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 selectthe 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.)

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.

8. We’re going to create the same variables in Jamovi that we did last week in RStudio so we can practice some data manipulation in Jamovi.

9. 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. Data – Compute – name the variable BMI – formula: weight/((height/100)^2)

10. 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. Data – Recode – name the variable lvst4\_recode – source variable: lvst4 – using transform: = 6 - $source (see https://blog.jamovi.org/2018/10/23/transforming-variables.html)

11. Create a new variable in Jamovi to calculate a final score for the Love of Statistics Scale for each participant called LOSS\_total.

You will need to create a computed variable for this. Data – Compute – name the variable LOSS\_total – formula: SUM(lvst1,lvst2,lvst3,lvst4\_recode)

12. Create a new variable in Jamovi to calculate a final score for the Love of Statistics Scale for each participant called LOSS\_total.

You will need to create a computed variable for this. Data – Compute – name the variable FOSS\_total – formula: SUM(frst1,frst2,frst3)

13. Create a new variable in Jamovi to calculate a final score for the Love of Statistics Scale for each participant called LOSS\_mean.

You will need to create a computed variable for this. Data – Compute – name the variable LOSS\_mean – formula: MEAN(lvst1,lvst2,lvst3,lvst4\_recode)

14. Create a new variable in Jamovi to calculate a final score for the Love of Statistics Scale for each participant called FOSS\_mean.

You will need to create a computed variable for this. Data – Compute – name the variable FOSS\_mean – formula: MEAN(frst1,frst2,frst3)

15. Create a new variable in Jamovi to calculate a z-score for LOSS\_total called LOSS\_total\_z.

You will need to create a computed variable for this. Data – Compute – name the variable LOSS\_total\_z – formula: Z(LOSS\_total)

16. Create a new variable in Jamovi to calculate a z-score for LOSS\_mean called LOSS\_mean\_z.

You will need to create a computed variable for this. Data – Compute – name the variable LOSS\_mean\_z – formula: Z(LOSS\_mean)

17. Create a new variable in Jamovi to calculate a z-score for FOSS\_total called FOSS\_total\_z.

You will need to create a computed variable for this. Data – Compute – name the variable FOSS\_total\_z – formula: Z(FOSS\_total)

18. Create a new variable in Jamovi to calculate a z-score for FOSS\_mean called FOSS\_mean\_z.

You will need to create a computed variable for this. Data – Compute – name the variable FOSS\_mean\_z – formula: Z(FOSS\_mean)

19. Click on the Analyses menu at the top next to Data. Click on Exploration - Descriptives. The Descriptives analysis options will display on the left. Output will appear on the right.

20. Get descriptive statistics for the categorical variables by dragging them to the Variables window or clicking on them to select and hitting the right arrow button. (You can also remove variables by dragging them back out or selecting them and clicking the arrow which points to the left when you click on variables in the Variables window.) (Which variables are categorical variables? What descriptive statistics and plots are you going to get?)

Categorical variables: sex, research, lvst1, lvst2, lvst3, lvst4, lvst4\_recode, frst1, frst2, frst3.

For the categorical variables we want frequencies so click on the Frequency tables check box. (You should see new tables appear in the output.)

For categorical variables we don't care about descriptive statistics like means. Remove those from the output by clicking on the Statistics bar to open it. Uncheck the boxes for Mean and Median. You can close the statistics bar by clicking on it again.

You can get a bar plot for each variable by opening the Plots bar and clicking on the Bar Plot check box.

You can keep the output for the categorical variables by clicking on the hide options arrow (arrow pointing to right with circle around it). The Descriptive options will disappear and the data window will show again. The output on the right will scroll up.

21. Get descriptive statistics for the continuous variables. (Which variables are the continuous variables? What descriptive statistics and plots are you going to get?)

Continuous variables: height, weight, BMI, the new variablesDrag the variables to the Variables window.

Open the Statistics bar and check the boxes for Std. deviation and Range. (You'll see them appear in the output table.)

Open the Plots bar and check the boxes for Histogram and Box plot.

Keep the output by clicking on the hide options arrow.

22. You can get the R code for most of the analyses you do in Jamovi.

1. Click on the three vertical dots at the top right of the Jamovi window.

2. Click on the Syntax mode check box at the bottom of the Results section.

3. Close the Settings window by clicking on the Hide Settings arrow at the top right of the settings menu.

4. you should now see the R code for each of the analyses you just ran.

23. Copy the code for the descriptive statistics for the categorical variables in Jamovi and run the code block in RStudio.

Right click on the code block in Jamovi.

Select Syntax - Copy

Paste it in an R code block in your R notebook in Rstudio.

You may need to change the name of the dataset to match the name of the data frame in RStudio. data = week3 (or the name of your data frame)

24. Copy the code block for the descriptive statistics for the continuous variables and run it here in RStudio.

25. Some notes on Jamovi

Some advantages of Jamovi

Jamovi takes some of the difficulty out of remembering commands to run statistical analyses by providing point and click menus.

Jamovi is freely available to anyone (and pretty easy to install) so it's easy to share output. (You can also export to other formats.)

The jmv package collects basic statistical anayses into a single package so it makes learning statistics in R a bit easier.

The Jamovi output is formatted very nicely. The tables are pretty much APA compliant.

Some limitations of Jamovi

Jamovi does not allow you to make notes or write in the output like you can in RStudio. Interpreting output is as important as getting output in the first place. Making some notes to go along with what you analyze is pretty useful and important.

Jamovi does not currently have the wide variety of packages available which can be run in RStudio. It's a good idea to know both.

Jamovi is still undergoing significant development. There are some parts which may still be buggy.

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

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

a. LOSS\_total – interval, continuous

b. FOSS\_total – interval, cotinuous

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

a. LOSS\_total – histogram and box plot

b. FOSS\_total – histogram and box plot

29. 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

30. What units are z-scores in?

Standard Deviations

31. 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

32. 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

33. 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

34. 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

35. 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 the text of the .Rmd file with the output into a Word or pdf file and turn in that.

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

Love the repetition. It helps. Disliked the error I kept getting with the jamovi code in R studio until I figured it out

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

some typos with the new variables above

38. 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.)