

SOCI 385 - Social Statistics

Fall 2021

Assignment 2

This assignment is due via Canvas by 10:00 AM on Monday, September 27, 2021.

Please show all your code and write your responses in complete sentences as appropriate. Submit your knitted pdf and the R Markdown (.Rmd) file for your notebook. Remember to include your name in the header.

This week you will continue practicing with descriptive statistics. This assignment requires the `assignment_02.csv` file which includes data from the 2008-2018 waves of the General Social Survey. Your dataset should have 9,569 observations of 4 variables.

The variables are:

`id`: a unique identifier for each respondent

`year`: the year the respondent completed the survey, from 2008-2018

`educ`: respondent's highest year of schooling completed, from 8-20

`agekdbrn`: respondent's age when their first child was born, from 12-57

Questions

1. What are the mean and median of `agekdbrn`?
2. Save the value of the 25th percentile of `agekdbrn` as an object, save the value of the 75th percentile of `agekdbrn` as an object, and find the difference between these two values using the object names and R's built-in calculator function. In words, describe what this range tells us. **[NOTE: Object names have to start with letters, not numbers!]**
3. What is the mode of `agekdbrn` for respondents who completed 12 or fewer years of education?
4. What proportion of respondents completed exactly 16 years of education?
5. Use the `dplyr` package (remember to load `tidyverse` first!) to create a new data frame with only the `agekdbrn` and `educ` variables, and that is limited to respondents who have 16 or more years of education. How many observations are in this new data frame?
6. What are the mean and median of `agekdbrn` for respondents in this new data frame?
7. Approximately how long did it take you to complete this assignment?