Jacob Sriraman

CSCI 136

Benford.docx

After completing the Benford.py code and running in on the four provided text files, I have gathered the following results for each of the data sets:

Test Data:

0s: 0, 1s: 3, 2s: 4, 3s: 1, 4s: 0, 5s: 2, 6s: 1, 7s: 0, 8s: 0, 9s: 0

This is a very small data set, so it is harder to make an accurate assumption, but it is safe to say that it roughly follows Benford’s law, as 1 shows up around 30% of the time, and the larger digits are less common. Therefore, we can conclude that this data is naturally occurring.

Library Books:

0s: 0, 1s: 3056, 2s: 1606, 3s: 1018, 4s: 801, 5s: 640, 6s: 560, 7s: 502, 8s: 503, 9s: 452

This data set is much larger than the previous one. Again, 1 appears around 30% of the time, and we can see that each digit appears less than the digit smaller than it (excluding 0), which once again shows that this is naturally occurring data that follows Benford’s law.

Live Journal:

0s: 0, 1s: 982, 2s: 276, 3s: 30, 4s: 38, 5s: 50, 6s: 91, 7s: 94, 8s: 121, 9s: 197

Despite 1 showing up often in this data set, this is an unnatural data set, because 1 occurs over 50% of the time, and the data is skewed in a bowl shape, instead of heavily to the left. We see this by observing that digits 3-9 increase likelihood as they get larger. This data does not follow Benford’s law.

Sun Spots:

0s: 87, 1s: 868, 2s: 369, 3s: 307, 4s: 318, 5s: 305, 6s: 257, 7s: 193, 8s: 196, 9s: 173

This data set is another large one, and 1 shows up close to 30% of the time, which each larger digit smaller than the last, which supports the conclusion that this is another naturally occurring data set, that follows Benford’s law.

For my own data source from the web, I am using a list of number of births from every county in the United States from 2016 through 2021. I found my data on the CDC Wonder database, which is an open-source, public database. There are 625 counties that had data available. I compiled these numbers into a text file to be analyzed. This data can be accessed at <https://wonder.cdc.gov/> .

Birth Counts:

0s: 0, 1s: 266, 2s: 130, 3s: 70, 4s: 34, 5s: 36, 6s: 24, 7s: 18, 8s: 22, 9s: 25,

After running the text file through my code, it is safe to say that this data set follows what we would have expected based on Benford’s law. The digit 1 occurs a little bit over 30%, with 2 at a rate half of that, and each digit after that getting smaller for the most part.

One more note: the test code gave me 0/2 for the output of my main function, but when I used the Anaconda Prompt command line, I was getting the correct output in the correct format, so I figure that my code is correct.