SOC6707 Intermediate Data Analysis, Winter 2022

Assignment 3

Due date: March 29 11:59pm

Details

There are **100 points** in total.

You will need to submit both your answers to the questions and accompanying R code. You should submit:

- your R Markdown file; and
- the knitted PDF resulting from your R Markdown file.

Please submit both files via Quercus.

Remember to:

- Label the answers to each question
- Label any graphs clearly with suitable axis labels and titles
- Comment your code so that it is easy to understand

Question 1

This question relates to the future fertility intentions of a sample of women in the US. The dataset `intentions` is a cleaned version of the National Survey of Family Growth 2015-2017.

a)

i) Filter the dataset to just include women aged 25 and above
ii) Create a new variable `bachelor_or_higher` that is equal to "yes" if a woman has a bachelor or post grad degree and "no" otherwise
iii) Use the `fct_relevel` function to make the baseline category of `marital_formal` equal to "Never Married"
iv) Report the resulting number of women in the sample

b)

i) Calculate the proportion of women who want more kids by age group (`age_gp`) and education (`bachelor_or_higher`)
ii) Create a bar plot that is the proportion of women who want more kids by age group (on the x-axis) and education (fill of bars). Use `position = "dodge"` in the `geom_bar()` function to plot the bars by education group next to each other. Interpret what you see.

c)

Do the same thing as b) but now by parity (current number of children) and marital status (`marital_formal`):

i) Calculate the proportion of women who want more kids by parity and marital status (`marital_formal`)
ii) Create a bar plot that is the proportion of women who want more kids by parity (on the x-axis) and marital status (fill of bars). Use `position = "dodge"` in the `geom_bar()` function to plot the bars by marital status group next to each other. Interpret what you see.

d)

i) Run a logistic regression with `wants_more_kids` as the dependent variable, and `parity`, `age`, `marital_formal` and `bachelor_or_higher` as the independent variables.
ii) Interpret the coefficient on parity
iii) Interpret the cofficient on marital status = "Married"

e)

Use the model in d) to estimate the probability that a married woman aged 35, with 2 kids and PhD[1], wants more kids.

[1] This is Monica.