Survey Data analysis –Peter Lugtig – last update: 3 Sep 2022

Take Home Exercise week 2 - Exercise on probability
(as with all take home exercises, finish this exercise before the start of the next class)

From today on, we will do exercises in R. We will do such R exercises almost every week from now on, both at home and in class. One reason is that we try to teach you R quickly, and the best way to learn a new software package is to use it a lot. A second reason is that you will need sophisticated software for the more complicated and real-life situations where you will want to draw samples or analyse complex survey datasets that we will encounter in this course. STATA and R are currently the best software packages for doing survey data analysis.

There is no need to hand in all your answers or syntax here. Just write down your answers for yourself (or bring your laptop to the next lecture), or – even better – annotate your R code with the #comment option. You'll have to hand in some answers for the final exercise though. Do this with the new excel sheet on Teams.

1. Download the file "THE_exercise_week_2.R" from Blackboard.
2. What you see is computer code (syntax), which tells R what to do. Some parts of the code may already be familiar to you, but some parts may be new. If so, try to figure out what is going on.

3. The goal of this exercise is to draw a certain number of cards from a deck of cards (52). For now, you have to imagine that the deck of 52 cards is our "population". Without running any code, what is the number of cards with the suit "spades" that you would expect when you draw 8 cards?

4. Run the whole code. Didactically, it is better to run it line by line, and then see what happens in the panel where the output appears, but that is not strictly necessary: you can also run everything in one go and scroll up and down in the output panel. Have a look what is being done on each line.

5. What happens to the number of spades you get when you sample without replacement (so not putting draws back in the deck)
   a. When you draw 10 cards? (search the syntax for the line of code that does this).
   b. When you draw 40 cards?

6. If you would draw cards one-by-one, and put them back after looking at them (sampling with replacement)
   c. when you draw 10 cards?
   d. When you draw 40 cards?

7. Do you find a difference when you look at the number of Spades?

8. Repeat questions 5 and 6 once, so you get 8 estimates in total. Bring the numbers to class

Optional:
9. Perhaps you are thinking that repeating the analyses just once is not very informative. The number of Spades will vary every time you take a draw. However, now imagine repeating the analyses 1000 times (or even more). You then get 1000 estimates of the number of spades in the deck of cards. This imaginary distribution of your 1000 estimates is called the "sampling distribution". We will illustrate the sampling distribution at the start of class in week 3. However, if you want to take a sneak peek:

download and run the file "simulation_SRS_cards.R".