**PROJECT 3**

**\*\*File to project 1 is attached, which you may or may not use.**

**College course: Statistics for Information Science**

**Watch the video for project 3 in which the data generation is discussed and the coding is explained step by step. (Link to video) \*\* details to access video in red text below**

**Using the dataset generated for you (as explained in the video), solve the following problems using the code given in the video (or you can write your own code if you like, it's your choice).**

**Use this answer sheet and rename it by replacing "testStudent" with the idName in your assigned csv file. (project3\_kkamara3.csv):**

[**project3\_testStudent.csv**](https://umd.instructure.com/courses/1288442/files/59838534/download?wrap=1) **(File attached)**

**Submit a zip file containing this csv answer sheet together with the R code you write based on the video (or your own code if you like).**

**\*\* ACCESS TO VIDEO**

**link to video**

**<https://umd.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=667c4dcb-8c87-47fa-8627-ac80000e1bfd>**

**Login required to watch video**

**Username: kkamara3**

**Password: Saloneboy97**

**\*\* after logging in, authentication method will pop up. Click on “Duo Push”. Afterwards I will allow authorization. Otherwise, if nothing happens, contact me at this number +1 240-353-8291 (WhatsApp preferred) so that I can give authorization to continue.**

**\*\*All files mentioned in the video are attached, contact me if anything is missing**

**QUESTIONS**

**Define group 1 to be all songs before 1980 (so, 1979 and earlier), and group 2 to be songs from 1990 onward. For brevity of expression, in this assignment we call group 1 "old songs" and group 2 "new songs". Throughout, we use significance level alpha=5%.**

Q1: calculate mean length of group1 songs minus mean length of group2 songs.

Q2: with the same order (group1 minus group 2), calculate the difference between median song lengths.

Q3: find difference between the mean vocabulary size of the two groups.

Q4: find the difference between median vocabulary size of the two groups.

Q5: We want to test the hypothesis: average length of old songs is different from average length of new songs.

conduct a permutation test with 50'000 permutations. What are the critical values?

Q6: We want to test the hypothesis: average length of old songs is less than average length of new songs.

conduct a permutation test with 50'000 permutations. What is the critical value?

Q7: We want to test the hypothesis: median length of old songs is less than median length of new songs.

conduct a permutation test with 50'000 permutations. What is the critical value?

Q8: Now let's focus on vocabulary size instead of song length.

We want to test the hypothesis: median vocabulary size of old songs is less than median vocabulary size of new songs.

conduct a permutation test with 50'000 permutations. What is the critical value?

**Now define a new variable, which we call "word density", to be: s o n g v o c a b u l a r y s i z e /s o n g l e n g t h**

Q9: We want to test the hypothesis: median word density of old songs is different from median word density of new songs.

conduct a permutation test with 50'000 permutations. What are the critical values?

Q10: We want to test the hypothesis: average word density of old songs is different from average word density of new songs.

conduct a permutation test with 50'000 permutations. What are the critical values?