Week 2 – Data exploration & cleaning

Question

Answer

1. What is your name?

Jennifer Harvey

2. When conducting statistical analyses, we should always keep in mind the larger context in which we are working with data. What are steps in the research process which are completed before we analyze data? (Hint: This was on the week 1 assignment.)

a. Generate research questions

b. Generate research hypotheses

c. Determine study design, variables, levels of measurement

d. Collect data

3. Data needs to be examined and any problems fixed before analyses can be done. In statistics, what are 4 main areas to examine as part of pre-analysis checking? (Hint: This was on the week 2 assignment.)

a. Data accuracy

b. Missing data

c. Outliers

d. Statistical assumptions

4. What are important steps to complete when checking data accuracy? (Hint: This was on the week 2 assignment.)

a. Make sure the data types are correct. (What measurement of each variable?)

b. Check the data for typos. (i.e. mailes for males, feemales for females)

c. Check the data for nonsensical values. (i.e. 1000 year old people, BMI of 500)

d. Check categories make sense. (i.e. variable is gender but has a category of blue)

e. Correct problems if possible or omit the data from the analyses. (Use filtering to select

the desired data and leave out the data not desired)

f. Reverse code instrument items if needed. (Make sure to do needed recoding before calculating summary scores.)

g. Calculate summary scores for any instruments that need it. (Make sure to use instrument scoring methods recommended by the instrument developer.)

h. Keep track of what you do so you can report it as part of the analysis steps you completed. (Be transparent. completed. à Make notes as you complete analysis steps. Keep the code you used with the output.)

5. Open the Week3.rds dataset in Jamovi. (Note any difficulties.)

thank goodness easy

6. What are the variables in the dataset and the level of measurement for each one? Make sure the level of measurement for each variable is correct in Jamovi. (Notice the first one is done for you.)

a. sex – nominal (no problems noted)

b. research - nominal - no problems noted

c. height- continuous (ratio) - no problems noted

d. weight - continuous (ratio)- no problems noted

e. lvst1 - ordinal - no problems noted

f. lvst2 - ordinal - no problems noted

g. lvst3 - ordinal - no problems noted

h. lvst4 - ordinal - no problems noted

i. frst1 - ordinal - no problems noted

j. frst2 - ordinal - no problems noted

k. frst3 - ordinal - no problems noted

7. Check the dataset for accuracy using visual inspection and descriptive statistics. Note any problems.

- All of the lvst and frst variables were listed as nominal but are ordinal - I changed each one

- No missing data. Weight of 9.2 appears to be an outlier. The outlier could be eliminated prior to running the data.

8. Create a new variable in Jamovi for BMI. The formula for BMI is weight (kg)/height (m)^2. Weight is already in kg and height is in cm.

You will need to create a computed variable for this.

done

9. Create a new variable in Jamovi for lvst4\_recode. Change the values in lvst4 according to the following pattern: 1=5, 2=4, 3=3, 4=2, 5=1.

You will need to create a transformed variable for this.

done

10. Save your Jamovi (.omv) file. You will turn it in.

done

11. Open the .Rmd file. Note any difficulties

done no difficulties

12. Run the code blocks. Note any difficulties.

done no difficulties

13. What level of measurement are the newly created LOSS and FOSS variables?

a. LOSS\_total - interval

b. FOSS\_total - interval

14. What descriptive statistics and plots should you use to examine the new variables?

a. LOSS\_total descriptives mean (sd)

b. LOSS\_total plots histogram and boxplot

c. FOSS\_total descriptives mean (sd)

d. FOSS\_total plots histogram and boxplot

15. What are some advantages of using an average to score an instrument over a sum?

- final score is in the same scale as the original items so interpreting can be a bit more straightforward

- possible to calculate the mean when there are missing data without changing the scale of the final score

16. What units are z-scores in?

Standard Deviations

17. What is the meaning of the following z-scores?

a. positive z-score The score is above average

b. negative z-score The score is below average

c. z-score of 0 The score is the same as average

d. z-score absolute value > 2 possible univariate outlier

18. Why are z-scores useful?

o z-scores allow for the direct comparison of variables measured in different units

o z-scores show how a score compares to the mean

o z-scores help identify univariate outliers

19. What impact does scoring an instrument using sum vs average have on a z-score?

- There is no impact on the z-score.

- z-scores will be the same for sums and averages

20. Are there any z-scores in the LOSS and FOSS variables with an absolute value > 2? If so, how many in each?

a. LOSS\_total – 4

b. FOSS\_total – 4

21. After you finished reading through the .Rmd file. Knit the file to markdown and save the .md file. You will turn the file in as part of your assignment. Describe any difficulties. If you encounter errors close all instances of RStudio and try opening it again. (Don’t save your workspace.) If you encounter errors while creating the file and can’t figure out how to fix them before the assignment is due. You can paste thetext of the .Rmd file with the output into a Word or pdf file and turn in that.

Initially I kept getting errors then watched the office hour recording, then also saw your post on the discussion board about the one chunk with the eval=false. Changed that and success. yayyyy.

22. What did you like/dislike about this assignment?

Love the repetition. It helps. Disliked the error I kept getting on the knit.

23. How would you change this assignment to make it better for future students?

Take out the eval=false

24. Complete the answer submission in Canvas and turn in your RStudio and Jamovi files. (The answer submission in Canvas allows multiple attempts. You may want to save the text you provide in open responses like the two questions above so you can just paste it into the answer field if you make additional attempts to improve your score.)