

Quantitative Data Analysis DSO2502

Midterm assignment

To complete this assignment you will need to download a Zip archive from canvas, containing the following files:

Data	qda2020.master.csv
Codebook	qda2020.pdf
Assignment text	qda2020.midterm.assignment.pdf (this document)
Answer sheet	qda.midterm.000000.docx*

Your submitted assignment should include:

Word	Answer sheet, with your answers and output
R	The script you have used to address the questions in the assignment and produce the output.

Please compress these files into a Zip archive and upload it to Canvas/assignments by Monday, 28 September 2020, before 16:00.

Tips:

1. Use the R code we provided and substitute variables with the ones you want to analyze;
2. Have your codebook at hand, to make sure you understand the attributes.

[Click here for a quick reminder on how to compress files.](#)

Good luck!

Please use the answer sheet and answer all of the following five questions

1. Create a frequency table for the variable q12 and answer the following questions:
 - a. What is the level of measurement of this variable?
 - b. **How many** respondent mentioned that they are very confident that they were able to identify fake news?
 - c. What **percentage** of respondents said that they were (“very” or “somewhat”) confident that they were able to identify fake news?
 - d. *There were **a little more** “not very confident” respondents than “very confident” respondents.* True or False?
- * **Paste the frequency table in the answer sheet**

* Make sure to change the 000000 to your UM ID number before saving the answer sheet.

2. Create a sample of respondents between the ages of 18 and 19 and answer the following questions:
 - a. What is the level of measurement of the age variable?
 - b. **How many** observations are included in this sample?
 - c. What **percentage** of these 18-19 year-olds come across fake news on a daily basis?
 - d. What **percentage** of these 18-19 year-olds are (very or somewhat) confident that they are able to identify fake news?

3. Go back to the main data frame and create a bar chart. The chart should show the **numbers of respondents** owning **each of the following**: Desktop, laptop, mobile phone, internet connection at home and a tablet computer; in addition add one bar for respondents who do not own any of these devices. The graph should have the following characteristics:
 - The plot should have a **main title** reading "Computing devices ownership", followed by **your name and UM ID number**;
 - The **x-axis** should be labelled "Number of devices";
 - The bars should be **red**.
 - * **Paste the chart in the answer sheet**

NB: To create this chart you will need to run frequency tables for all device-ownership variables (including owning none), write down the number of respondents owning each of these devices, and create a new vector with these values, to plot.

4. Create a boxplot for the number of sources respondents used, broken down by their desktop computer ownership (D1a). Which of the following is NOT true?
 - a. The minimum number of sources is **higher** for people who **own** a desktop computer.
 - b. The median number of sources is **higher** for people who **own** a desktop computer.
 - c. Outliers were identified for respondents who **owned desktop computers** but not for respondents who didn't.
 - d. 75% of all respondents reported using **five sources or less**.
 - * **Paste the chart in the answer sheet**

5. Create a crosstabulation of gender by Sum of devices R has. NB: your crosstab should present percentages, summarized for the row (gender). Answer the following questions:
 - a. What are the levels of measurement of each of these variables?
 - b. What percentage of men own all five devices? What percentage of women do?
 - c. What was the mode number of devices for men? And for women?
 - d. *Men are more likely to own 4 devices than woman are.* True or False?
 - * **Paste the frequency table in the answer sheet**