

Exercise 4

15:15-16:45 (Mon.) 08 Feb, 2021

Hi all,

This exercise 4 follows the lecture **Graphs**. However, you may also need to use some techniques from **Tables & interpreting results** as well. You are expected to capture when to use graphs or tables to demonstrate your study results, and also how to plot graphs with sufficient information.

Part 1: Types of graphs

Please open your `sphc_all.dta` and the codebook `SPHC_variable_list.xls`.

1. Make a **bar chart** to show how many participants joined the survey by year.
2. Make a **box plot** to demonstrate overall self-rated health by year (or stratified by any variable if you would like).
3. Following the question above, is bar chart a good way to show the results here? Why or why not? (Hint: think about the definitions of median and mean. And what statistics is bar chart based on?)
4. Following the question above, instead, how would you demonstrate the results? Try to write two sentences to describe your thought.

Part 2: twoway



Figure 1: Mean self-rated health by age group (18-29, 30-44, 45-64, 65-84) and by survey year (2002, 2006, 2010)

Please plot "Mean self-rated health by age group, in 2002 2006 2010", similar to the figure above.

Suggested steps:

1. Generate the mean of self-rated health by age group and by year. (Hint: `tabstat`)
2. Put the statistics generated by `tabstat` back to the dataset. (Hint: `help tabstat, return list`)

3. Then plot a **scatter plot** showing mean self-rated health by age group and by year. (Hint: `twoway scatter`) Think about why making a scatter plot instead of a line graph?
- * If you encounter difficulty in making this scatter plot, please at least try starting from Q2.3. Write some codes for the scatter plot, e.g., the legend, x-axis, y-axis, and titles.