Major Project on

FRAUD DETECTION AND ANALYSIS
using
BENFORD'S LAW

Submitted By:
Kapil Kumar Tyagi
Roll No.: 2K20/EMBA/18

Under the Guidance of:

Mr Yashdeep Singh

(Assistant Professor)

DELHI SCHOOL OF MANAGEMENT
Delhi Technological University
Bawana Road Delhi 110042

CERTIFICATE

This is to certify that the dissertation report titled "**Fraud Detection and Analysis using Benford's Law"** is a work carried out by **Mr. Kapil Kumar Tyagi** of **EMBA 2020-22** and submitted to Delhi School of Management, Delhi Technological University, Bawana Road, Delhi-42 in partial fulfillment of the requirement for the award of the Degree of Masters of Business Administration (Executive).

Signature of Guide | **Signature of Head (DSM)**

Place: | **Seal of Head**

Date:

DECLARATION

Myself, **Kapil Kumar Tyagi**, student of **EMBA 2020-22** of Delhi School of Management, Delhi Technological University, Bawana Road, Delhi – 42, hereby declare that the dissertation report "**Fraud Detection and Analysis using Benford's Law"** submitted in partial fulfillment of Degree of Masters of Business Administration (Executive) is the original workconducted by me.

The information and data given in the report is authentic to the best of my knowledge.

This report is not being submitted to any other University, for award of any other Degree, Diploma or Fellowship.

Kapil Kumar Tyagi

Place:
Date:

ACKNOWLEDGEMENT

I would like to express my sincere gratitude towards my Guide, Mr. Yashdeep Singh (Assistant Professor, Delhi School of Management, DTU) for the support and valuable guidance throughout the duration of theproject. I thank him for the constant encouragement and support at every stage.

My sincere gratitude goes out to my colleagues whose participation in the project gave many valuable inputs for its completion.

Kapil Kumar Tyagi

(2K20/EMBA/18)

ABSTRACT

Fraud is any risky activity that aims to cause financial loss to another person. Fraud occurs as a result of deliberately used data in a complex cyber-complex operation that is a difficult task to find or research agencies. However, a ban on risky fraud is the best way to reduce fraud. The complexity of the supply network details allows fraudsters to commit fraud beyond internal control. Detecting fraud by analysing large amounts of data is a difficult task to find or evaluate agencies. Intension to use Excel Sheet to conduct Benford's Law distribution statistics tests as an effective tool for detecting red flags on suspected data. Supportive tools, leads to inaccurate detection or hidden data patterns, and its in-depth analysis helps agencies to scientifically assess the feasibility of using a single trust-but-verification' approach to a delivery network using a possible distribution called Benford's Law distribution within a short timeframe. and to prevent fraudulent transactions.

Applications of Benford's Law

The basic application of Benford's law is in fraud detection. Here are some of the most popular fraud detection applications of the law:

- **Accounting**: The idea behind detecting using Benford's law is that, if data of a certain type is known to be close to Benford's law, the chi-squared test comes to the rescue to identify the fraud.
- **Election**: The law has been used in detecting election frauds in many country elections.
- **Banks**: Deutsche Bank said Benford's Law works on balance sheets. The bank said that the law applies equally well to balance sheets and income statements as it does on another dataset.
- **Approximation of mathematical models**: Tests for goodness-of-fit to Benford are also useful as a diagnostic tool for assessing the appropriateness of mathematical models.

Table of Contents

1 INTRODUCTION

1.1 Background

In 1881, astronomer and mathematician Simon Newcomb (1881) stumbled upon the fact that the first pages of a book with log tables wear out much faster than the last. People are more likely to look for numbers with numbers 1 and 2 than numbers that start with numbers 8 and 9. Based on his observations, Newcomb came to the following conclusions from the following series of positive mistakes: The mantissa of the logarithms have equal probabilities, so we can determine the percentage of numbers whose first digit is from 1 to 9. Similarly, we can continue up to the nth number, such as determining the percentage of the second number (from 0 to 9). . Engineer and physicist Frank Benford set out to experimentally test Newcomb's hypothesis. He collected many data sets, including population data, addresses, newspaper articles, mortality rates, baseball statistics, elemental atomic weights, and numbers from the Readers Digest article (Benford, 1938). His analysis showed good agreement with Newcomb's distribution of leading digits. This phenomenon got to known as Benford's law. So, while human intuition dictates that each number in the set occurs with equal probability, Benford's law says that the number 1 is ahead about 30% of the time, the frequency of each successive number decreases, and the number 9 occurs less often.

1.2 Objective of Study

Modern corporate systems can register thousands of transactions every day. This makes it difficult to find multiple instances of anomalous activity in legitimate transactions. For large organizations operating in emerging global digital markets, this means monitoring hundreds of thousands of transactions and scrutinizing suspicious transactions. This can be expensive (Singh et al., 2014). The typical organization loses about 5% of its annual income due to fraud (ACFE, 2016b). Most victim organizations are unable to recover the losses from fraudulent activity, so proactive measures are needed to prevent and detect fraud. As a result, the need for continuous auditing and continuous monitoring

(CACM) in the global digital economy has increased as organizations become more complex and require more integrated business processes.

Digital research and detection of financial fraud in the huge datasets are complex tasks, because fraudsters use to create fraud by creating false organizations, e.g. goods / service evaluation or purchase price of orders or payments from merchants, and use fraudulent records to their advantage. It is a very difficult task to find or research agencies to find unusual duplicates of certain digits, digital combinations, certain numbers and circular numbers or other types of data manipulation. The principles of Benford's law were first published in the American Journal of Mathematics in 1881. Drs. Frank Benford (1938) argued that most numbers have smaller lead numbers, such as 1 or 2 than larger leading numbers. The distribution of opportunities in Benford is assessed by numerical tables (Goutsmit and Furry, 1944), body measurements (Knuth, 1969 and Burke and Kincanon, 1991), half live on radiation radio (Buck et al., 1993), interest rates (Becker, 1982), population (Hill, 1999), socio-economic data (Varian, 1972), stock exchange data (Ley, 1996, and Pietronero et al., 2001), law enforcement on biological discovery (Hoyle et al., 2002)), Newton's method (Berger et al., 2005), incorporates global scientific data (Niger and Miller, 2007), election results (Torres et al., 2007, etc. Recent research papers have reported exciting new areas, where Benford's Law is true, but audits using auditors to identify fraudulent changes by looking at digital tax waves or other financial information against the distribution of Benford (Nigrini, 1996). provides the auditor with a tool simple and effective in detecting fraud (Durtsch i, Hillison and P Pacini, 2004).

Fraud can be found in many ways or at least one trying to find it in many ways. There are many different ways to analyze active forensic data. All methods are based on legal principles and basic information. Other studies have reported success on similar effects of Benford's law to be a useful tool for identifying suspected accounts for further analysis (Skousen, et al., 2004); at prices on ebay auctions (Giles, 2005) etc. Distribution of Benford opportunities to detect available fraud chain to identify abnormal transactions of small data, leading to red flags (Varma, T. N. and Khan, D. A., 2012). In this study they found that full payment details of all merchants for a period of time corresponded to this payment but in applying this rule to each seller's payment details from a set data set, it was found that there were violations in Benford distribution of individual merchant details and this pattern.

1.3 Scope of Study

Benford Analysis is a powerful and easy-to-use digital analytical method that specifies the distribution of digital values in many common scenarios by identifying expected pattern data to create a new way to detect fraud with the help of an Excel spreadsheet. Benford's distribution as indicators of unpopular behavior that are strong indicators of fraud. This document demonstrates an effective way to search for hidden pattern or error or fiction data in a large test set from the Supply Chain Network using statistical testing in the Benford distribution. Identifying mismanaged data and conducting in-depth analysis of that data will always lead to the detection and prevention of fraudulent activities. The tool and technology outlined in this article will assist the auditor.

2 LITERATURE

The principles of Benford's law were first published in the American Journal of Mathematics in 1881. Simon Newcomb published an article in an American magazine and concluded that there are more numbers available, starting with the first number than the other numbers. Drs. Frank Benford (1938) argued that most numbers have smaller lead numbers, such as 1 or 2, than larger leading numbers. Durtschi et al. (2004) review the types of accounting information that may be relevant to the Benford Act and the circumstances in which ‘‘ Benford analysis ’’ may be helpful. Benford's law such as checking the validity of data is not limited to internal audit and the evidence is valid. Hoyle et al. (2002) applied Benford Law on biological discovery and Niger and Miller (2007) applied Benford Law to ground science data. The mathematical theory underpinning Benford Law is still developing. Examples include Berger et al. (2005), Kontorovich and Miller (2005), Berger and Hill (2007), Miller and Nigrini (2008) and Jang et al. (2009). Recent mathematical papers highlight new interesting situations in which Benford's law is true, but the experiments used by auditors and published studies are similar experiments proposed in Niger and Mittermaier (1997). is called the " Benford Set 'in Niger (2000, 12). Hurlimann's bibliography (2006) lists 350 copies of Benford's law between 1881 and 2006, 166 of which appeared between 2000 and 2006. Today, Benford's law has been widely used in many areas, from the efficient use of computer algorithms (Gent and Walsh, 2001) to eBay auction reviews (Giles, 2007). In business and administration, it is gaining increasing importance as a way to control doping 'datasets. Although sample size is important in applying the Act to technical data, the minimum size required for digital analysis has not yet been established - unless it must be large (Nigrini, 1996a; Hill, 1995). Samples as small as 100 were tested with minimal success. About 500 samples gave mixed results, and those over 1000 gave the best results when used with the right types of data (Hales etal., 2008).

3 LITERATURE REVIEW

This study often explores how a new analytical tool, Benford Law, can help in finding fraudulent jobs in various datasets and data/transactions. Documents can confirm Benford's terms and conditions. Benford's law as a test of data validity is not limited to internal audit and evidence work but can also be assessed as an indication of fraud in SCM using the 'trust but assurance' approach recommended in staff manuals.

4 METHODOLOGY

To evaluate the effectiveness and feasibility of digital analysis of Benford' s law on asset management, a sample of paid vendor data was used. First, we took the original data set in the excel sheet i.e.; all payment data of merchants of a particular department and extract the first digit of this data at their frequency. We compared the first digit waves with the Benford law predicting the first digit waves as shown in Table 1. We noted that this data set follows the distribution of Benford opportunities shown in Figure 1. To test our assumptions, even statistical deviations, we used Chi-square value assessments as shown in Table 2. and we found data confirming the distribution of Benford.

We then performed a similar exercise on each vendor data set and found that there was a violation of Benford' s distribution of single trader data as shown in Figure 2 and these patterns strongly rejected Benford' s rule as the main Chi-square test, shown in Table 3. This test shows a strong suspension for this trader. Finally we recorded the data, in which the difference between the observed frequency and the expected frequency percentage was high and low and we analyzed that data to detect fraud.

4.1 Selection and Import of Data

Sample data for currency exchange transactions made, e.g. Invoice prices for all vendors in the first column as shown in Figure 1 of the Microsoft Excel Worksheet took some time for the organization. Data entry in sheet columns is done manually or pasted from other sheets (CTR-C copy and then CTR-V attachment), or imported data with the help of data connection wizards. It must be ensured that all data in this column must be numbered. Bedford's law does not apply to all numerical persons such as pre-given numbers, naturally random, unstructured minimum or maximum value, aggregation of numbers around a certain value, etc.

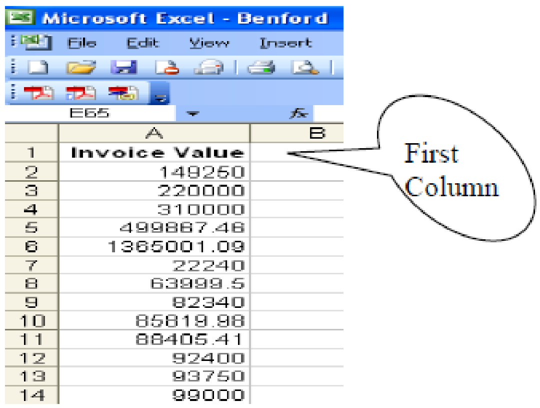


Figure 1 Vendor's Invoice Value

4.2 Extraction of First Significant Digit

The distribution of opportunities in Benford is based on the distribution of leading digits in a set of natural numbers. These lead numbers can be easily extracted using the LEFT Excel function, which has Syntax LEFT (text, [num_chars]), where text is a text string containing characters to be extracted, and Num_chars specifies the number of characters as needed. Multi_ numbers must be greater than or equal to zero. If the numbers are_a longer text length, the LEFT returns all text and is left out, assuming it is 1. In this case, most of the numbers on the left can be deducted using B2 cell formula such as = LEFT (A2,1) or by formula = LEFT (A2) and repeat this formula above the remaining lines at the end of the transaction.

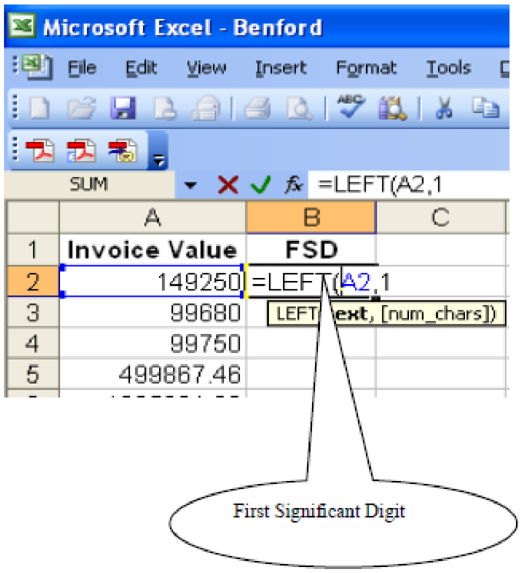
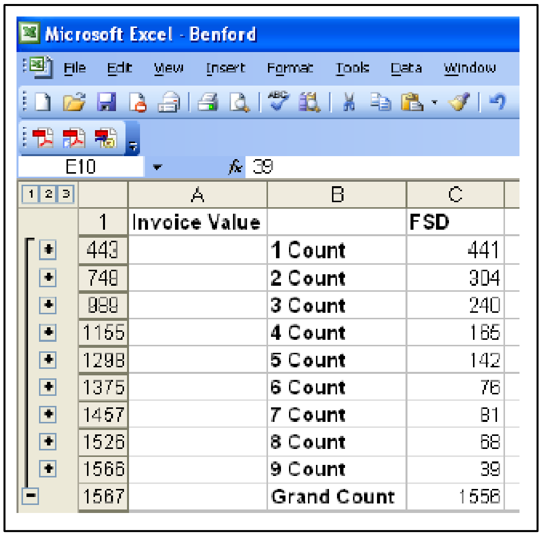


Figure 2 LEFT Function

4.3 Observed Frequency of First Significant Digit

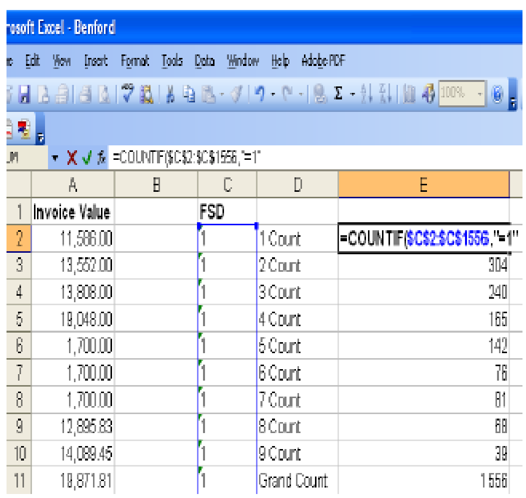
This function can be performed using the SUBTOTAL function. First, sort the First Significant Digit list, which was stored in the second column so that the linear particles are grouped together, and the rows, containing zeros in this column, are removed. Microsoft Excel can automatically calculate subtotals and large total numbers in a list. The list is set so that the summary report can be created by clicking on the frame symbols to show and hide the details lines for each theme. While it is possible to use the formula COUNTIF, which is Syntax COUNTIF (width, process) in which Width is the width of cells to be counted, and Terms are a way of determining the number, number, or text that defines which cells are counted. For example, to count all the ‘1s’ in C2: C1556, the formula would be = COUNTIF ($ C $ 2: $ C $ 1556, ”= 1”). This formula should be multiplied by more than eight cells, by 2 to 9 digits.



Microsoft Excel - Benford

E10 = 39

	A	B	C
1	Invoice Value		FSD
443		1 Count	441
748		2 Count	304
988		3 Count	240
1155		4 Count	165
1298		5 Count	142
1375		6 Count	76
1457		7 Count	81
1526		8 Count	68
1566		9 Count	39
1567		Grand Count	1556



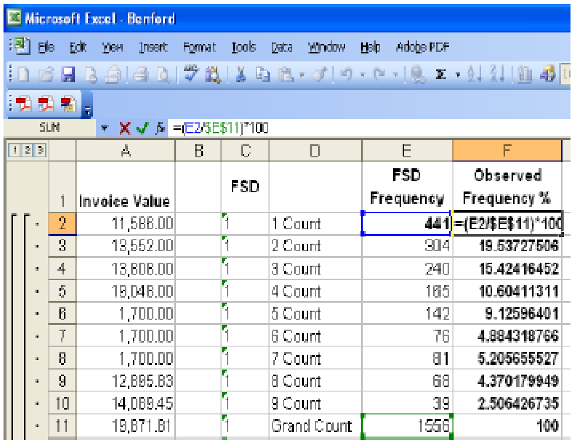
=COUNTIF($C$2:$C$1556,"=1"

	A	B	C	D	E
1	Invoice Value		FSD		
2	11,586.00		1	1 Count	=COUNTIF($C$2:$C$1556,"=1"
3	13,552.00		1	2 Count	304
4	13,808.00		1	3 Count	240
5	18,048.00		1	4 Count	165
6	1,700.00		1	5 Count	142
7	1,700.00		1	6 Count	76
8	1,700.00		1	7 Count	81
9	12,895.83		1	8 Count	68
10	14,089.45		1	9 Count	39
11	18,871.81		1	Grand Count	1556

Figure 3a Use of SUBTOTAL Function
Figure 3b Use of COUNTIF Function

4.4 Percentage Frequency Distribution of observed FSD

The frequency of all FSDs can be calculated in previous steps using the SUBTOTAL or COUNTIF formula in column E as appears in Figure 3a or 3b. To calculate the percentage of visual or actual data, the formula = (E2 / $ E $ 11) * 100 can be used, whereas the E2 cell contains the frequency of FSD 1 and the total recognition value of the E11 cell, shown in Figure 4.

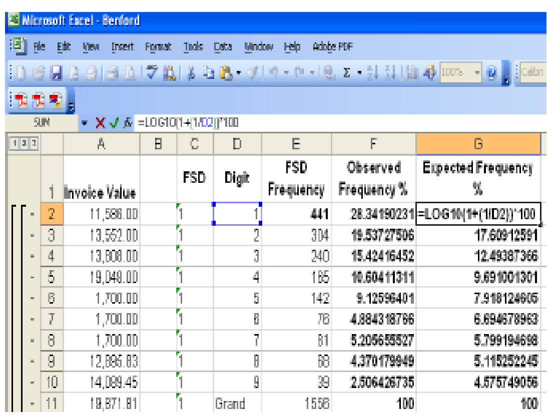


	A	B	C	D	E	F
1	Invoice Value		FSD		FSD Frequency	Observed Frequency %
2	11,586.00		1	1 Count	441	=(E2/$E$11)*100
3	13,552.00		1	2 Count	304	19.53727506
4	13,808.00		1	3 Count	240	15.42416452
5	19,048.00		1	4 Count	165	10.60411311
6	1,700.00		1	5 Count	142	9.12596401
7	1,700.00		1	6 Count	76	4.884318766
8	1,700.00		1	7 Count	81	5.205655527
9	12,885.83		1	8 Count	68	4.370179949
10	14,089.45		1	9 Count	39	2.506426735
11	19,871.81		1	Grand Count	1556	100

Figure 4 Actual Frequency % of FSD 1-9 in Cell E2-E10

4.5 Percentage Frequency Distribution of expected FSD

The frequency of all FSDs (1-9) can be calculated in the previous steps from column E as shown in Figure 3a or 3b. To delete Count text in Column D, this column selection and Count (CTR-F) can be used, and it will all be empty. Expected frequency percentages can be calculated using Benford' s Excel formula for Probability Distribution, i.e. = LOG10 (1+ (1 / D2)) * 100) digit 1. This formula should be multiplied by the other First Significant Digits 2 to 9 as shown in Figure. 5..



	A	B	C	D	E	F	G
1	Invoice Value		FSD	Digit	FSD Frequency	Observed Frequency %	Expected Frequency %
2	11,586.00		1	1	441	28.34190231	=LOG10(1+(1/D2))*100
3	13,552.00		1	2	304	19.53727506	17.60912591
4	13,808.00		1	3	240	15.42416452	12.49387366
5	19,048.00		1	4	165	10.60411311	9.691001301
6	1,700.00		1	5	142	9.12596401	7.918124605
7	1,700.00		1	6	76	4.884318766	6.694678963
8	1,700.00		1	7	81	5.205655527	5.799194698
9	12,886.83		1	8	68	4.370179949	5.115252245
10	14,089.45		1	9	39	2.506426735	4.575749056
11	19,871.81		1	Grand	1556	100	100

Figure 5 Benford Frequency % in Expected Frequency %

4.6 *Chi Square Test*

If detected or actual and expected values are obtained, the square of chi can be calculated as the sum of (O-E) * (O-E) / E. In this formula, O is the number of percent of First Detected digits and E is the expected number (computerized) percentage of frequency. A total of nine combined values gives the value of Chi Square. This amount can be evaluated by 8 degrees of freedom to determine whether Benford Act is accepted or rejected.

Digit	FSD Frequency	Observed Frequency %	Expected Frequency %	(O-E)	$(O-E)^2$	$(O-E)^2 / E$
1	441	28.34190231	30.1029996	-1.7611	3.101464	0.103028
2	304	19.53727506	17.6091259	1.928149	3.717759	0.211127
3	240	15.42416452	12.4938737	2.930291	8.586605	0.687265
4	165	10.60411311	9.6910013	0.913112	0.833773	0.086036
5	142	9.12596401	7.9181246	1.207839	1.458876	0.184245
6	76	4.884318766	6.69467896	-1.81036	3.277404	0.489554
7	81	5.205655527	5.7991947	-0.59354	0.352289	0.060748
8	68	4.370179949	5.11525224	-0.74507	0.555133	0.108525
9	39	2.506426735	4.57574906	-2.06932	4.282095	0.935824
Grand Total	1556					2.866352

Figure 6 Chi Square Test for all invoice data set

The critical values set for the distribution of the Chi square at 1%, 5% and 10% are the mean value of 8 degrees of freedom of 20.090, 15.507 and 13.362 respectively. If this number is greater than the calculated value then it accepts the absurd assumptions, otherwise it rejects Benford law, where there is a possibility of data manipulation by fraudsters. In Figure 6 it is clear that the calculated value is much smaller than the critical value set in the tables, which is why it receives a meaningless view, and the distribution is similar to Benford' s potential distribution of opportunities.

4.7 *Graph*

Select data from the visual frequency and the expected percentage column and then after inserting the chart wizard use chart tools such as design, layout, format etc., its output is shown below in figure 7. In the first chart series it is used as the actual frequency bar and the second frequency line as the second frequency line. look.

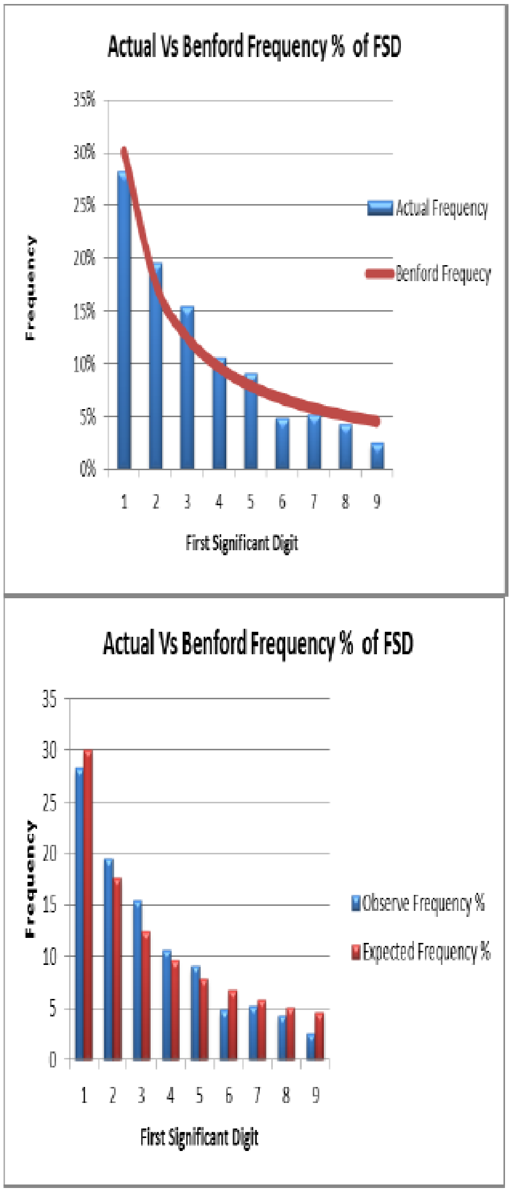


Figure 7 Actual (observed frequencies shown as line) and Benford frequency distribution %s (the expected frequencies from Benford's Law [in Eq. (1)] as bar) on y-axis and FSD on x-axis

5 ANALYSIS, DISCUSSION AND RECOMMENDATIONS

5.1 *Benford's Law and Applicability*

It has been observed that the first pages of the standard logarithm table show more aging than the last pages, indicating that the most commonly used numbers start with 1 digit rather than 9 digits (Newcomb, 1881; Benford, 1938). Benford has compiled more than 20,000 initial digits taken from a wide variety of sources. Sources ranged from random numbers to official statistical tables. Analysis of these numbers reveals that there is a logarithmic distribution of the original digits when the numbers are composed of two or more digits. Numbers taken from unrelated studies show a better fit with logarithmic distribution than numbers from statistical tables or official data. The first digit (D1) distribution second (D2) and first two digits (D1D2) in this data may be closely related to subsequent logarithmic distribution (Nigrini and Mittermaier, 1997)

$$P(D_1 = d_1) = log_{10}\left(1 + \tfrac{1}{d_1}\right) \quad d_1 \in \{1, ...9\} \tag{1}$$

$$P(D_2 = d_2) = \sum_{d_1=1}^{9} log_{10}\left(1 + \frac{1}{d_1 d_2}\right) \quad d_1 \in \{1, ... 9\}\ d_2 \in \{0, 1, ... 9\} \tag{2}$$

$$P(D_1 D_2 = d_1 d_2) = log_{10}\left[1 + \left(\sum d_i\ 10^{k-i}\right)^{-1}\right] \quad d_1 d_2 \in \{10, 11, 12\ ... 99\} \tag{3}$$

where P represents the expected probability of the digit in parenthesis. Using equation 3, the expected probability of the first-two digits 75 would be calculated as follows:

$$P(D_1 D_2 = 75) = \log_{10}\left(1 + \left(\frac{1}{75}\right)\right) = 0.005752 \tag{4}$$

The expected distribution of digits in the first, second, third, and fourth positions is shown in Table 1. It may be noted that high-quality digital distribution is increasingly approaching the same distribution. It follows that when it comes to numbers that are three digits or more than the digits on the right it is expected to be evenly distributed.

D_i	$P(D_1)$	$P(D_2)$	$P(D_3)$	$P(D_4)$
0	-	0.11968	0.10178	0.10018
1	0.30103	0.11389	0.10138	0.10014
2	0.17609	0.10882	0.10097	0.10010
3	0.12494	0.10433	0.10057	0.10006
4	0.09691	0.10031	0.10018	0.10002
5	0.07918	0.09668	0.09979	0.09998
6	0.06695	0.09337	0.09940	0.09994
7	0.05799	0.09035	0.09902	0.09990
8	0.05115	0.08757	0.09864	0.09986
9	0.04576	0.08500	0.09827	0.09982

Table 1. Benford's Law Expected Digit Distributions

(*Source*: Nigrini & Mittermaier, 1997)

Benford (1938) noted that the observed opportunities were closely related to events rather than a system of natural numbers. People tend to label things as 1, 2, 3, 4,… The idea that 1, 2, 4, 8,… being a more natural arrangement is not easily accepted. However, the latter occurs with an unexpectedly large number of events, for example, an increase in light sensitivity and increased light, or a sense of loud noise by two logarithmic operating conditions. Benford also noted that one of the best similarities in a logarithmic pattern was in data where the numbers did not have a relation to each other. In addition, distribution does not occur if the observed values are from a small range, if the numbers are assigned, or made by people (Slijepcevic and Blaskovic, 2014). This leads to the conclusion that the logarithmic relationship is the Unknown Numbers Act (Benford, 1938).

The fact that the series of random numbers is in line with Benford's law suggests that its use is conducive to the discovery of counterfeit data in financial transactions (Nigrini and Mittermaier, 1997; Durtschi et al., 2004; Ciaponi and Mandanici, 2015). By comparing the expected digital frequencies on invoices or payments; an unusual rise may be an indication of fraudulent transactions and requires further investigation by research staff.

Many fraudsters fail to take into account Benford's legal pattern when creating false invoices or payments. Therefore, analyzing data sets for the occurrence or non-occurrence of predictable patterns may reveal asset disposal conditions such as shell company plans, fraudulent payment schemes, misappropriation of assets or fraud, skiing or cash payments and segmentation of debt or other schemes involving fraud. predetermined performance limits (Lanza, 2003; ACFE, 2016a). This study automatically uses the Benford Act to monitor and analyze paid account activity data for confusing or 'red flags' in invoices and payments.

5.2 Data Compliance/non-compliance

Types of data that confirm the Benford Law:

i. Set of number that result from mathematical combinations of numbers - Results comes from two distributions. For ex: Accounts receivable (number sold * price), Account payable (number bought * price)
ii. Transaction level data, no need to sample. For ex: Disbursements, Sales, expenses.
iii. On large datasets, the more observations, the better. For ex: Full year's transactions
iv. Accounts that appear to confirm- when the mean of a set of numbers is greater than median and the skewness is positive. For ex: Most set of accounting numbers.

Types of data that do not confirm the Benford Law:

i. Dataset is comprised of assigned numbers. For ex: check numbers, invoice numbers, zipcodes.
ii. Numbers that are influenced by Human thoughts. For ex: Prices set at threshold levels ($199, ATM withdrawals)
iii. Accounts with a large number of firm-specific numbers. For ex: An account specifically setup to record $100 refunds.
iv. Accounts with a built in maximum and minimum. For ex: Set of assets that must meet a threshold to be recorded.
v. Where no transaction is recorded. For ex: Thefts, Kickbacks, Contract rigging.

5.3 Dataset Analysis

By repeating steps 4.1 - 4.7 for each set of invoice data set for vendors to find the value of each chi square, it was observed that there was a significant difference between the original and Benford Frequency% as shown in Figure 8 of FSD 9, where the potential for fraudulent data exchange exists in -Supply Chain Network.

Digit	FSD Frequency	Observed Frequency %	Expected Frequency %	(O-E)	$(O-E)^2$	$(O-E)^2 / E$
1	8	7.4766355	30.1029996	-22.6264	511.9524	17.00669
2	8	7.4766355	17.6091259	-10.1325	102.6674	5.83035
3	5	4.6728972	12.4938737	-7.82098	61.16767	4.895813
4	6	5.6074766	9.6910013	-4.08352	16.67517	1.720686
5	2	1.8691589	7.9181246	-6.04897	36.58999	4.621042
6	1	0.9345794	6.69467896	-5.7601	33.17875	4.955988
7	3	2.8037383	5.7991947	-2.99546	8.972759	1.547242
8	21	19.626168	5.11525224	14.51092	210.5667	41.16448
9	53	49.53271	4.57574906	44.95696	2021.128	441.7044
Grand	107					523.4467

Figure 8 Chi Square Test for a data subset

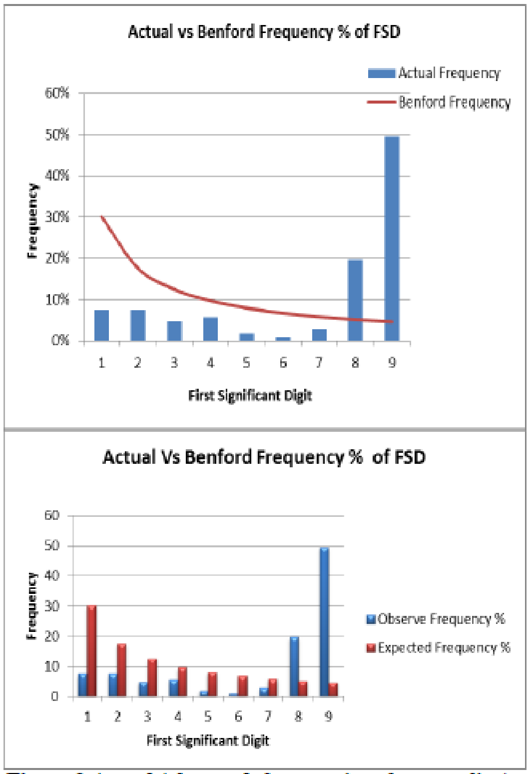


Figure 9 Actual (observed frequencies shown as line) and Benford frequency distribution %s (the expected frequencies from Benford's Law [in Eq. (1)] as bar) on y-axis and FSD on x-axis

Github repository: https://github.com/entrmatrix/Data_Analysis-Benford_Law

5.4 Result and Discussion

5.4.1 Invoice data analysis for all merchants

As of Figure 6, this data set appears to be consistent with Benford' s distribution of opportunities because the frequency of FSD is much higher than recommended by Benford' s law. The visual analysis of Figure 7 makes it very tempting to argue that the distribution of the original invoices for the original values is almost the same as predicted by Benford' s Probability Distribution, and Chi-square the fair value corresponds to Benford' s Probability Distribution. of Chi at 1%, 5% and 10% value of 8-digit freedom, 20.090, 15.507 and 13.362 respectively; that is why he accepts a point of view. Therefore, it leads to a strong integration of this pattern with Benford' s distribution.

5.4.2 *Analysis of data set of individual vendors*

In the big data analysis, it was found that the invoice data set for all vendors complied with Benford Distribution. By repeating the steps as discussed in 3.1 to 3.7, it is assumed that the data of one trader violates Benford' s distribution because the difference between the percentage of significant digit frequency and the expected distribution frequency is very high as shown in Figure 8, e.g. In this case the significant value of the Chi square distribution is 1%, 5% and 10% the value of 8 degrees of freedom is very high, which strongly prevents Benford' s distribution, and scares the red flags in the Supply Chain Network.

5.4.3 *Fraud Analysis*

By analyzing the invoice data of this particular vendor FSD differs from him, the difference between the observed frequency and the expected percentage was found to be very high (FSD 9). These transactions occur as most payments are made between Rs. 90000 to Rs. 9999 to avoid a high level of authority, namely one lakh, which was a fraudulent generation of procurement. This was due to the division of purchase orders, and repeated orders were given to merchants to use payment for the wrong reasons. In further analysis, it was observed that these fraudulent activities were carried out due to incomplete or unspecified job details, vendor selection without proper competency testing, and incorrect inclusion / incomplete data in the negotiation sheet to highlight the seller's ability, etc.

6 CONCLUSION

In the case of fraudulent transactions or data fraud in the Supply Chain cycle, data analysis can be used to detect fraud through the use of testing software. This study suggests that fraud detection can be one of the most effective ways to help auditors or find organizations to identify informal transactions, and to help them perform their tasks efficiently, efficiently and economically in the short term with the help of Simple Excel Tasks, which will reduce fraud risk.

7 REFERENCES

1. http://www.rspublication.com/ijrm/ijrm_index.htm.
2. Benford, Frank. "The Law of Anomalous Numbers." The American Philosophical Society78.4 (1938): 551-72. JSTOR. Web. http://www.jstor.org (July 15,2012)
3. Dr. L. Kailasam. " Benford Distribution - An Effective Audit Tool. " The Chartered Accountant. November 2011 : 716-720
4. Durtschi, Cindy, William Hillison, and Carl Pacini. "The Effective Use of Benford's Law to Assist in Detective Fraud in Accounting Data." V.1524-5586 (2004): 17-34. UIC. Journal of Forensic Accounting. Web http://www.uic.edu/
5. Saville, AD. "Using Benford's Law to Detect Data Error and Fraud: An Examination of Companies Listed on the Johannesburg Stock Exchange." SAJEMS 9.3 (2006): 341-
6. 54. Scribd. Scribd. Web.http://www.scribd.com/doc/47789223/Saville-Using-282006-29
7. Benford, Frank (1938), 'The Law of Anomalous Numbers', The American Philosophical Society Vol.7,No.4 ,pp. 551-72. JSTOR. Web. http://www.jstor.org (Accessed, July 15,2013)
8. Berger, Arno, and Theodore P. Hill. (2007), 'Newton's Method Obeys Benford's Law', American Mathematical Monthly Vol. 114, No.7, pp.588-60I.
9. Knuth, D. (1969), The Art of Computer Programming, vol. 2, 219-229. (2nd ed.).Addison-Wesley, Reading, MA, 239-249. (3rd ed.) (1998), 254-262.
10. Miller, S.J. and M.J. Nigrini (2006), ' Differences between independent variables and almost Benford behaviour. Arxiv preprint math.PR/0601344 at www.arxiv.org.
11. Nigrini, Mark J. Datas 2009 for Excel 2007 Program Details. Web. http://www.nigrini.com/datas_software.htm (Accessed ,,July 12,2013)
12. http://office.microsoft.com/en-us/excel/ (Accessed, May 8 ,2013).