HW2: Class One Survey Analysis

INSTRUCTIONS

Please complete tasks 1-10. Use R syntax to solve all problems (i.e. do not manually change values/variables in datasets or check them manually to make sure things are correct except as a double or triple check). Clearly label what your syntax does for each problem. For example, for problem 2a., include a comment that preceeds or immediately following the syntax similar to or exactly like this: *#2a. Determine the number of observations in the C1survey dataframe.* Save your R script file and submit it on blackboard under the HW2 assignment.

1. Import class one survey data from our Github site (The dataset called 'Class_One_Survey.csv' is located in the class one folder), calling the R dataframe that you create *C1survey*.

2. Determine a. the number of observations (i.e. the number of people who filled out the survey) and b. the number of variables in the dataframe.

3. Generate and display a list of column names, calling your list *varlist*.

4. a. Rename the column variables to something shorter and that is descriptive of what the variable is about (for example *like_dogs* for the 'Do you like dogs?' question variable) and b. write code to display that they are renamed.

5. Write code to determine and display the number of factor, integer, numerical, and character variables there are in the *C1survey* dataset.

6. a. Using code, check the *height* and *weight* variables for any unusual or missing values. If you find any, b. describe what you will do with the unusual values in a comment before or immediately following the code that does it and c. after you have cleaned up any unusual values, find the mean height in cm and weight in kg.

7. a. Create new variables called *weight_kg* and *height_m* that gives weight in kg instead of grams and height in meters instead of cm and b. determine the mean weight in kg and the mean height in meters.

8. a. Derive a BMI variable (kg/m2) from the height_m and weight_kg variables called *BMI* (adding it to the C1 survey dataset). b. Determine the median BMI. c. Make another variable called *BMI_cat* (adding it to the C1 survey dataset) that divides BMI into >median and <=median BMI. d. Label the levels. e. Determine how many people are in each category of *BMI_cat* using the table function.

9. a. Create a new dataset called *C1survey_BMI_below* that includes only individuals with BMI below the median and b. write code to check that your dataset only includes individuals with BMI below the median.

10. Pick your favorite variable to analyze, come up with a question you want to answer with that variable, generate the code, and provide an answer to your question. Describe what you did using comments (i.e. #'s).