Homework 1

Anna Ziff

R Workflow for Economists

1. Define two variables: $x = 0.2 - 0.1$ and $y = 0.3 - 0.2$. Using `==`, test if x and y are equal. What do you get? Read the documentation for the function `all.equal()`. Now use this function to test x and y are equal. What does the `tolerance` argument allow for? This issue is called a floating point issue and stems from how computers store doubles.

2. Read about `rnorm`. Create a vector with 100 random draws of the the normal distribution with mean 0 and standard deviation 1. Read about `rbinom`. Create another vector with 100 random draws of the Bernoulli distribution with a probability of success of 0.5. (Be mindful of the `size` argument). Calculate the means of these two variables.

3. A useful function to create factors is `gl()`. Read the documentation and create an ordered factor with 4 values, each repeated 20 times. The labels should be `"Freshman"`, `"Sophomore"`, `"Junior"`, `"Senior"`. (These are the names of the four years of high school/college in the U.S.)

4. Write a function for question 1. It should take as an input the number of random draws, the type of distribution, and any parameters associated with the distribution, and should output the mean of 100 random draws. Using this function and an apply-like function (or loop), redo the previous question 100 times. Output one vector (of length 100) of the means of the normal random draws and another vector (of length 100) of the means of the Bernoulli random draws.

5. Random numbers are chosen based on a random number generator that takes as input some seed. Edit your function from question 2 to have an additional argument be the seed. Set the seed in the body of your function using `set.seed()`. Confirm that your code can be replicated. That is, confirm that if you run the code with the same seed, you will get the same result.

6. The U.S. Census publishes the Public Elementary-Secondary Education Finance Data collected through the Annual Survey of School System Finances (PESEFD). The 2019 dataset and codebook are available in this Dropbox folder. Import the data to R. How many observations and variables does the dataset have?

7. Only keep the following variables: IDCENSUS, NAME, ENROLL, TOTALREV, TFEDREV, TSTREV, TLOCREV, TOTALEXP, PPCSTOT. Change the variable names so that they are lower case. Only keep observations with `enroll` greater than 0. Sort the dataset in descending order by enrollment. How many observations are there now? What are the 5 school districts with the largest enrollment in the data?

8. The dollar amounts are from 2019. For all variables with dollar values, inflate the dollar values to 2021 dollars (approximate that \$1 in 2019 is worth \$1.05 in 2021).

9. The National Center for Education Statistics publishes information on the Common Core of Data (CCD). The 2019 dataset is available in this Dropbox folder. Import the data to R. How many observations and variables does the dataset have? What does the warning message tell you about the variable `UNION`? How can you change how you read your data to avoid this issue?

10. If you wanted to merge the CCD and PESEFD datasets, what challenges might you encounter? This is a hypothetical question. That is, do not actually attempt to merge the datasets.