**IST687 – Sampling 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:
   1. Mean
   2. Median
   3. Min & max
   4. Standard deviation
   5. Length
3. Test the function with a vector that has (1,2,3,4,5,6,7,8,90,91,100). You should see something such as:

[1] "mean: xxx"

[1] "median: 6"

[1] "min: 1 max: 100"

[1] "sd: xx"

[1] "Length: 11”

**Step 2: Explore the airquality dataset**

1. Store the ‘airquality’ dataset into a temporary variable
2. Clean the dataset (i.e. remove the NAs)
3. Sample the ozone within the dataset 10 times, replacing the sampled item after each sample. Use your printVecInfo to see information of the samples. Also generate a histogram of the samples.
4. Sample the ozone again (same as described in the previous step), but this time, compute the mean of the sample.
5. Now, do the sampling 20 times, using the ‘replicate’ command. This should generate a list of 20 numbers. Each number is the mean of the 10 ozone samples. Use your printVecInfo to see information of the samples. Also generate a histogram of these means.
6. Repeat the previous step, but this time, sample the ozone 100 times. You should get 20 numbers, this time each number represents the mean of the 100 ozone samples. Use your printVecInfo to see information of the samples. Also generate a histogram of the samples.
7. Repeat the previous step, but this time, replicate the sampling 1000 times. You should get 1000 numbers, this time each number represents the mean of the 100 ozone samples. Use your printVecInfo to see information of the samples. Also generate a histogram of the samples.
8. How does the distribution of the Ozone samples change during each of the previous steps? Which more representative of the actual Ozone values?