IST687 – Samples HW

Let's continue our exploration of sampling.

Step 1: Write a summarizing function to understand the distribution of a vector

1. The function, call it 'printVecInfo' should take a vector as input
2. The function should print the following information:
    a. Mean
    b. Median
    c. Min & max
    d. Standard deviation
    e. Quantiles (at 0.05 and 0.95)
    f. Skewness

    Note for skewness, you can use the function in the 'moments' library.
3. Test the function with a vector that has (1,2,3,4,5,6,7,8,9,10,50). You should see something such as:

    [1] "mean: 9.54545454545454"
    [1] "median: 6"
    [1] "min: 1 max: 50"
    [1] "sd: 13.7212509368762"
    [1] "quantile (0.05 - 0.95): 1.5 -- 30"
    [1] "skewness: 2.62039633563579"

Step 2: Creating Samples in a Jar

4. Create a variable 'jar' that has 50 red and 50 blue marbles
   (hint: the jar can have strings as objects, with some of the strings being 'red' and some of the strings being 'blue'
5. Confirm there are 50 reds by summing the samples that are red
6. Sample 10 'marbles' (really strings) from the jar. How many are red? What was the percentage of red marbles?

7. Do the sampling 20 times, using the 'replicate' command. This should generate a list of 20 numbers. Each number is the mean of how many reds there were in 10 samples. Use your printVecInfo to see information of the samples. Also generate a histogram of the samples.
8. Repeat #7, but this time, sample the jar 100 times. You should get 20 numbers, this time each number represents the mean of how many reds there were in the 100

samples. Use your printVecInfo to see information of the samples. Also generate a histogram of the samples.

9. Repeat #8, but this time, replicate the sampling 100 times. You should get 100 numbers, this time each number represents the mean of how many reds there were in the 100 samples. Use your printVecInfo to see information of the samples. Also generate a histogram of the samples.

Step 3: Explore the airquality dataset

10. Store the 'airquality' dataset into a temporary variable
11. Clean the dataset (i.e. remove the NAs)
12. Explore Ozone, Wind and Temp by doing a 'printVecInfo' on each as well as generating a histogram for each