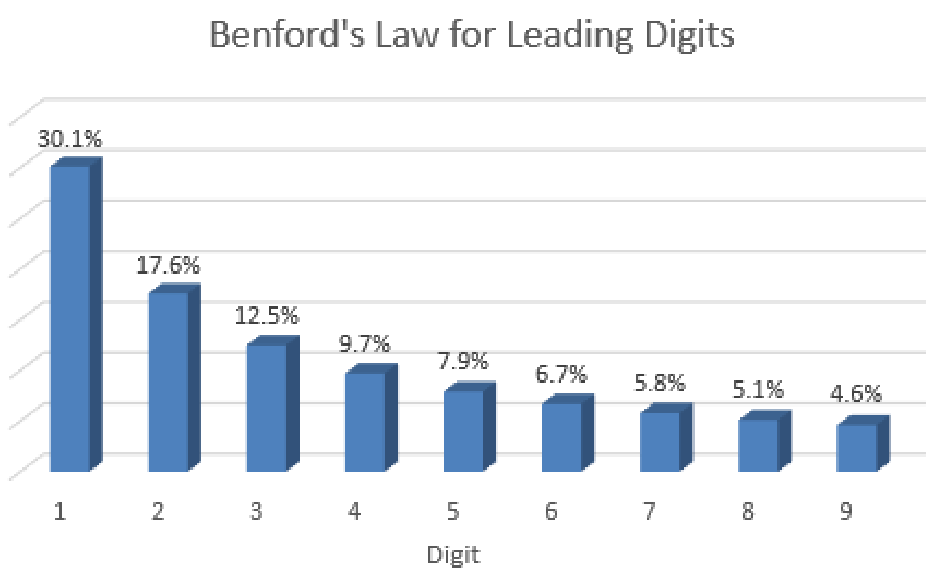
Daniel Leal
April, 2025

Exploring Benford's Law in Everyday Data

Let's imagine you open up your favorite social media app, like instagram or youtube. As you scroll though your feed, you notice the number of likes in each post and decide to keep track of all the likes. Specifically the leading digits which are first digits when reading a number from left to right. At first glance, looking at all the numbers, you might think that the data collected from your feed would be random with no distinguishable pattern. However, extensive research done by analytical experts as well as a personal research I conducted reveal a surprising trend. After collecting data and analyzing every leading digit, we find that likes beginning with a digit of 1 appear far more often than any other digit. With a chance of appearing of 30%. This is followed by the digit 2 with a 17% chance of appearing so on and so forth, with bigger leading digits having a smaller chance of appearing. Thus, the number with the lowest chance of appearing belongs to number 9 with only a 5% chance. Challenging the assumption of randomness within a dataset. In this document we will further explore this phenomenon, its history and how it is used in our day to day lives.

As stated before, when we go through each leading digit from 1 through 9; The chances of smaller leading digits appearing is greater than the next number. This is particularly true with datasets containing numerical data in large quantities, a wide range of values and being uniformly distributed. You can refer to the graph below to better understand how benford law behaves and what are the chances for each leading digit to appear ranging from 1 through 9:



Credit: Amanda Montañez; Source: "Note on the Frequency of Use of the Different Digits in Natural Numbers," by Simon Newcomb in American Journal of Mathematics, Vol. 4, No. 1; 1881 (data)

WHO DISCOVERED BENFORD'S LAW?

Benford's law which is also known as the first digit law has been mathematically proven time and time again thanks to all the mathematicians who devoted time and effort on exploring this unique phenomenon which we were able to visualize thanks to the graph above. The first person to point out this phenomenon was an astronomer and applied mathematician named Simon Newcomb in the year 1881. The story goes that Simon had a book filled with data, and while studying, he noticed that the tables in which he used to keep all of the data organized seemed to contain a leading digit of 1 far more frequently than any other larger, leading digits such as a seven, eight or nine. Though his theory had a strong foundation and evidence to prove his initial solution, no one really cared to investigate further and come to a more definitive conclusion. Or at least it was true until an American engineer and physicist named Frank Benford decided to test Newcomb's theory in 1938 with a survey/research he made named "The Law of Anomalous Numbers" consisting in collecting more than 20,000 data points from a variety of sources such as the molecular weight of chemical compounds, address numbers, population size among other things. By the end of his research, he concluded that he had sufficient proof to deem Newcomb's theory valid and true. Later being backed up by more brilliant minds such as Theodore Hill in 1995 when he also concluded that the "numbers in data series following Benford's Law are, in effect, second generation distributions, i.e. combinations of other distributions." [1]

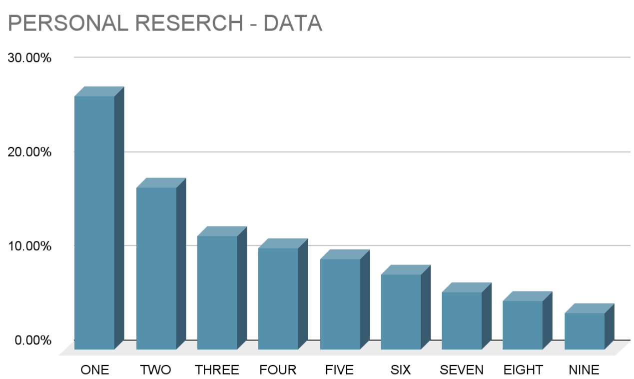
THE RESEARCH:

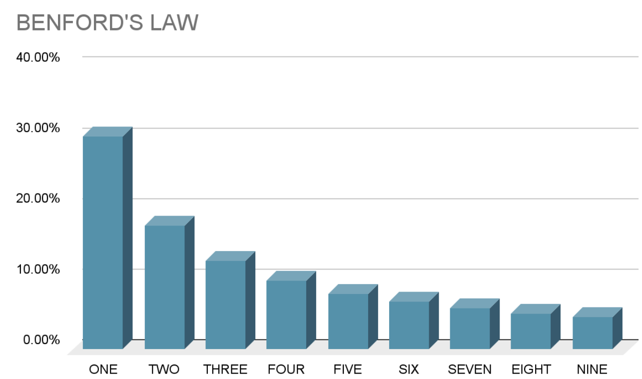
After learning about the origins of Benford's law, and how researchers have collected data which shows a very unique pattern within datasets. I wanted to see if I can collect data in a personal research project to see Benford's law in full effect for myself. I embarked on a journey motivated to see results for myself. First I had to think about a data pool containing data in large quantities, a wide range of values and being uniformly distributed. Which is how we landed on collecting data from an instagram feed. Taking note of the leading digit each likes has. Once the data sample is large enough, percentages for each leading digit will be taken using an algorithm in a self made application. This app will be able to show the total number of leading digits, how they are distributed and the comparison to actual Benford's law statistics.

–Go to following page for results–

[1] Sarkar, T., & Urwin, M. (2025, March 6). What is Benford's law and why is it important? Built In.

After hours of scrolling through my instagram feed and collecting over 1,000 leading digits. The data is plugged into a graph to see if our perusal dataset follows Benford's natural progression. Looking at the graphs below, you can see the similarity between our own research at the left hand side. And Benford's law at the right. Including the percentages obtained in the tables.





– Refer to graphs below in order to find specifics and percentages –

LEADING DIGITS	**ONE**	**TWO**	**THREE**	**FOUR**	**FIVE**	**SIX**	**SEVEN**	**EIGHT**	**NINE**
TIMES APPEARED	294	189	133	120	104	85	69	55	42

My Percentages:	26.9%	17.3%	12.2%	11.0%	9.5%	7.8%	6.3%	5.0%	3.9%
Known Percentages	30.10%	17.60%	12.50%	9.7%	7.9%	6.7%	5.8%	5.1%	4.6%

RESEARCH CONCLUSION AND HOW IT TIES TO OUR DAY TO DAY LIVES:

Based on my extensive research, I was able to see how Benford's law is part of our day to day lives even if we don't know it's there. Something as simple as the likes on your instagram feed will follow this peculiar mathematical law. The functionality of this unique law however; does not stop at being able to predict the percentages of leading digits within a dataset. There are real life examples of this law being used to be part of something for the greater good of society. Data Analysts have built algorithms used by accountants and financial institutions in order to easily detect fraudulent activity. Certain algorithms are used to monitor income tax, company sales, payment methods and monitor account activity to avoid any illegal activity involving money. Let's say for example; that someone is trying to commit fraud by cashing void checks. We can assume that this person is most likely to write checks with large quantities of money rather than write small checks for $100 or $200. If that person were to do that enough times then the natural order of numbers will be altered. Thinking back to how Bendford law behaves, We know that a large leading digit such as 9 only has a 5% chance of appearing. So if there are many checks being written worth $900 compared to a check for $100. The algorithm will probably be able spot certain inconsistencies and flag such activity as possible fraud thanks to the numbers not following their natural order. [2]

NOTE: If you would like to collect data on your own and see if Benford's law follows your personal research, please refer to this link in order to download our data collection app in the play store.

FINAL THOUGHTS

Benford's Law reveals a fascinating and counterintuitive pattern that can be found in large data pools across a diverse number of fields. This unique pattern was first discovered by Simon Newcomb in 1881. With Frank Benford further proving its legitimacy in 1938 as he further expanded on research regarding Benford law. In today's day and age, we can observe this pattern when collecting Instagram data such as likes or views on posts. We have demonstrated that seemingly random numbers align with how Benford's law behaves basen on our research. The results highlight not only the significance of this mathematical law in theoretical terms but also its practical applications in everyday life for example in finance and accounting fields where accountants use Benford's Law to detect anomalies and flag possible fraudulent activity occurring within bank accounts. Increasing the integrity of the data and ensuring a better secure system when it comes to detecting fraudulent activity. Thanks to the overlap in mathematics, data analysis and real-world implications we can firmly state that Benford's Law is part of our daily life. This simple algorithm will most likely remain a key tool when it comes to breaking down and examining complex datasets, even when we don't see it but know it's there.

[2] Eckhartt, G. M., & Ruxton, G. D. (2023, April 11). Investigating and preventing scientific misconduct using Benford's law. Research Integrity and Peer Review, 8(1).

SOURCES:

- Sarkar, T., & Urwin, M. (2025, March 6). What is Benford’s law and why is it important? Built In. https://builtin.com/data-science/benfords-law
- Simkin, M. G. (2011, June 27). Using spreadsheets and Benford’s law to test accounting data. ISACA Journal, 3. https://www.isacajournal-digital.org/isacajournal/2011vol3?article_id=1078244&pg=NaN#pgNaN
- Eckhartt, G. M., & Ruxton, G. D. (2023, April 11). Investigating and preventing scientific misconduct using Benford’s law. Research Integrity and Peer Review, 8(1). https://doi.org/10.1186/s41073-022-00126-w
- Clark, H. (2010, December 1). The history and application of Benford’s law. Boise State University. https://scholarworks.boisestate.edu/cgi/viewcontent.cgi?article=1008&context=mss_2020#:~:text=o%20Benford's%20Law%20is%20also,numbers%20as%20a%20leading%20digit.&text=o%20Simon%20Newcomb%20was%20for,decided%20to%20test%20Newcomb's%20theory.
- Murtagh, J., & Bryner, J. (2023, May 8). What is Benford’s law? Why this unexpected pattern of numbers is everywhere. Scientific American. https://www.scientificamerican.com/article/what-is-benfords-law-why-this-unexpected-pattern-of-numbers-is-everywhere/
- Smith, S. W. (1997). The scientist and engineer's guide to digital signal processing: Chapter 34: Explaining Benford's law. DSPGuide. https://www.dspguide.com/pdfbook.htm
- Mark J. Nigrini (1999, April 30) *I've Got Your Number* journalofaccountancy https://www.journalofaccountancy.com/issues/1999/may/nigrini.html