INFO 201
Software for Data Analysis
Module 2 Homework A

Instructions

A. Download the data file *bank* from ELO. *Bank* lists 2,500+ payment amounts over a 5-year period, and is conceptually similar to the *transactions* data file we used in class for Case Study II.

B. Modify your Case Study II code to perform a similar analysis for the *bank* data file that we performed in class for the *transactions* file.

C. Hint: The sections of your code will be reasonably similar to Case Study II. While you can organize your code in any way you want, you may find the following headings to be helpful:

1. Plot original data by transaction amount
2. Compute and plot expected Benford distribution
3. Compute actual distribution by first digit
4. Join and plot actual vs. expected distribution
5. Additional analysis by year
6. Rank Chi2 by year
7. Detailed investigation for one year (identify at least one year that could be problematic)
8. New section: Create one new block of code to identify any potentially problematic transactions during the year you identified

D. Save your code as *yourlastnameHW2A*, and submit to ELO by the due date/time.

E. This homework assignment will be scored as follows:

1. Subsections C.1 - C.6 will be worth 6 points. To earn these 6 points, your code will be tailored correctly to the new dataset, run smoothly with no errors, and produce the correct output.
2. Subsections C.7 and C.8 will be worth 6 points. This part will be more challenging than the first part above. Your analysis can take the form of at least one substantive additional numerical table and/or additional graphic. You are not required to create any new variables (other than those already created by our code in class), and you are not required to prove the existence of fraud. However, your additional analysis should help the reader better understand the year in question, beyond the analysis we performed in class. Please add a 1-2 sentence comment directly above your new code to describe your analysis and conclusions.

Honor Code: You are allowed to discuss this homework assignment with other students in the course, and you are allowed to work on this homework assignment side-by-side in the computer lab with other students. However, each student must type and submit their own code. You may not submit code that was partly or entirely typed by another student, which would be an Honor Code violation that will result in a course grade of 'F.'