

SOC 4930/5050: Problem Set 02 - Descriptive Statistics

Christopher Prener, Ph.D.

September 11th, 2017

Directions

Complete all of the following questions using the data from the `testDriveR` package. Your well-formatted R Notebook source (the `.Rmd` file) and `html` output should be uploaded to your assignments repository by 4:15PM on Monday, September 18th, 2017.

Part 1: Descriptive Statistics

Use the `gss16` data frame saved in the `testDriveR` and calculate the most appropriate descriptive statistics for the following variables. Statistics you should consider are: mode, median, mean, standard deviation, range, and inter-quartile range.

1. Keep only the following variables - `ID_`, `SEX`, `HRS1`, `INCOME16`, `RACE`, `WRKSTAT`
2. Keep only observations for respondents who work full time, i.e. where `WRKSTAT == 1`
3. Calculate the appropriate descriptive statistics for the following:
 - (a) `RACE` for full time workers - where 1 is “white”, 2 is “black”, and 3 is “other”
 - (b) `HRS1` for full time workers - the number of hours worked during the last week
 - (c) `SEX` for full time workers - where 1 is “male” and 2 is “female”

Part 2: Plotting Data

Use the same GSS data to produce the following plots:.

4. Create a bar plot using the “race” variable you renamed
5. Create a scatter plot of last year’s income by hours worked for full time workers with the points colored using the “race” variable you renamed