Assignment 2 -Normalization methods

* 1. Generate 3 variables, 10000 samples each
     1. B: Gaussian mean 5 sd 2
     2. I : Power law , use scipy.stats.powerlaw.rvs with a= 0.3,
     3. H : Geometric p= 0.005
     4. You are encouraged to change the parameter values and observe change in shape, skew etc.
  2. Compare above variables in single box plot
  3. Try following normalization methods
     1. Divide each variable by max
     2. Divide variable by sum of its values
     3. Convert each variable into z score using respective mean and sd
     4. For each variable , convert the values in percentiles
     5. Make medians of all the variables same
        1. Calculate median of each variable
        2. Calculate mean of these medians say m1
        3. Generate a multiplier for each of the variables so that median value of each variable becomes m1
     6. Quantile normalize the data using off the shelf library function
  4. For each of the above method,
     1. Compare original distribution with its normalized version in single histogram
     2. Compare all the normalized variables in single box plot