GV207 , 2020
Instructor: Ryan Bakker

Homework assignment 3: 2020/2021 Reassessment

Write an R code file named (`gv207-HW3.R`) to complete the following tasks.

Rules–2pts each

- Submit two files, and two files only. That is, submit (1) the coversheet and (2) your R code file (`gv207-HW3.R`). Don't submit your graph or other outputs.
- Make sure that you delete your name from your R code file.
- Execute everything before you submit (e.g., CTRL + A & CTRL + Return on a Windows PC; Command + A & Command + Enter on a Mac machine), and make sure your file runs without an error.
- Your file must have a proper header.
- Add comments and annotations to everything you do. Try to make your code file look like my code file. Don't copy and paste all the questions into your R code file, but do show me the question number for each question.

Tasks

1. You are going to use the following 6 `R` packages: `ggplot2`, `gmodels`, `Hmisc`, `stargazer`, `effects`, and `gridExtra`. Load all the packages. [2 points]
2. Load the World Value Surveyl dataset available on Moodle (`wvs.csv`), and store it as an object named `wd`. [2 points]
3. The unit of observation is an individual survey respondent. How many respondents does the dataset have? That is, how many rows are there in the dataset? Provide a command to get the answer. Also, write your answer in a comment line. [2 points]
4. The dataset contains various information on survey respondents, including whether or not they feel people can be trusted or not (`peoptrust`). . Create a simple frequency table of the `peoptrust` variable (that is, there is no need to change the column names or to obtain relative percentages for this task; one line of command is sufficient) to see how many respondents trust people and how many feel you need to be careful around other people. Provide a command to create a frequency table of this variable (no need to write a comment). [2 points]
5. Calculate the percentage of those who are trusting of other people. Provide command(s) to calculate it. Also, write your answer in a comment line. [2 points]
6. For the next few tasks, we will analyze the relationship between trust and religiosity using one of the three bivariate hypothesis testing methods you have learned in Weeks 6 & 7. The question we ask here is: how does self-reported religiosity effect whether people are trusting of others? Let's say we hypothesize that being religious is positively associated

with trusting people. First, what are the dependent and independent variables in our investigation? Provide your answers in a comment line. [2 points]

7. Second, create a two-way frequency table (a.k.a, cross tabulation) of the two variables, `relig` and `peoptrust`. Provide a command to create such a table (no need to write a comment). Make sure that (1) values of the dependent variable are shown in rows and the independent variable in columns, (2) your table shows column percentages but not row percentages, cell percentages, or χ^2 contributions, and (3) your table produces a χ^2 test statistic. [6 points]

8. Read the table you produced the above and answer a few questions. (a) What is the trusting percentage among the religious respondents? (b) What is the trusting percentage among the non-religious respondents? (c) What is the trusting percentage among the atheists? Provide your answers in comment lines. [3 points]

9. Would you say that the relationship between `peoptrust` and `relig` is consistent with our hypothesis? Why or why not? There is no need to comment on statistical significance, but do comment on the pattern observed in the sample. Provide your answers in a comment line. [3 points]

10. The dataset contains a variable named `suicide`, which asks people whether or not suicide is something a person has the right to do. It is measured on a 10 point scale with higher values repsenting more support for people having the right to end their own lives and lower values meanign people are opposed to the right to end one's own life. Are female respondents more or less supportive of peoples' rights to end their own lives? Explore the relationship between the `sex` variable (coded as "Female" for female passengers and "Male" for male passengers) and `suicide`. Choose an appropriate bivariate statistical testing method for these two variables from the three methods you have learned in Weeks 6 & 7, and perform the test. Provide command(s) to perform the analysis. (Hint: I am not asking you to run a regression.) [5 points]

11. Interpret the results of the bivariate test you performed above and answer the question (do female passengers tend to support the right to commit suicide more than men?). Comment on the observed pattern in the sample as well as the statistical significance, and draw a conclusion (i.e., answer the question posed here). Your answers must have up to three sentences. [6 points]

12. The dataset contains a variable named `age` (age of the respondent). Do older respondents tend to have a different view of suicide compared with younger respondents? Explore the relationship between `age` and `suicide` graphically. That is, create a plot that shows the relationship between these two variables using the `ggplot` function. Provide commands to create the plot. [5 points]

13. Perform an appropriate bivariate statistical test (again, choose one from the three methods covered in Weeks 6 & 7) to explore the relationship between `age` and `suicide`. Provide command(s) to perform the analysis. (Hint: I am not asking you to run a regression.) [5 points]

14. Interpret the results of the bivariate test you performed above and answer the question (do older respondents tend to have a different view toward suicide than younger respondents?). Comment on the observed pattern in the sample as well as the statistical significance, and draw a conclusion (i.e., answer the question posed here). Your written answers must have up to three sentences. [6 points]

15. Regress `suicide` on `age`, and produce a regression table using the `stargazer` function. [4 points]

16. Create a plot that illustrates the estimated effect of `age` on `suicide` based on the model you estimated above. [3 points]

17. Judging from the numerical and graphical results of the regression analysis, would you say that `age` and `suicide` are positively or negatively related? Comment on the observed pattern in the sample as well as the statistical significance, and draw a conclusion. [6 points]

18. Graphically explore the relationship between `age` and `suicide`, holding constant the `sex` variable. That is, create a plot using the `ggplot` function that shows the relationship between `age` and `suicide` for female and male respondents separately. Try to have one plot that has two panels (one for female and one for male). [5 points]

19. Regress `suicide` on `age` and `female`, and produce a regression table using the `stargazer` function that summarizes the results of this model and the model you estimated in task 15. [4 points]

20. Which one of the two regression models performs better? Provide at least two pieces of information to justify your response. [6 points]

21. Create a plot that illustrates the estimated effect of `female` on `suicide` based on the second regression model you estimated. [3 points]

22. Create a plot that illustrates the estimated effect of `age` on `suicide` for male and female respondents separately based on the second regression model you estimated. Try to have one plot that has two panels (one for female and one for male). [6 points]

23. It may be the case that there the effect of age is conditional on sex. To test this, estimate a model that includes an interaction term between age and sex. That is, include age, sex, and the interaction as indepedent variables. Produce a stargazer table that presents the results. [4pts]

24. Produce a graph that displays the relationship between age and suicide for men and for women. Include a brief description of what this graph tells us about the effect of age on attitude toward suicide for the 2 groups. (Is the effect of age different between men and women? If so, how is it different?). [4pts]

25. . Write out the implied regression equation for men and for women (2 equations). Be sure to comment these out. [4pts]

End of file