Fraud Detection in Supply Chain using Excel Sheet

T N Varma and D A Khan, Ph.D
Department of Computer Applications,
National Institute of Technology, Jamshedpur, India

ABSTRACT
Fraud occurs due to intentionally manipulated data in complex and dynamic supply chains cyberspace being a complicated task for detecting or auditing agencies. However, prevention from vulnerable manipulation is the best way to reduce frauds. The application of Excel sheet as decision supporting tools, leads to identify abnormally mismatch or hidden pattern of data, and its depth analysis helps agencies to scientifically examine the feasibility of implementing one 'trust-but-verify' method in supply chain network using a probability distribution called Benford's distribution within a short span of time to detect and prevent fraudulent transactions. This paper demonstrates how to use Excel Sheet to perform Benford distribution statistical test as an effective tool for locating red flags in suspected data from decision-making data-set of supply chain network.

Keywords
Benford Distribution, Excel sheet, Fraud, Supply Chain Management

1. INTRODUCTION
Digital auditing and financial fraud detection in Supply Chain Network are complex tasks, because fraudsters use to commit fraud by creating fictitious entities, e.g. material/ service rate or purchase Order value or vendor payments, and then manipulate the fictitious records to their advantage. It becomes very challenging job for detecting or auditing agencies to find abnormal duplications of specific digits, digit combinations, specific numbers and round numbers or other forms manipulation in data. The principles of Benford law were first published in American Journal of Mathematics during 1881. Dr. Frank Benford (1938) became convinced that more numbers have small leading digits, like 1 or 2 than large leading digits. Benford Probability Distribution has been tested for numbers in statistical tables (Goutsmit and Furry,1944), physical constants (Knuth,1969and Burke and Kincanon,1991), half lives in radioactive decay (Buck et al. ,1993), failure rates (Becker,1982), population volumes (Hill ,1999) , socio-economic data (Varian ,1972), stock exchange data (Ley,1996, and Pietronero et al.,2001), application of law to biological findings (Hoyle et al. ,2002), Newton's Method (Berger et al.,2005) , apply for earth science data (Nigrini and Miller ,2007), election results (Torres et al.,2007, etc. Recent research papers have reported interesting new areas, where Benford's Law holds true, yet the tests are used by auditors in identifying fraudulent transitions with a check of digital frequencies in tax or other financial data against the Benford distribution (Nigrini, 1996).Benford's Law has been promoted for providing the auditor with a tool that is simple and effective for the detection of fraud(Durtschi, Hillison and Pacini ,2004). Some studies have also reported success in similar results for Benford's law being a useful tool for identifying suspect accounts for further analysis(Skousen, et al.,2004); for prices in ebay auctions(Giles ,2005) etc. The Benford Probability Distribution for fraud detection in supply chain to identify the abnormal transactions of small data set, which lead to red flags (Varma,T. N. and Khan, D. A.,2012). In this study they found that the total payment data of all vendors of a particular period of an organisation obey this distribution but on applying this law on individual vendor payment data from the sub set of data set, it was observed that there was violation of Benford probability distribution against one vendor data and this pattern strongly rejects the Benford's Law as per Chi-square test of significance. On further analysis, it was observed that the red flags of these fraudulent activities in Supply Chain Management were lack of clarity in job specification, selection of vendor without capability assessment and wrong inputs/ incomplete data in negotiation sheet to highlight capability of vendor etc. Computer-assisted audit techniques, like IDEA and ACL are most popular amongst auditors or detectors for applying Benford's Law for digital analysis on entire population of transactional data but it is costly and not easily available to all members of Supply Chain Networks of an organization. Although Microsoft Excel is an easily available software in any IT based organization, which has many powerful formulae, and it is very easy to run it as the features of an audit tool without any add-ins with limitation of 65,536 rows or records of data, 'batch' audit routines, easy to access logs for later reference, relating tables in data management tests etc without any extra cost on Supply Chain Management. Where Microsoft spread sheet is not available then it can be easily applicable with open source software.This article explains and demonstrates a practical approach of an effective method for searching hidden pattern or error, or manipulated data in a data-set from Supply Chain Network by using Excel Sheet or open source spread sheets and formula without any add-ins to perform the tasks required to Benford distribution test.

2. BENFORD'S LAW
Benford empirical studies lead to propose that in many real world applications the first digit D follows the probability distribution:

$$P\,(D=d_i) = \log(1 + 1/d_i) \quad d_i \in \{1, 2, ..., 9\} \qquad (1)$$

Table 1 shows the distribution of first significant digits with their expectation [E (d)] and variance [Var (d)]. The probabilities of first digit decline monotonically.

Table 1 Benford distribution of first and second digits

Digit (D)	1	2	3	4	5	6	7	8	9	E(d)	Var(d)
Probability % of 1st digit P(D)	30.1	17.6	12.5	09.7	07.9	06.7	05.8	05.1	04.6	3.440	6.057

A derivation of this formula is given in Cohen (1976). Hill (1995) extends this to the general law given by

$$P=(D1 \cdot\cdot\cdot Di = d1 \cdot\cdot\cdot di) = \log10 (1 + 1/(d1 \ldots di)) \quad (2)$$

Using the above formulae, it is observed that approximately 30.1% of the numbers have first digit one, while only 4.6% of the numbers have first digit nine. Actually, once the position of the digit in a number in increased, the gap converges to zero.

3. METHODOLOGY

Digital analysis (also first digit law or Benford's law) is a statistical technique regularly discussed in the professional guidance on fraud detection in general (Benford 1938; Nigrini 2000; Nigrini and Mittermaier 1997; Tackett 2007). A sample data has been taken from procurement cycle of Supply Chain Network, i.e. invoiced data of all vendors and number of invoices, which are popular targets for fraudsters. Although the amount of individual invoices is proprietary of the organization, it has been examined, whether the data collected for vendor payment was consistent with Benford's law or not. For this, simple formula has been applied in the Excel Sheet with following steps.

3.1 Selection and Import of Data

A sample data of financial transition of payment made, i.e. invoice values to all vendors in first column as shown in figure 1 of a Microsoft Excel Worksheet has been taken for a specified period of an organization. Entry of data in columns of sheet is be performed manually or pasting from another sheets (CTR- C for copying and then CTR- V for pasting), or importing data by help of data connection wizards. It is to be made sure that all data in this column must be numeric. Benford's Law does not apply to all numeric population as pre-assigned numbers, random in nature, non-established minimum or maximum value, clustering of values around a particular amount, etc.

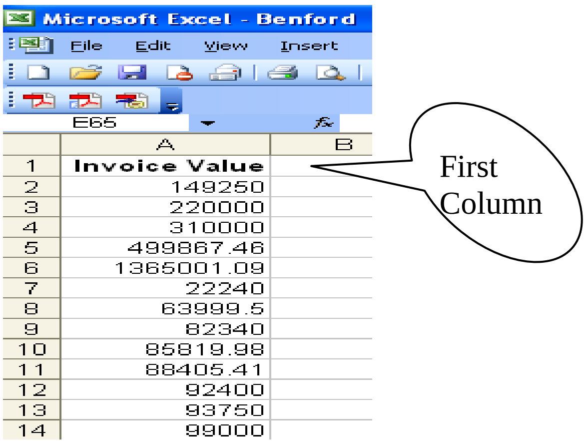


Figure 1 Vendor's Invoice Value

3.2 Extraction of First Significant Digit

Benford Probability Distribution is based upon the distribution of lead digit in set of naturally occurring numbers. These lead digits can be easily extracted by using Excel's LEFT function, whose Syntax is LEFT (text, [num_chars]), where text is the text string that contains the characters to be extracted, and Num_chars specifies the number of characters as required. Num_chars must be greater than or equal to zero. If num_chars is greater than the length of text, LEFT returns all of text and it is omitted, assumed to be 1. In this case, left most digit can be extracted by using formula for B2 cell as =LEFT(A2,1) or simply by formula =LEFT(A2) and then replicate this formula over the remaining rows at the end of the transactions.

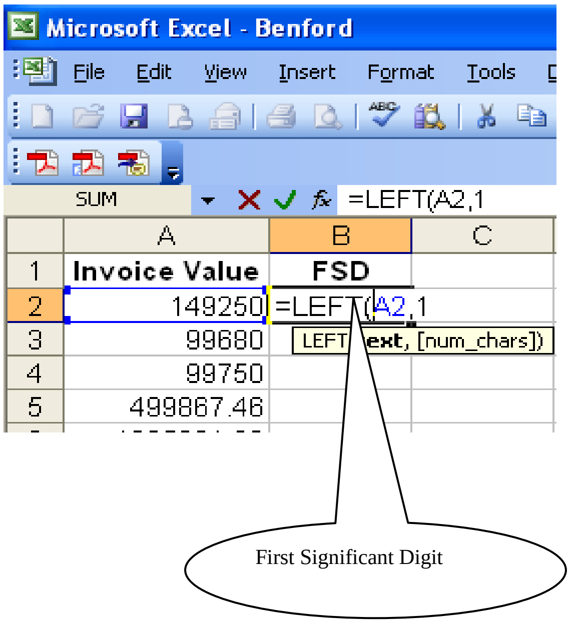
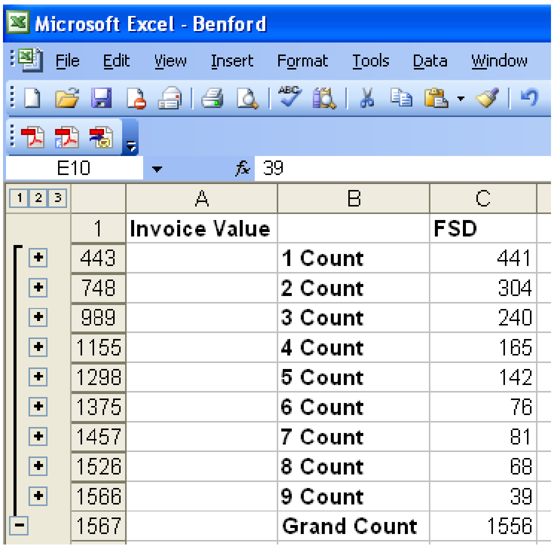


Figure 2 LEFT Function

3.3 Observed Frequency of First Significant Digit

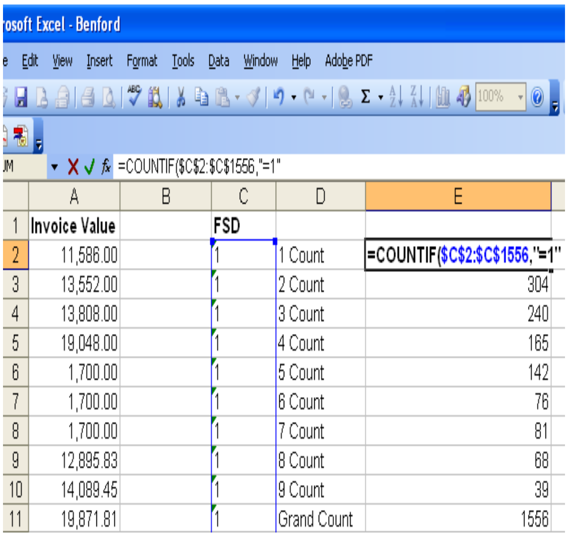
This task can be performed by using SUBTOTAL function. Firstly, sort out the First Significant Digit list, which was stored in second column so that the rows subtotals are grouped together, and the rows, which contain zeros in this column be eliminated. Microsoft Excel can automatically calculate subtotals and grand total values in a list. The list is outlined so that a summary report can be created by clicking the outline symbols to display and hide the detail rows for each subtotal. Although it is easily possible to use COUNTIF formula, which Syntax COUNTIF(range, criteria) where Range is the range of cells to be counted, and Criteria is the criteria in the form of a number, expression, or text that defines which cells will be counted. For example, to count all the '1s' in the range C2:C1556, the formula would be =COUNTIF($C$2:$C$1556,"=1"). This formula should then be replicated over the eight more cells, with counts for the digits 2 to 9.



Microsoft Excel - Benford

E10 | 39

	A	B	C
1	Invoice Value		FSD
443		1 Count	441
748		2 Count	304
989		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



Microsoft Excel - Benford

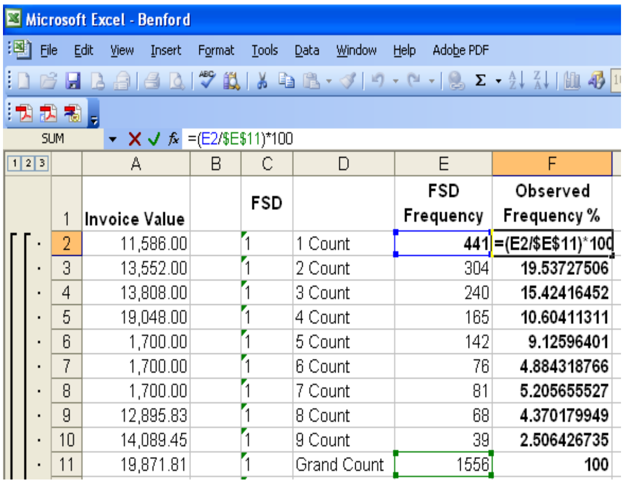
=COUNTIF($C$2:$C$1556,"=1"

	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	19,048.00		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	19,871.81		1	Grand Count	1556

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

3.3.1 Percentage Frequency Distribution of observed FSD

The frequency of all FSD can be calculated in previous steps by using SUBTOTAL or COUNTIF formula in Column E as appeared in figure 3a or 3b. For calculating percentage frequency of observed or actual data, the formula =(E2/$E$11)*100 can be used, where E2 cell contains frequency of FSD 1 and value of total observation in E11 cell, which is shown in Figure 4.



Microsoft Excel - Benford

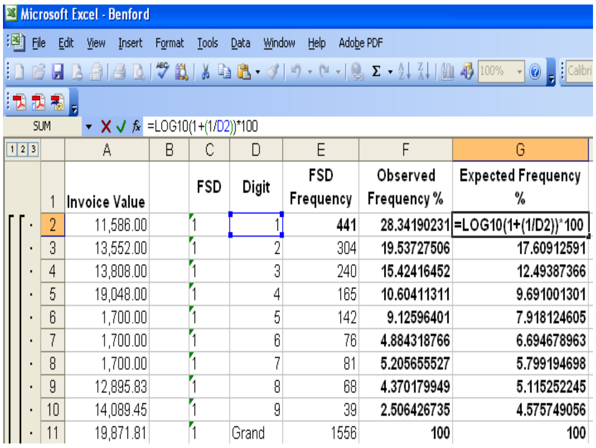
SUM | =(E2/$E$11)*100

	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	12,895.83		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

3.4 Percentage Frequency Distribution of expected FSD

The frequency of all FSD (1-9) can be calculated in previous steps appeared in Column E as shown in figure 3a or 3b. For removing Count text from Column D, selection of this column and finding Count (CTR-F) can be used, then replacing all by blank. The expected frequency percentage can be computed by using the Excel formula for Benford's Probability Distribution, i.e. =LOG10(1+(1/D2))*100) for digit 1. This formula should be replicated over the other First Significant Digits 2 to 9 as shown in Figure 5.



Microsoft Excel - Benford

SUM | =LOG10(1+(1/D2))*100

	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	12,895.83		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 %

3.5 Chi Square Test

Once the observed or actual and expected values have been obtained, the chi square can be computed as the sum of (O-E)*(O-E)/E. In this formula, O is the number of First Significant digits percentage observed and E is the number expected (which was computed) frequency percentage. The sum of each of these nine computed values provides the Chi Square value. This amount can be evaluated for 8 degrees of freedom to determine the acceptance or rejection of Benford's Law.

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 tabulated critical values for Chi square distribution at 1%, 5% and 10% level of significance for 8 degrees of freedom are 20.090, 15.507 and 13.362 respectively. If this value is greater than calculated value then it accepst the null hypothesis, otherwise it rejects the Benford's rule, where lies the chance of manipulation of data by fraudsters. By Figure 6 it is observed that the calculated value is much less than tabulated critical value, hence it accepts the null hypothesis, and the distribution is as per Benford's Probability Distribution.

3.6 Graph

Select the data from observed frequency and expected frequency percentage column and then after inserting chart wizard use chart tools as design, layout, format etc., whose output is shown below in figure 7. In first chart series being used as bar for actual frequency percentage and line for Benford frequency percentage and in second graph both distributions as bar chart.

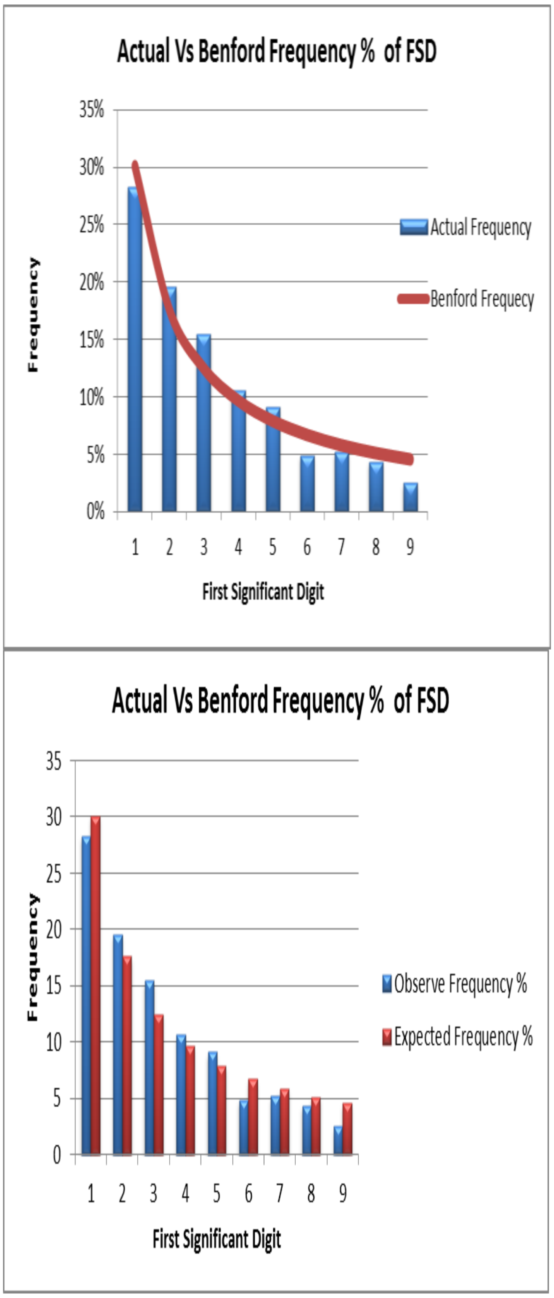
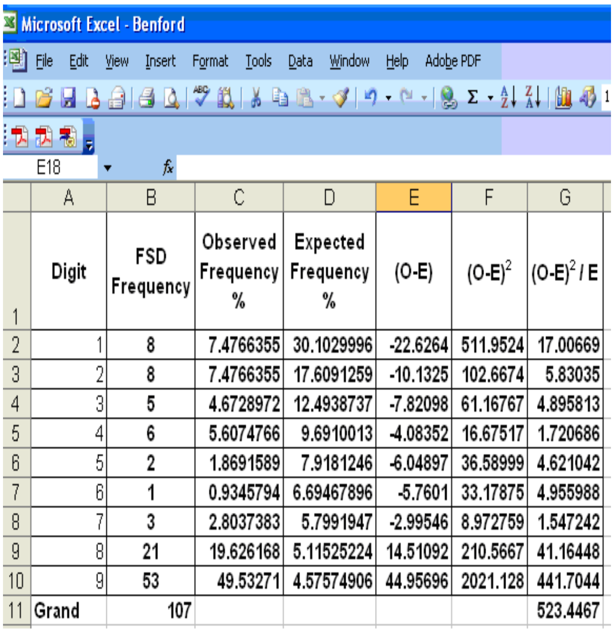


Figure 7 Actual (observed frequencies shown as line) and Benford frequency distribution %s (the expected frequencies from Benford's Law [in Eq. (1)] as bar) on y-axis and FSD on x-axis

3.7 Analysis of data Subset

On repeating steps 3.1 – 3.7 for each subset of invoice data set of vendors in order to determine the chi square values for each, it was observed that there was a huge difference between actual and Benford Frequency % as shown in Figure 8 for FSD 9, where probability of fraudulent transition of data is there in the Supply Chain Network.



	Digit	FSD Frequency	Observed Frequency %	Expected Frequency %	(O-E)	$(O-E)^2$	$(O-E)^2 / E$
2	1	8	7.4766355	30.1029996	-22.6264	511.9524	17.00669
3	2	8	7.4766355	17.6091259	-10.1325	102.6674	5.83035
4	3	5	4.6728972	12.4938737	-7.82098	61.16767	4.895813
5	4	6	5.6074766	9.6910013	-4.08352	16.67517	1.720686
6	5	2	1.8691589	7.9181246	-6.04897	36.58999	4.621042
7	6	1	0.9345794	6.69467896	-5.7601	33.17875	4.955988
8	7	3	2.8037383	5.7991947	-2.99546	8.972759	1.547242
9	8	21	19.626168	5.11525224	14.51092	210.5667	41.16448
10	9	53	49.53271	4.57574906	44.95696	2021.128	441.7044
11	Grand	107					523.4467

Figure 8 Chi Square Test for a data subset

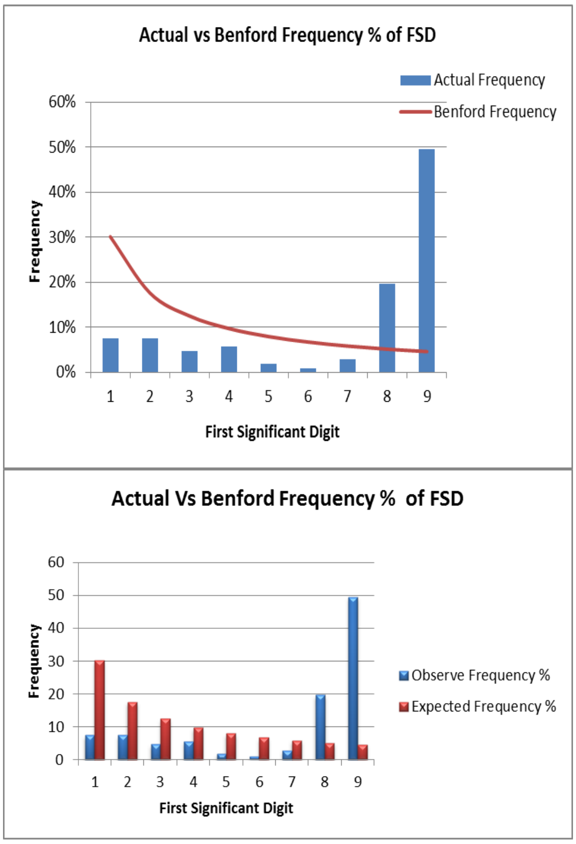


Figure 9 Actual (observed frequencies shown as line) and Benford frequency distribution %s (the expected frequencies from Benford's Law [in Eq. (1)] as bar) on y-axis and FSD on x-axis

4. RESULTS AND DISCUSSION

Analysis of data set of all vendors' invoice

From figure 6, it appears that this data set conforms to the Benford's probability distribution because the frequency of FSD is much more common than suggested by Benford's law. Visual inspection of Figure 7 makes it very tempting to argue that the distribution of vendor' invoice first digit percentages are almost identical with percentages predicted by the Benford's Probability Distribution, and Chi-square goodness-of-fit test conforms to Benford's Probability Distribution, which is of statistical significance.The calculated critical value is 2.866, which is much less than critical value for Chi square distribution at 1%, 5% and 10% level of significance for 8 degrees of freedom, which are 20.090, 15.507 and 13.362 respectively; hence, it accepts the null hypothesis. Therefore, it leads to correlate this pattern strongly with Benford's Distribution.

Analysis of data set of individual vendor

On analysis of large data set, it was found that the invoice data set of all vendors is obeying the Benford's Distribution. By repeating steps as discussed in 3.1 to 3.7, it has been observed that one vendor's data is violating Benford distribution because the difference between first significant digits frequency percentage and expected frequency distribution is very high as shown in figure 8, i.e. against digit 9. In this case the critical value for Chi square distribution at 1%, 5% and 10% level of significance for 8 degrees of freedom is very high, which strongly rejects Benford's distribution, and which is alarming for red flags in Supply Chain Network.

Fraud Analysis

By analysing the invoice data of this particular vendor for which FSD, the difference between observed frequency and expected frequency percentage has been found very high (FSD 9). The phenomena happened as most of payments were made between Rs.90000 to Rs.99999 to avoid the higher approval authorities limit, i.e. one lakh, which was the generation of procurement fraud. This occurred due to splitting of purchase orders, and the repeat orders were awarded to the vendors to manipulate payment with ulterior motives. On further analysis, it was observed that the these fraudulent activities were made due to incomplete or vague job specification, selection of vendor without proper capability assessment, and wrong inputs/ incomplete data in negotiation sheet to highlight capability of vendor, etc.

5. CONCLUSION

In the fraudulent transactions scenario or manipulation of data in Supply Chain cycle, data analysis can be applied to detect frauds with application of auditing software. This study demonstrates that the fraud detection can be as one of the proactive activities to help auditors or detecting agencies to identify the abnormal transactions, and to assist them in performing their tasks more effectively, efficiently and economically within a short span of time with help of simple Excel Functions, which will reduce the risk of fraud in supply chain.

6. REFERENCE

[1] Becker, P. (1982), ' Patterns in listings of failure-rate and MTTF values and listings of other data', IEEE Transactions on Reliability, Vol.31,pp. 132-134.

[2] 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)

[3] Berger, Arno, and Theodore P. Hill. (2007), 'Newton's Method Obeys Benford's Law', American Mathematical Monthly Vol. 114, No.7, pp.588-60I.

[4] Buck, B., Merchant, A. and S. Perez (1993), ' An illustration of Benford's first digit law using alpha decay half lives', European Journal of Physics, Vol. 14, pp. 59-63.

[5] Burke, J. and E. Kincanon (1991), ' Benford's law and physical constants : the distribution of initial digits', American Journal of Physics, Vol. 59, pp. 952.

[6] Dr. L. Kailasam (2011), 'Benford Distribution - An Effective Audit Tool' ,The Chartered Accountant, pp. 716-720

[7] Durtschi, Cindy, William Hillison, and Carl Pacini (2004), 'The Effective Use of Benford's Law to Assist in Detective Fraud in Accounting Data', UIC. Journal of Forensic Accounting, pp.17-34,Web http://www.uic.edu/ (Accessed,July 1,2013)

[8] Goutsmit, S.A. and W.H. Furry (1944). Significant figures of numbers in statistical tables. Vol. 154, pp. 800-801.

[9] Hoyle, D.C., Rattray, M., Jupp, R. and A. Brass (2002), ' Making sense of microarray data distributions', Bioinformatics, Vol. 18(4), pp. 576-584.

[10] 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.

[11] Newcomb, Simon.(1881), 'Note on the Frequency of Use of the Different Digits in Natural Numbers.' American Journal of Mathematics, Vol.4.1, pp. 39-40, Web http://www.jstor.org/ (Accessed, July 1,2013)

[12] 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.

[13] Nigrini, M.J., and S. J. Miller. (2007), 'Benford's Law applied to hydrology data - results and relevance to other geophysical data', Mathematical Geology, Vol. 39 (5),pp. 469-490.

[14] Nigrini, Mark J. Datas 2009 for Excel 2007 Program Details. Web. http://www.nigrini.com/datas_software.htm (Accessed ,,July 12,2013)

[15] Nigrini, Mark J. (1996), 'Taxpayer compliance application of Benford's law' , Journal of the American Taxation Association, Vol. 18(1), pp. 27-92

[16] Pietronero, L., Tosatti, E., Tosatti, V. and A. Vespignani (2001), 'Explaining the uneven distribution of numbers in nature', Physica A, Vol. 293,pp.297-304.

[17] Roukema, Bouewjin. (2009), 'Benford's Law Anomalies in the 2009 Iranian Presidential Election', Annals of Applied StatisticsWeb.http://arxiv.org/PS_cache/arxiv/pdf/0906/0906.2789v1.pdf (Accessed, July 12,2013)

[18] Saville, AD.(2006), 'Using Benford's Law to Detect Data Error and Fraud: An Examination of Companies Listed on the Johannesburg Stock Exchange', SAJEMS, Vol. 9.3,pp. 341-54. Web.http://www.scribd.com/doc/47789223/Saville-Using-282006-29

[19] Sinn, Hans-Werner (2002), 'Weber's Law and the Biological Evolution of Risk Preferences', http://www.huebnergeneva.org/documents/Sinn.pdf (Accessed,July 10,2013)

[20] Skousen, Christopher J., Guan, Liming and Wetzel, T. Sterling (2004), 'Anomalies and unusual patterns in reported eamings: Japanese managers round eamings', Journal of International FinancialManagement and Accounting, Vol. 15, Issue 3, pp. 212-234

[21] Tam Cho, Wendy K. and Gaines, Brian J. (2007), 'Breaking the (Benford's) Law: Statistical Fraud Detection in Campaign Finance' American Statistician, Vol. 61, Issue 3, August, p. 218-223

[22] Thomas, Jacob K. (1989), 'Unusual Patterns in Reported Earnings', The Accounting Review, Vol. LXIV.4 pp.773-87.

[23] Varian, H.R. (1972), 'Benford's Law', American Statistician, Vol. 26, pp. 65-66

[24] Varma, T.N. and Khan , D.A. (2012) , 'Fraud Detection in Supply Chain using Benford Distribution.', International Journal of Research in Management, Vol. 5 No. 2, pp.90-96

[25] Weiss, Neil A. (1999), Introductory Statistics, Fifth Edition, Addison Wesley, Reading, MA, pp.A-16

[26] Wikipedia (2012), Weber-Fechner law, http://en.wikipedia.org/wiki/Weber-Fechner_Law (Accessed, July 12,2013)

[27] http://office.microsoft.com/en-us/excel/ (Accessed, May 8 ,2013).