

Report

By: Tyler McElroy

The purpose of this analysis is to apply Benford's Law on several Healthcare companies with the goal of identifying potential fraud. According to Benford's Law, a dataset column with naturally occurring numerical features is expected to have lower digits as the leading digit more frequently than higher digits. The expected distribution is as follows.

Leading Digit Expected Percentage

Leading Digit	Expected Percentage
1	30.1%
2	17.6%
3	12.5%
4	9.7%
5	7.9%
6	6.7%
7	5.8%
8	5.1%
9	4.6%

Within each Excel spreadsheet, I analyzed the three numeric-based columns to identify deviations from Benford's Law -- that is, I analyzed data to see if the probability distributions for leading digits differed significantly from the expected percentages listed above. The three numeric-based columns are as follows: (1) aggregate number of cases by day, (2) total cost per testing by day, and (3) total insurance claims billed by day.

Refer below for sections describing the results for each column and recommendations regarding further action.

1.) aggregate number of cases by day:

The column "aggregate number of cases by day" has several deviations from the expected percentages for each of the three companies. For Advance Genomic Diagnostics, the observed frequency of leading digits contains spikes for the digits 3 and 7; additionally, digit 2 contains substantially less observations than expected by Benford's law. For Exelonixx Labs Inc., the observed frequency of leading digits contains spikes for the digits 4, 5, 6, and 8. And lastly, California Molecular Testing Inc. roughly follows the expected percentages detailed by Benford's Law; however, there is a spike of observations for leading digits 8 and 9.

2.) total cost per testing by day:

The column "total cost per testing by day" differs significantly from the expected percentages for each of the three companies. Each company has leading digits that are completely absent within the data. Also, the companies Advance Genomic Diagnostics and Exelonixx Labs Inc. contain many observations for leading digits 1 and 2; however, observations largely decrease after that with some digits having no observations at all.

Note: digit 5 contains a noticeable increase for the company Exelonixx Labs Inc. Lastly, the company California Molecular Testing Inc. only contains observations for digits 2, 3, 4, and 5. Note: digit 5 is the only digit that conforms to the expected percentage value as outlined in Benford's Law.

3.) total insurance claims billed by day:

The column "total insurance claims billed by day" roughly follows the percentages expected by Benford's Law. As such, it is unlikely that the numbers in this column were modified by any of the three companies mentioned in this analysis. No further investigation is needed for the column "total insurance claims billed by day."

4.) Recommendations:

Benford's Law is used as a screening tool to identify potential fraud. As such, other audit and risk management work will need to be performed to reach a final verdict. Given the analysis contained within this report, my recommendations are as follows:

1.) Identify why several dates are duplicated within each dataset.

2.) Identify why several values are missing within each dataset (See: Blank Values, NA, N/A).

3.) Investigate each of the three companies further -- especially regarding the columns "number of cases per day" and "total cost of testing by day." It is suspicious that these columns do not follow the expected percentage distribution, but the column "total insurance claims billed by day" does follow the expected percentage distribution.