Exercise 2   **15:15-16:45 (Mon.) 01 Feb, 2021**

Hi all,

This exercise 2 follows the lecture **Data clearance (2) & Summary statistics**. For this practice, you will get your hands dirty with the already appended Stockholm Public Health Cohort data. (Not yet done? Follow exercise 1 to append your data.) Through this exercise, you are expected to practice and review what you learned previously in your Biostat courses or in this lecture to manage variables and learn about the data.

Part 1: Manage variables (ref. Data clearance: Manage variables)

Please open your sphc_all.dta and look up the variables in the codebook SPHC_variable_list.xls. (Supposedly, you have already appended sphc_2002.dta, sphc_2006.dta, sphc_2010.dta.)

1. Label the data sphc_all.dta.

2. Rename variables: `kon` → `sex`, `aldkl4` → `agegrp4`, `f302` → `srh` . E.g., rename `kon` into `sex`.

3. Open your codebook (SPHC_variable_list.xls), and add one column entitled "Renamed variable". And add those three variables' new names.

4. Label the 3 variables above. E.g., `label sex sex`.

5. Label values the 3 variables above. E.g., label values `sex` with `1=man, 2=woman`.

6. Generate a new variable `female`, where `female=1` is woman, and `female=0` is man. And also label that variable. (Hint: think about what is the difference between `=` and `==`)

7. Generate dummy variables for `agegrp4`.

8. Generate a new variable `nsrh`, of which self-rated health=5 implies excellent, ..., and =1 implies very poor. Caution: the original self-rated health variable (`f302`; `srh`) is in reverse order.

* Avoid using `recode`. Otherwise, you have to make sure the codebook is correct.

Part 2: Learn about the data

1. Describe the data, sphc_all.dta.

   - No. of variables, observations. (ref. Data clearance: Get to know the data)
   - Choose a variable to describe the mean, standard deviation, variance, median, min, max, range. (ref. Summary statistics)

2. Stratification

   - What is the overall proportion of living alone? What about if your separate by 4 age groups (`aldkl4`; `agegrp4`)? (ref. Summary statistics)
   - What is the proportion of females by year? (ref. Summary statistics)

Part 3: Homework

1. Try to make some graphs or tables according to your research question(s). Bring them to our next session.

2. Rename and label the other variables.

* If you rename variables, you have to go back to your codebook to fill in the new name. Otherwise, you might forget what variable names you have changed.