N736 Homework 01

Gaea Daniel

September 6, 2017

## N736 Homework 01 - Due Sept 6, 2017

1. Using your statistical software of choice, read in the dataset "Dataset\_02" from the \*.CSV, \*.XLS or \*.XLSX format.

#Import data after edit (q2 input shifted up by one cell)  
library(readxl)  
Dataset\_02\_Edit <- read\_excel("~/RStudio Stuff/N736Testdataset/Dataset\_02\_Edit.xlsx")  
View(Dataset\_02\_Edit)

1. Create an analysis report that explains **every step you did** - including anything you did manually. Ideally, everything you do to the dataset for fixing the errors or identifying the issues you spot in the data should be done using CODE (aka SYNTAX) so that someone else can take your code and reproduce your results. *[Remember, that "someone else" might be you in a few weeks when you've forgotten what you did earlier...]*

Corrections made manually in RCmdr: GenderSTR formatting => all changed "f" or "m" q4 - outlier 40 => changed to 4 q1 - outlier 11 => changed to 1 q2-q5 - contains 9's => changed to NA GenderCoded - missing values WeightPRE - outliers 68, 60 (typo or lbs vs kg) => changed 68 to 168 and 60 to 160 Height - 2.6 => changed to 6.2 Age - outlier 99 => changed to NA 9's could indicate missing, refused to anser, N/A, I don't know q2 - shifted down by 1 cell => shifted up by 1 cell q3,4,5 missing for last 3 submects (skipped items maybe on back or swithched to short form) All missing data changed to NA

1. Provide comments and an explanation in your analysis report and in your code for each issue you find and fix.

Corrections: GenderSTR formatting => all changed "f" or "m" for consistent formatting q4 - outlier 40 => changed to 4 because the scale is 1-5; data entry mistake assumed q1 - outlier 11 => changed to 1 because the scale is 1-5; data entry mistake assumed q2-q5 - contains 9's => changed to NA for missing data since 9's are commonly entered for missing data GenderCoded - missing values changed to NA WeightPRE - outliers 68, 60 (typo or lbs vs kg) => changed 68 to 168 and 60 to 160; data entry mistake assumed Height - 2.6 => changed to 6.2; data entry mistake assumed considering the weight Age - outlier 99 => changed to NA for missing data 9's could indicate missing, refused to anser, N/A, I don't know q2 - shifted down by 1 cell => shifted up by 1 cell q3,4,5 missing for last 3 submects (skipped items maybe on back or swithched to short form); kept in the dataset because more than half of the information was provided All missing data changed to NA

1. HINT: There are 10 issues to identify or problems to fix in this dataset. One of the issues might have to be done manually, but you should still explain what you did to fix it, even if you did it manually.
2. Turn in - upload to CANVAS no later than Sept 6, 2017 11:59pm:
   1. An analysis/data preparation report (in either DOC or PDF format) - don't forget to format your report nicely, make sure your name and date is included and
   2. your code/syntax with your comments and commands used to reproduce your results (you can upload your code in either \*.R, \*.sas, \*.sps or \*.txt (i.e. plain text) formats)