Statistic Workshop 1

Bolun

2021-01-12

Welcome

Welcome to the statistic workshop. My name is Bolun and I am going to help you to navigate this field in the sequence of SOCG205 and SOCG206.

You can contact me via email: boz023@ucsd.edu, I have a regular office hour as well.

- This workshop follows the philosophy of "learning by doing".
- Mainly you will participate in group activities in the workshop.
- Self-introduction

Group activity: your story related to statistics

Here I want to introduce the main model of group activities we are going to use through the workshops.

We are going to divide into 3 groups, and will discuss the following questions:

- Share your previous experience about statistics with other members in your group. What do you like or dislike about statistics?
- Which part you think is most difficult to learn?

After the discussion, one member from each group will brief their discussion to other people.

Group Activities: Max Weber and Statisitcs

Reading the footnotes from the first chapter of Protestant Ethics, finish the following tasks:

- Identify three different kinds of variables in these footnotes.
- Reframe Max Weber's research question using the data he presents in the footnotes.

Statistics is a way of storytelling

Statistics is a way of storytelling, just like other narrative methods. We should not see quantitative studies and qualitative studies as two kinds. They are just using different methods to answer our questions about the social world.

The overall learning goal of this course

Along this line, there are three things that I will try my best to let you get during the course.

- Knowledge of statistics and skills of coding that are covered in SOCG205 and SOCG206.
- Basic logic of statistical research.
- Skills that you can learn quantitative method yourself outside the course sequence.

Tables and Visualization: Present Different Kinds of Data

Please go to https://rstudio.cloud/, and click on “Get Started For Free”, register an account.

Then, click on “File” on the left top corner, create a new r script file.

After that, please go to the following link to copy all the codes into the script.

Tables and Visualization (1)

Categorical Variable: Gender as an example

- Bar plot
- Frequency table

Tables and Visualization (2)

Discrete Variables: years of education

- box plot

Tables and Visualization (3)

Continuous variables: income (note: in social science research, real contiunous variable are very rare.)

- Histogram
- Frequency polygon
- Scatter plot for income and age

Group Activities: try visualization

Group activity: choose three different types of variables from "World" dataset(see pdf for codebooks). and visualize then by modifying the code in the google docs. Present the results to the rest of the class.

Use the following code to load the data.

```
library(poliscidata)
data <- world
```

You can use the following code to check the codebook of a variable

```
memisc::codebook(data$dem_level4)
```

Summary

- What we learn today: types of variables, how to represent them in tables and graphs.
- Next workshop: Frequency distribution.
- Further information on installing R and Rstudio on your own computer.
- Redo the visualization using ggplot2.
- If you choose to learn it in a hard way, I have a datacamp assignment. You can find it in datacamp.com.
- adding your questions, answers and tips to https://docs.google.com/document/d/16z74gTLKkR7ujRjTvBDIfDn6oE_GyWNIatYvir746CY/edit?usp=sharing