Do Pandemic Related Datasets with High Artificial Control Still Follow the Benford's Law?

C. Kalpani Dissanayake
Assistant Professor, Department of Business and Economics,
Pennsylvania State University,
25 Yearsley Mill Rd., Media, PA 19063 USA
ckd5250@psu.edu

Jay Daniel
Program Leader, MSc Global Operations and Supply Chain Management
Derby Business School, University of Derby
Kedleston Rd, Derby DE22 1GB, UK
j.daniel@derby.ac.uk

Abstract

Benford's Law (BL) is being used extensively in research for several purposes including for the detection of potential manipulations of the data to detect fraud since datasets tend to follow the Benford's distribution when they occur naturally without artificial control. The COVID-19 pandemic has heavily impacted business and non-business-related activities. Datasets related to the pandemic are being used in many different analyses to arrive at different conclusions. However, the credibility of the results and conclusions depend heavily on the accuracy of the datasets. The COVID-19 related datasets are obvious results of intense human intervention and artificial control efforts; therefore, the question arises as to whether Benford's analysis can still be used to detect anomalous datasets among them? This research uses several publicly available datasets and uses predictive analytics to perform the Benford's analysis. The applicability of BL is first verified using a regular dataset occurred prior to the pandemic, and then applied on COVID-19 related datasets to test the research hypothesis. The results demonstrate that even the datasets with sufficiently large sample sizes with considerable human intervention and artificial control follow the Benford's distribution and that Benford's analysis can still detect the anomalous datasets. The findings are anticipated to be useful for the data analysts and researchers and adds to the current literature gap. This paper may also serve as a class case study for the academia teaching data analytics.

Keywords
Benford's Law, Data Anomaly Detection, SAP Predictive Analytics, Fraud Signal, Pandemic Datasets

1. Introduction
According to Lee et al., (2020), Benford's Law (BL) is an empirically discovered pattern for the frequency distribution of first digits in many real-life datasets including forensic analysis for potential manipulations of the data to detect fraud; applied to genome data; the half-lives of unstable nuclei; self-reported toxic emissions data; tax auditing; accounting; election data; stock markets; regression coefficients; inflation data; World Wide Web; religions; birth data; river data; first letter words; elementary particle; decay rates; and astrophysical measurements. According to Kraus et al. (2014), BL may be categorized as a descriptive data mining method, as it discriminates data, but also as predictive, as it identifies characteristics of datasets that may help to predict future schemas. Further, the authors mention that the applicability of the law is verified, but also controversially discussed in numerous papers. By observing the literature, it is obvious that the BL is applied in multiple research. According to Nigrini (1999), Benford's analysis might be helpful within the supply chain processes for estimations in the general ledger; the relative size of inventory unit prices among locations; check for duplicate payments; and check on customer refunds. Gauvrit et al. (2017) state various different applications of the law such as the distance between earth and known stars, crime statistics, the number of daily-recorded religious activities, earthquake depths, financial variables, study of gambling behaviors, and brain activity recordings. It is also important to note that the BL is very much applied as a fraud

indicator in supply chain data by using the 'trust-but-verify' approach that is advocated in the practitioner literature (Hales et al., 2009).

However, the literature states limitations to the application of BL, mainly depending on the sample size, and the fact that the dataset should occur naturally without any intervention to artificially control the data (Hales et al., 2009). The COVID-19 pandemic has been shattering the world, heavily impacting all business and non-business transactions. Datasets related to the pandemic are being used in many different analyses to arrive at different conclusions. However, the credibility of the results and conclusions depend heavily on the accuracy of the datasets. Much research is done using publicly available pandemic related datasets that are both directly and indirectly impacted by the high human control interventions. Therefore, it is important to detect any anomaly in datasets before they are utilized. According to a study by Lee et al. (2020), if the COVID-19 epidemic growth curve follows an exponential distribution, the number of infections and deaths will obey BL. Their conclusion is that it is possible that when the degree of intervention is high, the growth of death or infection rates may not obey BL. Thus, they state that BL testing alone would not be sufficient to detect potential manipulations of the growth of the COVID-19 death rate. This leaves much doubt that is worth of more detailed investigation as almost all datasets available with the pandemic are at least of some degree of artificial control due to the sudden, unexpected responses needed.

1.1 Objective

This research study attempted to answer the question "Do the Pandemic Related Datasets with High Artificial Control Still Follow Benford's Law?". The findings are anticipated to be useful for the data analysts and researchers while it adds to the current literature gap. This paper may also serve as a class case study for the academia teaching data analytics and SAP predictive analytics software.

2. Literature Review

2.1 Benford's Law

According to Nigrini (2012), Benford found out that numbers with low first digits occurred more often and then derived the expected frequencies of the digits as in: $p(d = d_1) = log_{10}\left(1 + \frac{1}{d_1}\right)\ for\ d\epsilon\{1,2,\dots,9\}$; where d is a number {1,2,9} and p is the probability. Table 1 lists the percentages expected on each of the four digits as per the BL. The BL is also referred to as the First-Digit law, or the Newcomb-BL to honor Newcomb (1881) who first found the phenomenon (Gauvrit et al., 2017).

Table 1. Benford's distribution of first, second, third and fourth digits

First Digit	Expected 1st digit, %	Expected 2nd digit, %	Expected 3rd digit, %	Expected 4th digit, %
0	-	11.97	10.18	10.02
1	30.10	11.39	10.14	10.01
2	17.61	10.88	10.10	10.01
3	12.49	10.43	10.06	10.01
4	9.69	10.03	10.02	10.00
5	7.90	9.67	9.98	10.00
6	6.70	9.34	9.94	9.99
7	5.80	9.04	9.90	9.99
8	5.10	8.76	9.86	9.99
9	4.60	8.50	9.83	9.98

Although frequencies for several leading digits were proposed, the Benford's first digit law is the most common application. The first two-digit test is regarded as a more focused test than the first digit test and is to detect abnormal duplications of digits and possible biases in the data (Kraus and Valverde, 2014; Nigrini, 2012).

2.2 Use of Benford's Law for Data Anomaly Detection in Supply Chains

A supply chain is an integrated process wherein many business entities (i.e., suppliers, manufacturers, distributors, and retailers) work together to acquire raw materials, convert these raw materials into specified final products, and deliver these final products to retailers (Beamon, 1992; Dissanayake and Cross, 2018). Managing supply chains has become even more challenging as its scope change dynamically due to many factors including changes brought about by global customer preferences. With its immense scope overwhelmed with information, supply chain management naturally calls for the need for managing big data in all of its functions. According to Gawankar et al. (2020), the data anomalies can be caused by many different factors including wrong procedures, wrong measurements, human and machine errors and fraud. As mentioned by Verma and Khan (2012), there is no effective audit tool available to date for identification of all types of mistakes/ frauds/ irregularities. According to Kraus and Valverde (2014), large data volumes and the inability to analyze them may lead to fraud and increased costs of supply chain management. With the increase of cyber-attacks and the hackers' use of sophisticated tools for malicious attacks, detecting such fraudulent and suspicious activities become even more challenging. Kraus and Valverde (2014) suggest that a fraud detection mechanism is necessary to reduce the risk of fraud especially in the supply chain area. While There are many different approaches in forensic data analysis to detect data anomalies, all approaches are based on a basic set of principles and knowledge (Varma and Khan, 2012). Authors further state that the BL is one such approach that is used to quickly spot potential anomalies and mismatches in data. The in-depth analysis of those data leads to the root causes of problems, therefore the application of the BL help managers to formulate potential anomaly/fraud preventive actions in a more effective manner.

2.3 Use of Predictive Analytics Tools

More high-quality information is advantageous as it leads to better informed managerial decisions. The digital world today is inundated with information. According to DuttaRoy (2016), 'business analytics' has become an art of combining every data footprint to gain actionable insights from huge data, and can help businesses grow, improve customer experience, and generate more revenue. SAP Expert Analytics that comes with SAP Predictive Analytics software is one of these statistical analysis and data mining toolset that enables the building of predictive models to discover hidden insights and relationships in data. (SAP Analytics 3.3 online help guide, 2020) Further, features in it include the ability to perform various types of data analyses; ability to visualize the quality of models from training datasets; ability to support a range of predictive algorithms; ability to support the use of the R; and its in-memory data mining capabilities for handling large volume data analysis efficiently (SAP Analytics 3.3 online help guide, 2020).

According to McLeod et al. (2017), analytical tools offer the greatest benefit, and the organizations use predictive analysis to gain insight, expose opportunities by building models, and for in-depth understanding. As stated by Menon (2020), SAP Predictive Analytics is a powerful system that leverages in memory computing capabilities of the HANA database and offers the business value of running predictive models on production data that feeds the operational systems. SAP Expert Analytics is one of many SAP Predictive Analytics tools that has many functions including the retraining of the models to insure their performance level and accuracy, as well as the detection of the model's deviations. However, as stated by McLeod et al. (2017), since these products are relatively new products, the curricula for these big data and analytics tools are still developing, and it is important to add to the available literature to support the widespread application of these tools.

3. Methods

This research used datasets from several data sources to cover different time periods and sample sizes: None-COVID-19 related data are assumed to have no or insignificant artificial control, whereas COVID-19 related data are influenced by considerable effort made towards controlling the final outputs. SAP Expert Analytics tool pack was used for performing the Benford's analysis and obtain the graphs as shown in figure 1, 2, and 3 below. The procedure was as per the instructions published by Kale, and Jones in chapter 8 of their 'Predictive Analytics' textbook (Kale, N., & Jones, N., 2020). To evaluate the reliability of adhering to the BL, visual observation coupled with Pearson's chi-square test for goodness of fit, and the Kupier's test were performed. The Pearson's goodness-of-fit Chi-square statistic, χ^2, computes the sum of the squared deviations of the empirical number of observations Nj (with $\chi^2=\sum_{j=1}^{m}\frac{(Nj-nj)^2}{nj}$ Nj) for each of the first digit j ∈ {1, 2, 3 . . ., 9} and the expected frequency n_j = (p (j) x n) as proposed by the BL. The Kuiper's test statistic is a rotation-invariant Kolmogorov-type test statistic. The critical values of a modified Kuiper's test statistic are used according to the guidelines given in Stephens (1970): for n>8, 1.537 for 15% significance, 1.620 for 10% significance, 1.757 for 5% significance, and 2.001 for 1% significance (Koch and

Okamura 2020). Chi-Square test is extremely sensitive for large sample sizes and tends to reject statistical significance even for small differences (Koch and Okamura 2020), however Kupier test is not that sensitive. In the Kupier test, the test statistic, T is calculated as: $\mathrm{T} = (\mathrm{D}_\mathrm{n}^{+} + \mathrm{D}_\mathrm{n}^{-})[\sqrt{\mathrm{n}} + 0.155 + \frac{0.24}{\sqrt{\mathrm{n}}}]$ where, $\mathrm{D}_\mathrm{n}^{+} = \sup\,(\mathrm{H}_\mathrm{d} - \mathrm{P}_\mathrm{d})$ and $\mathrm{D}_\mathrm{n}^{-} = \sup\,(\mathrm{P}_\mathrm{d} - \mathrm{H}_\mathrm{d})$; where H_d stands for the cumulative frequency of the first digit d in the observed data, and P_d that of the Benford distribution.

The following hypothesis are used in this study.
H_0: The observed distribution follows the theoretical Benford's distribution.
H_a: The observed distribution does not follow the theoretical Benford's distribution.

Null hypothesis is not rejected if the test statistic does not exceed the corresponding critical values: 20.09 for $\alpha = 0.01$, and 15.51 (13.36) for $\alpha = 0.05$ ($\alpha = 0.10$) respectively.

4. Data Collection

This research study used several different datasets from several sources, several time periods, and several sample sizes. They include an USAID supply chain health commodity shipment and pricing dataset with regular data before the COVID-19 pandemic, and datasets related to the COVID-19 pandemic from several sources. Table 2 reports the sample sizes after removing non-data and zeroth digit rows. According to Hales et. al. (2009), literature do not provide any firm rules on the sample sizes for Benford's Analysis. The literature suggests that the primary qualifications to apply the BL is for the dataset to be large enough, and generated naturally (i.e., without the intervention or artificial limitations that prohibit digits from taking on values from 1 to 9) (Hales et al., 2009). However, it is also found that the artificial limitations assumption can be relaxed in certain contexts (Hales et al., 2009). As published by Hales et. al. (2009), the minimum size necessary to conduct digital analysis has not been established except that it must be large. The authors further state that the sample sizes small as 100 have been tested with little success, sample sizes around 500 have provided mixed results, while those above1000 have provided the best results when used with appropriate data types.

Table 2: Size of the datasets and their time periods

Item	Before COVID-19 Datasets				After COVID-19 Datasets							
	Shipping Value	Packing Price	Freight Quantity	Freight Cost	COVID-19 Deaths	COVID-19 Confirmed Positives	COVID-19 Recovered	PPE Sales in USA	Chinese_Total Cases	Total Cases	US_Total Cases	World_Total Death
Final sample size after removing zero/no data rows	10324				266			44532	331	54188	331	46132
Period	2006-06-02 to 2015-08-31				2020-3-25 to 2020-12-16			2020-7-28 to 2020-11-25	2020-1-21 to 2020-12-17			

5. Results and Discussion

5.1 Numerical Results

Table 3 shows the observed first digit distribution percentages for shipping value, packing price, freight cost, and freight quantity of health commodity shipments prior to the pandemic, and their chi-square and Kupier test statistics. By observing the Chi-square statistics and Kupier statistics in table 3, it can be inferred that the frequency distribution percentages of the first digit of all the observations failed to reject the null hypothesis at a significance level of 10% for the Chi- Square test, and at a significance level of 15% for the Kupier test. Thus, this indicates that there are no red flags, and no indication of any data irregularity or anomaly. This dataset demonstrates strong adherence to the Benford's law and verifies regular datasets without human intervention follows the suggested distribution.

Table 3. Test results for regular before COVID-19 datasets

Benford's Exp%	Shipping Value,	Exp%-Obs%	Squared Dif./Exp	Pack Price, Obs%	Exp%-Obs%	Squared Dif./Exp	Freight Cost, Obs%	Exp%-Obs%	Squared Dif./Exp	Quantity, Obs %	Exp%-Obs%	Squared Dif./Exp
30.1	29.0	1.1	0.0	25.0	5.1	0.9	32.0	-1.9	0.1	29.0	1.1	0.0
17.6	19.0	-1.4	0.1	19.0	-1.4	0.1	18.0	-0.4	0.0	18.0	-0.4	0.0
12.5	12.0	0.5	0.0	14.0	-1.5	0.2	11.0	1.5	0.2	13.0	-0.5	0.0
9.7	10.0	-0.3	0.0	7.0	2.7	0.7	9.0	0.7	0.0	10.0	-0.3	0.0
7.9	8.0	-0.1	0.0	5.0	2.9	1.1	7.0	0.9	0.1	9.0	-1.1	0.2
6.7	7.0	-0.3	0.0	5.0	1.7	0.4	6.0	0.7	0.1	7.0	-0.3	0.0
5.8	6.0	-0.2	0.0	9.0	-3.2	1.8	6.0	-0.2	0.0	5.0	0.8	0.1
5.1	5.0	0.1	0.0	11.0	-5.9	6.8	5.0	0.1	0.0	4.0	1.1	0.2
4.6	4.0	0.6	0.1	4.0	0.6	0.1	5.0	-0.4	0.0	4.0	0.6	0.1
Chi Squared Test Statistic			0.28*			12.07*			0.57*			0.67*
D+.	%	1.39			5.9			1.9			1.1	
D-	%	1.1			5.1			1.49			1.1	
Kupier Statistic		0.08*			0.36*			0.11*			0.07*	

Table 4: Test results for COVID-19 pandemic related datasets with high intervention

Benford Exp%	COVID Death, Obs%	Exp%-Obs% Difference	Squared Dif./Exp	COVID positives, Obs %	Exp%-Obs% Difference	Squared Dif./Exp	COVID recovered, Obs %	Exp%-Obs% Difference	Squared Dif./Exp	PPE sales, Obs %	Exp%-Obs% Difference	Squared Dif./Exp
30.1	50.0	-19.9	13.2	33.0	-2.9	0.3	24.0	6.1	1.2	29.0	1.1	0.0
17.6	31.0	-13.4	10.2	11.0	6.6	2.5	23.0	-5.4	1.6	24.0	-6.4	2.3
12.5	2.0	10.5	8.8	7.0	5.5	2.4	17.0	-4.5	1.6	9.0	3.5	1.0
9.7	2.0	7.7	6.1	7.0	2.7	0.7	9.0	0.7	0.0	10.0	-0.3	0.0
7.9	3.0	4.9	3.0	10.0	-2.1	0.6	9.0	-1.1	0.2	11.0	-3.1	1.2
6.7	3.0	3.7	2.0	10.0	-3.3	1.6	8.0	-1.3	0.3	8.0	-1.3	0.3
5.8	3.0	2.8	1.4	10.0	-4.2	3.0	4.0	1.8	0.6	3.0	2.8	1.4
5.1	3.0	2.1	0.9	6.0	-0.9	0.2	2.0	3.1	1.9	4.0	1.1	0.2
4.6	4.0	0.6	0.1	5.0	-0.4	0.0	4.0	0.6	0.1	3.0	1.6	0.6
Chi Squared Test Statistic			45.63			11.34*			7.50*			6.96*
D+	%	19.9			4.2			6.1			6.39	
D-	%	10.49			6.61			5.39			3.49	
Kupier Statistic		0.98*			0.35*			0.37*			0.32*	

Table 5: Test results for after COVID-19 pandemic related datasets with high intervention

Benford Exp%	Chinese_Total Cases, Obs%	Exp%-Obs% Difference	Squared Dif./Exp	US_Total Cases, Obs%	Exp%-Obs% Difference	Squared Dif./Exp	All Cases, Obs%	Exp%-Obs% Differenc	Squared Dif./Exp	All COVID Death, Obs%	Exp%-Obs% Difference	Squared Dif./Exp
30.1	1	29.1	28.1	37.0	-6.9	1.6	31.0	-0.9	0.0	31.0	-0.9	0.0
17.6	1	16.6	15.7	11.0	6.6	2.5	17.0	0.6	0.0	16.0	1.6	0.1
12.5	1	11.5	10.6	6.0	6.5	3.4	12.0	0.5	0.0	12.0	0.5	0.0
9.7	1	8.7	7.8	7.0	2.7	0.7	9.0	0.7	0.0	9.0	0.7	0.0
7.9	1	6.9	6.0	10.0	-2.1	0.6	8.0	-0.1	0.0	8.0	-0.1	0.0
6.7	1	5.7	4.8	10.0	-3.3	1.6	6.0	0.7	0.1	7.0	-0.3	0.0
5.8	5	0.8	0.1	8.0	-2.2	0.8	6.0	-0.2	0.0	6.0	-0.2	0.0
5.1	56	-50.9	508.0	6.0	-0.9	0.2	6.0	-0.9	0.2	5.0	0.1	0.0
4.6	32	-27.4	163.2	4.0	0.6	0.1	5.0	-0.4	0.0	5.0	-0.4	0.0
Chi Squared Test Statistic			744.36			11.44*			0.39*			0.30*
D+	%	50.9			6.9			0.7			0.9	
D-	%	29.1			6.61			0.9			1.61	
Kupier Statistic		2.59			0.44*			0.05*			0.08*	

Chi-Square Critical Values	Kupier Critical Values
*10% significance 13.36; **5% significance 15.51; ***1% significance 20.09	*15%significance 1.537; **10% significance 1.620; ***5% significance 1.757; ****1% significance 2.001

Table 4 and Table 5 are showing observed first digit distribution percentages and the test results for datasets related to the COVID-19 pandemic. These datasets had both big and small sample sizes of >100. It is clear from the results that the Chinese_Total Cases data in table 5 fails to meet the critical values for both Chi-Square and Kupier tests at significance levels of 10% and 15% respectively. Failing both statistical tests, it can be inferred that the null hypothesis is rejected in favor of the alternative hypothesis for the Chinese COVID death related dataset, while all other COVID 19 related datasets pass the statistical tests. Therefore, the Chinese COVID 19 death related dataset is peculiar and raises the red flag for the presence of possible data anomaly since it does not follow the Benford's distribution. On the other hand, the rest of the pandemic related datasets do follow the Benford's distribution, even if they are highly controlled artificially.

5.2 Graphical Results

Figure 1 below presents the visualization of the observed first digit distribution against the Benford's expected distribution for the before pandemic dataset. Figures 2 and 3 visualize the pandemic related dataset with high intervention. It is clear that except for the Chinese_COVID 19 death related data distribution, all the other graphs do follow a distribution very close to the exponential distribution . This can be verified using literature as Lee et al. (2020) states that the Benford's distribution is typically closer to that of an exponential distribution specially with big datasets.

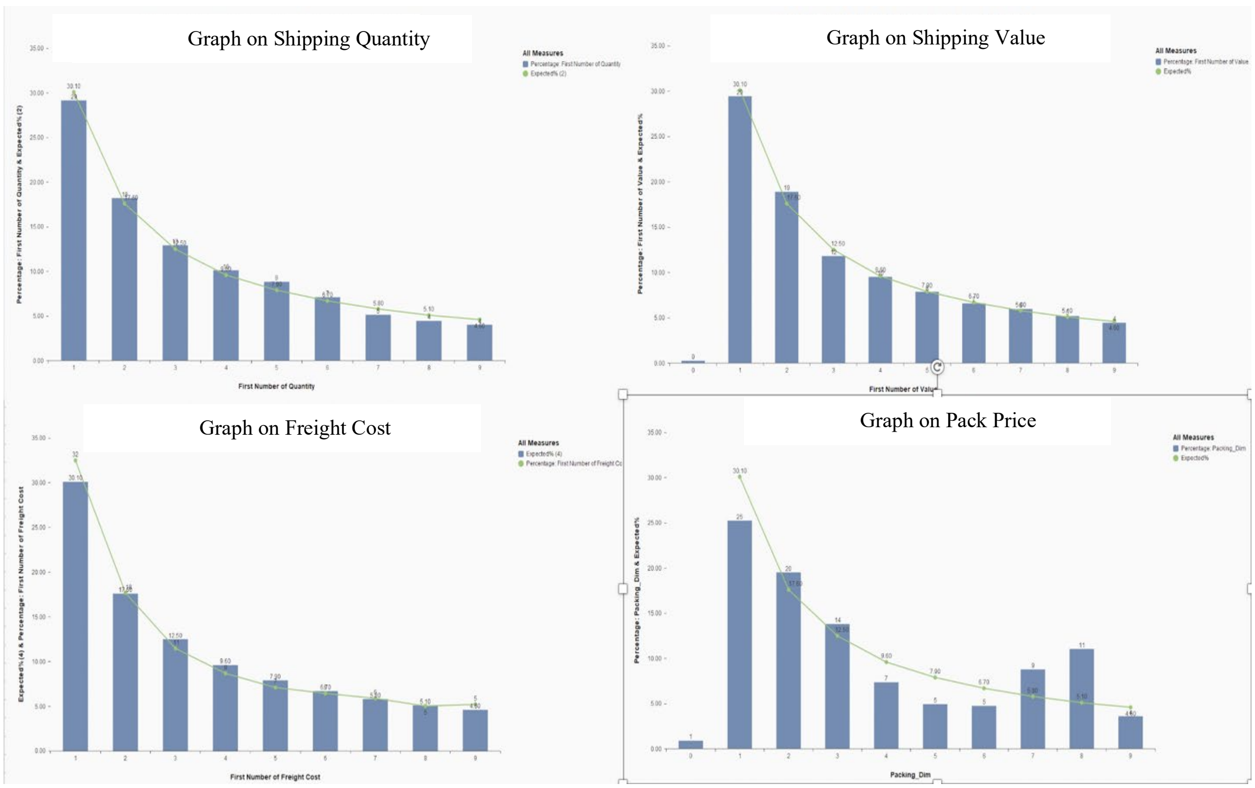

Figure 1. Observed first digit distributions Vs Benford's expected distribution for before COVID-19 dataset

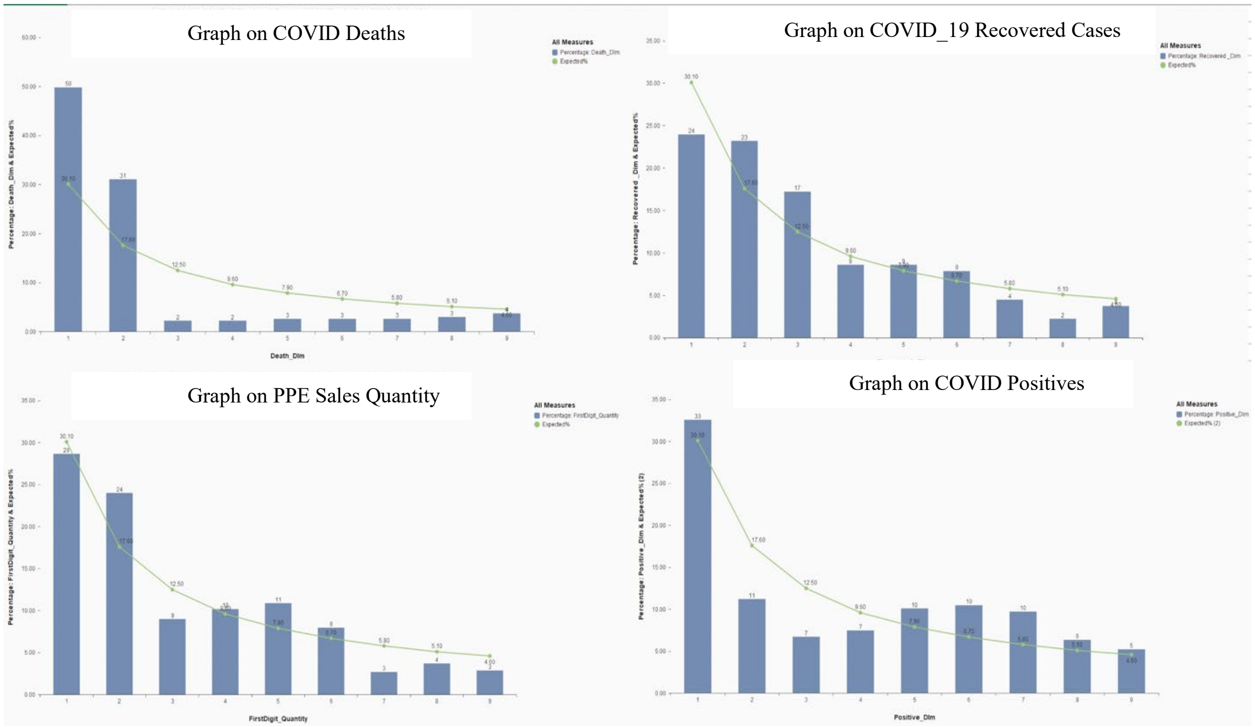

Figure 2. Observed first digit distributions vs Benford's expected distribution for after COVID-19 datasets reported in table 4

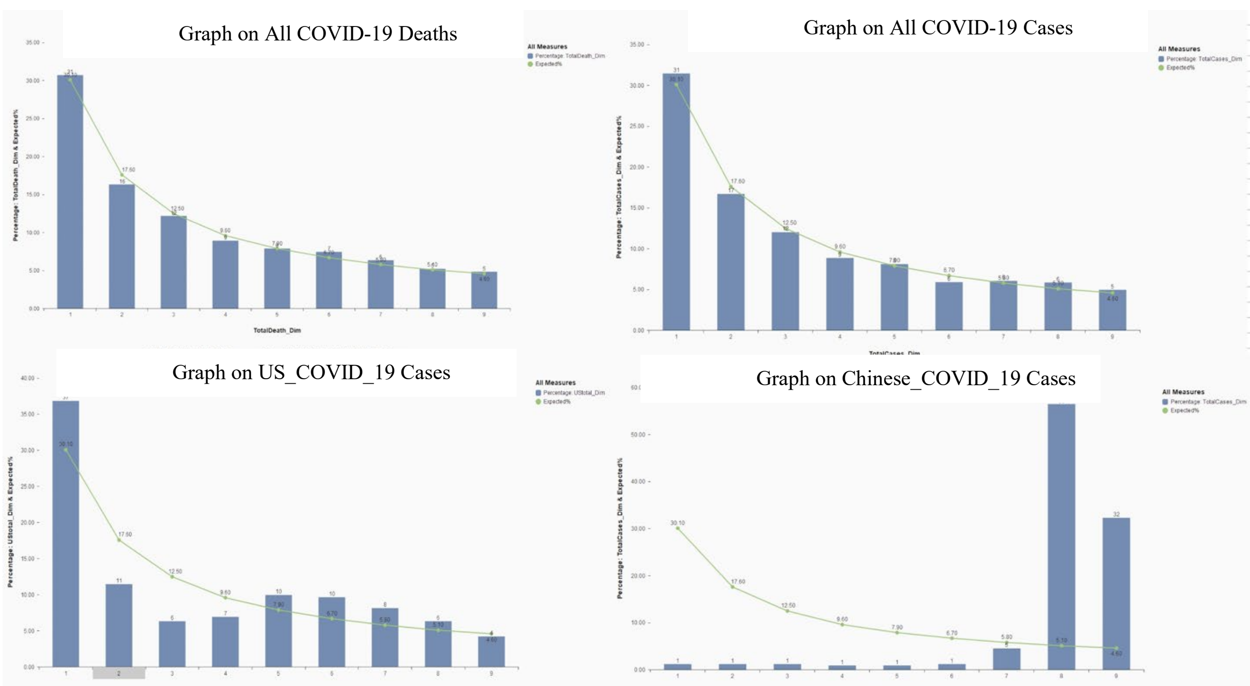

Figure 3. Observed first digit distributions vs Benford's expected distribution for after COVID-19 datasets reported in table 5

6. Conclusion

This study was conducted to answer the questions on "Do pandemic related datasets with high artificial control still follow the Benford's law?" The study used several datasets, all >100 in sample size from several data sources. The findings suggest that any dataset with reasonably big enough sample size both with or without human intervention, obey the BL, and that Benford's analysis can be used to identify anomalous datasets. The conformity can be visualized using graphs, and verified using statistical tests, and this study used the Chi-square and the Kupier tests. One of the post COVID-19 datasets failed both tests, and when its graph for the observed % of first digit vs the expected % in Benford's was observed, this non-conformity was apparent because the distribution was not following the expected shape of an exponential curve.

As per Todter (2009), the BL is a potentially useful instrument to discover fraud and manipulation, however there may be many plausible reasons for deviations from the BL, such as insufficient variability of the underlying data, rounding effects or other irregularities (Kraus and Valverde, 2014). According to Kraus and Valverde (2014), supply chain data as well as other data requires a detailed business know how for interpreting to prevent misinterpretations. Gaining insight into data requires careful extraction, presenting of information in a meaningful way, and the transformation of findings into actionable insights (Gole and Shiralkar, 2020). Therefore, the red flag raised from Benford's analysis is another strong signal that caution should be practiced before making use of this dataset further. It is suggested in literature that any red-flags such as this should follow the 'trust-but-verify' approach before making conclusions. According to Koch, & Okamura (2020), the media frequently claim that the Chinese government has understated the numbers of those affected, and that it can be because the data sharing practices China had at the early stages of the pandemic were inadequate since they were affected first. The authors further state that the Chinese government was unable to test those who did not present at hospitals, and the testing capacities were limited. The results of this study showing anomaly in Chinese data therefore is justified, and it is clear that by applying BL on any dataset of reasonable size, presence of anomaly and irregularities can be identified.

This study demonstrated the procedure to detect potential data anomalies using SAP Predictive Analytics tools. Since the process is simple, effective, and quick for the supply chain managers or auditors to detect red flags in their supply chain big data, they can follow in-depth to detect anomalies including fraud. The findings will support the managers/auditors with their preventive measures and/or the formulation of organizational policies. As it is more efficient to run the checks and store the results in the cloud with SAP Expert Analytics, analysts can perform more frequent checks on big data from the business world. While this article adds to the current lacuna in literature in this field, it may also be utilized as an SAP analytics software training guideline. Since the procedure is well detailed, and the dataset is openly accessible, the supply chain academia may use this as one of the class case studies.

More future research on effective sample sizes, and a clear taxonomy may help the generation of a clear set of guidelines for organizations to apply the BL. Such initiative can help the organizations to avoid the unnecessary investigations and expenditure on 'false-alarms'. With Benford's analysis becoming more common in fraud detection, new complementary analyses are already introduced making it more difficult, even for informed swindlers intentionally conforming to the law to remain undetected (Gauvrit et al., 2017). These complementary analyses are also a another area generating more future research.

References

Benford, Frank. July 15,2012. The Law of Anomalous Numbers." *The American Philosophical Society*78.4 (1938): 551-72. *JSTOR*. Web. http://www.jstor.org

Beamon, B. M.1998. Supply chain design and analysis: Models and methods. *International journal of production economics*, *55*(3), 281-294.

Dissanayake, C. K., & Cross, J. A. 2018. Systematic mechanism for identifying the relative impact of supply chain performance areas on the overall supply chain performance using SCOR model and SEM. *International Journal of Production Economics*, *201*, 102-115.

DuttaRoy, S. 2016. SAP Business Analytics: *A Best Practices Guide for Implementing Business Analytics Using SAP. Apress.*

Gawankar, S. A., Gunasekaran, A., & Kamble, S. 2020. A study on investments in the big data-driven supply chain, performance measures and organizational performance in Indian retail 4.0 context. *International Journal of Production Research*, *58*(5), 1574-1593.

Gauvrit, N. G., Houillon, J. C., & Delahaye, J. P.2017. Generalized BL as a lie detector. *Advances in cognitive psychology*, *13*(2), 121.

Gole, V., & Shiralkar, S. 2020. Empower Decision Makers with SAP Analytics Cloud. *Springer Books*.

Hales, D. N., Chakravorty, S. S., & Sridharan, V. 2009. Testing BL for improving supply chain decision-making: A field experiment. *International Journal of Production Economics*, *122*(2), 606-618

Kalé, N., & Jones, N. 2020. Practical analytics 2nd edi, *Epistemy Press,* ISBN: 9780997209242

Kraus, C., & Valverde, R. 2014. A data warehouse design for the detection of fraud in the supply chain by using the BL. *American Journal of Applied Sciences*, *11*(9), 1507-1518.

Koch, C., & Okamura, K. (2020). Benford's Law and COVID-19 Reporting. Available at SSRN 3586413.

Lee, K. B., Han, S., & Jeong, Y. 2020. COVID-19, flattening the curve, and BL. *Physical A: Statistical Mechanics and its Applications*, *559*, 125090.

McLeod, A. J., Bliemel, M., & Jones, N. 2017. Examining the adoption of big data and analytics curriculum. *Business Process Management Journal, 23*(3), 506-517. Doi:http://dx.doi.org.ezaccess.libraries.psu.edu/10.1108/BPMJ-12-2015-0174

Menon, A. 2020. Time Series Analysis in SAP Predictive Analytics. *Available at SSRN 3677420*.

Nigrini, M.J.1996. A taxpayer compliance application of BL. *J. Am. Taxat. Assoc.,* 18: 72-91

Nigrini, M. J., Mittermaier, L. J. 1997. The use of BL as an aid in analytical procedures. *Auditing,* 16: 52-67.

Nigrini, M.J. 1999. 'I've Got Your Number'. *J. Accountancy*, 187: 79-83.

Nigrini, M., 2012. BL: Applications for Forensic Accounting, Auditing and Fraud Detection. 1st Edn., *John Wiley and Sons, Hoboken, New Jersey*, ISBN-10: 1118152859, pp: 330.

SAP Predictive Analytics 3.3 Online Help-
https://help.sap.com/viewer/94dbf2ba9d4047618880187451c3b253/3.3/en-US

Stephens, M. (1970). Use of the Kolmogorov-Smirnov, Cramer-von Mises and related statistics without extensive tables. Journal of the Royal Statistical Society, B32, 115-122.

Tödter, K. H. (2009). BL as an Indicator of Fraud in Economics. *German Economic Review,* 10(3), 339-351.

Varma, D. T., & Khan, D. A. (2012). Fraud Detection in Supply Chain using Benford Distribution. *International Journal of Research in management*, *5*(2).

USAID Dataset on supply chain health commodity shipment and pricing data:source: https://catalog.data.gov/dataset/supply-chain-shipment-pricing-data, accessed on 10.1.2020

PPE sales dataset: source: https://data.ca.gov/dataset/covid-19-ppe-logistics/resource/7d2f11a4-cc0f-4189-8ba4-8bee05493af1_Accessed on 11.28.2020

COVID-19 Dataset: source: https://covidtracking.com/data/download_Accessed on 12.19.2020

Biographies

Dr. Kalpani Dissanayake is an Assistant Professor at the Pennsylvania State University, USA involved in their Project and Supply Chain Management degree program. She received her Ph.D. in Systems and Engineering Management from the Texas Tech University (TTU) and obtained her B.Sc. in Engineering and the M.B.A. degrees from the University of Peradeniya, Sri Lanka. Dr. Dissanayake was honored as the "Banner Bearer for the Graduate School" at the TTU commencement ceremony in August 2017 for her best all-around achievements during the Ph.D. program. She has also won several other academic awards including the 'Best Dissertation Award' at the American Society of Engineering Management(ASEM) conference 2018, the 'Merl Baker Award for the Best Student Paper' at the ASEM Annual Conference 2015, the 'J.T. and Margaret Talkington Fellowship Award 2015/2017' from TTU, and the 'Doctoral Degree Scholarship Award from the Ministry of Higher Education in Sri Lanka, 2013'. Prior to joining PennState, Dr. Dissanayake taught in several other universities including TTU, and has also worked in the private sector. Her current research interests include application of business analytics for organizational problem solving, performance improvement in supply chains, and teaching pedagogy. Dr. Dissanayake also holds the PMP certification from the Project Management Institute and the Engineer-In-Training (EIT) license from the Texas Board of Professional Engineers.

Dr Jay Daniel is a Program Leader for MSc Global Operations and Supply Chain Management and Senior Lecturer in the Derby Business School at University of Derby. Before joining the Derby Business School, he was a Lecturer (Assistant Professor) in Supply Chain and Information Systems at University of Technology Sydney (UTS), Australia. Previously with DB Schenker, Australia, and Alliance International Registrar, Asia Pacific, he held positions of Senior Management Consultant, Supply Chain Solution Analyst, Project Manager, Industry Trainer and Lead Auditor. He have made contributions to multiple research areas in the context of logistics and supply chain management with demonstrated practical applications across a wide range of industries. His primary areas of research focus are: Business Analytics and Supply Chain Management, Information Systems and Sustainable Supply Chain, Decision Making in Logistics and Supply Chain and Healthcare Supply Chain Management. He has been invited as a keynote speaker/invited speaker at international industry and academic workshops and conferences such as Keynote Speaker in Oracle Modern Business Experience Conference, etc. around the globe. An expert in applied and problem-driven research, he has used analytical tools and innovative optimization approaches to help managers create efficient, resilient and sustainable supply chains. He has been engaged in consulting to wide range of industries and organization structures, from small and medium size Australian companies to Fortune 500 corporations.