MTH 109: Benford's Law Project

Due October 12, 2015

You will complete this project in groups of 4 or less. You may **not** have a group of more than 5 people. This project consists of four parts and should be typed. The project is worth 220 points and each person in the group will receive the same grade as the others in the group.

Each group needs to email me on the following dates to let me know the progress of the project:

- October 2 by midnight
- October 9 by midnight

Only one person from the group needs to email me. Each email will be worth 10 points. In each email make sure to include the names of everyone in your group.

Part 1

Do some research on Benford's law. There is an online book available through the library. The title of the book is "Benford's Law" and is written by Steven J. Miller. The first chapter of this book might be a good place to start your research. After doing the research, your group will write a one page summary (double-spaced) about Benford's law. Make sure to include citations. There is no required style of citations, just be consistent.

Part 2

In this part you should collect two real-world datasets and determine if Benford's law applies to it. Make sure it is an appropriate dataset to use. In particular, the dataset should be large (at least 100 observations) and shouldn't be something where certain digits never can occur. For example, you should not use a dataset which is the population of cities that range from 3,000 to 9,000 as the number "1" will never appear as the first digit. Some possible choices for this section include: populations of countries in the world or atomic masses of elements. You can use one of these choices, but not both.

For each dataset, calculate the number of times $1, 2, 3, 4, 5, 6, 7, 8, 9$ appear as the first digit. If you are using Excel to do this part, you can use the command INT(LEFT(X,1)) where X is the cell that you want to know what the first digit is. My suggestion would be to have one column with the actual data and a column next to it with this command for each cell. Once you have calculated how many times each number appears, determine what proportion of observations have each digit as its first digit. Compare this with what is predicted by Benford's law. Make sure you include what dataset you are using and cite where you obtained this dataset.

Part 3

In this part you will do the same as in part 2, except you will use mathematically defined sequences instead of real-world data. You need to choose two datasets which are defined mathematically. For example, you could choose to look at squares of positive integers. Thus, the first six numbers of the sequence would be $1, 4, 9, 16, 25, 36$. Another interesting example to look at is the Fibonacci sequence. It is completely up to your group which sequence to use, you don't have to do either of the examples I have given here. Your sequences should have 100 data points. So for the squares example, your list will consist of the numbers 1 through 100 squared.

Part 4

In this you will once again determine if Benford's Law applies. You will create a list of 100 random numbers between 10 and $100,000$ and determine if Benford's Law applies. You can do this in Excel by using the command RandBetween(10,100000). Explain why or why not Benford's Law should apply in this situation.