

SOCI 385 - Social Statistics

Fall 2023

Assignment 1

This assignment is due via Canvas by 10:00 AM on Tuesday, September 26, 2023.

Please show all your code and write your responses in complete sentences as appropriate. Use PDF as your output format when you set up your Quarto file. Submit your rendered PDF and the Quarto (.qmd) 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_01.csv` file which includes data from the 2012-2021 waves of the General Social Survey. The dataset is available at [this link](#). Your dataset should have 9,731 observations of 4 variables.

The variables are:

`id`: a unique identifier for each respondent

`year`: the year the respondent completed the survey, from 2012-2021

`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. Find the value of the 33rd and 67th percentiles of `agekdbrn`. In words, describe what this range tells us.
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 tidyverse functions (remember to load the package first!) to create a new data frame with only the `agekdbrn` and `educ` variables, and that is limited to respondents who have 14 or more years of education. How many observations are in this new data frame?
6. Use tidyverse functions and the new data frame you created in the previous question to find the mean of `agekdbrn` for respondents with each year of highest schooling completed.
7. Approximately how long did it take you to complete this assignment?