week.
Make some plots or tables of stratification variables, psus, design weights, analysis weights or other variables if they are available
 - d. Using both the survey documentation that may be available online and/or the datafile to you work out the sampling design? It helps here again to draw a figure/ add numbers if you are dealing with strata or clusters, so that you understand how many strata or clusters there are.

- e. 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 week1, 3. We will continue working on this document in the next weeks, and use that as the basis for assignment 2.

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