REGIONAL TRAINING INSTITUTE KOLKATA
Indian Audit & Accounts Department

सत्यमेव जयते

USING BENFORD'S LAW IN AUDIT

RESEARCH PAPER

Introduces the basic concept of Benford's law, its application in audit, especially in detecting frauds and provides step by step guide to use Benford's law with MS Excel & IDEA software.

पूर्णमदः पूर्णमिदं पूर्णात्पूर्णमुदच्यते ।
पूर्णस्य पूर्णमादाय पूर्णमेवावशिष्यते ॥

3rd MSO building, CGO Complex, 5th Floor ("A" Wing) DF Block, Salt Lake, Sector-I, Kolkata 700 064
Tel +91 33 2321 3907 / +91 33 2321 6708
Fax (033) 2321 6709
http:/rtikolkata.cag.gov.in

TABLE OF CONTENTS

Contents

Preface

Benford's law, also called the first-digit law, was made famous in 1938 by Physicist Frank Benford, who after observing sets of naturally occurring numbers, discovered a surprising pattern in the occurrence frequency of the digits one through nine as the first number in a list. In essence, the law states that in numbered lists providing real-life data (e.g., a journal of cash disbursements and receipts, contract payments, or credit card charges), the leading digit is one almost 33 percent (i.e., one third) of the time. On the other hand, larger numbers occur as the leading digit with less frequency as they grow in magnitude to the point that nine is the first digit less than 5 percent of the time.

In the 1970s, Hal Varian, a professor at the University of California's Berkeley School of Information, suggested that the law could be used to detect possible fraud in lists providing socioeconomic information. Since then, Benford's law has been applied to large numbers of data to detect unusual patterns that are often the result of errors or, worse, fraud. As part of their work, auditors often employ tools and scientific methods that enable them to detect instances of fraud. Although the use of Benford's law might seem daunting at first, auditors don't need to have advanced degrees or an expensive data analysis tool to use Benford's law as part of their fraud investigations — this task can be effectively and efficiently performed using Microsoft Excel and IDEA.

In this research paper, we introduce the Benford’s law phenomenon in simple terms and explain how we can apply it with IDEA or Microsoft Excel. To drive home our point, we have analysed three different datasets to show how Benford law analysis could make sense in audit.

Shri Sandeep Singh
Principal Director
February 29, 2016

Background

Benford‘s Law is a wonderful, albeit underutilized, tool in the auditor‘s arsenal. It looks at data in a way that is different from the normal auditor‘s point of view. Instead of looking for the largest amounts or a sufficient amount of coverage, Benford‘s Law allows the auditor to evaluate the digits that make up the numbers themselves. Benford‘s Law is not used to detect defective deliveries, contract rigging or off-book transactions, bribes, kickbacks, or asset thefts—but it is used to find payroll, expense, sales, accounts receivable, fixed asset, and journal entry anomalies as well as industry-specific account anomalies.

Benford‘s Law is based on the fact that many numbers normally used in business (and elsewhere) are not random, but rather follow some ordered progression. For example, a chart showing wealth will show that it is not uniformly distributed; a few people have much wealth and many people have less wealth. Sales, inventory, and disbursements are also not uniformly distributed. Benford‘s Law uses this fact to help point to fraud, inefficiencies, and other forms of data manipulation.

Back in the 1800s an astronomer, Simon Newcomb, noticed that the earlier pages in books of logarithm tables were more worn than the later pages. Logarithm books were used to multiply (and divide) large numbers. Newcomb posited that numbers beginning with lower digits were used more often than numbers beginning with higher digits. He published ―Note on the Frequency of Use of the Different Digits in Natural Numbers in 1881. He offered his observation but gave neither a use nor a proof. The article was promptly forgotten.

In the years following the Great Depression, without apparent knowledge of Newcomb or his article, Frank Benford noticed the same thing. He was working for General Electric and had a lot of time on his hands. He decided to test his hypothesis. Benford analyzed 20 lists of large data sets (total of 20,229 data points) and 10 lists of smaller data sets (total of 2,968 data points). These lists came from random sources, such as the numbers in an issue of a magazine and death rates, as well as from sources that were not random—populations, horsepower lost, and molecular weight, and so on.

Benford published his observation and proof as —The Law of Anomalous Numbers in 1938. Though he did not identify any uses, Benford‘s article had a better reception than Newcomb‘s and we now have Benford‘s Law.

It was not until 1988 that Benford's Law was cited in a survey by Charles Carslaw. Carslaw assumed that managers round off the company's earnings if these are slightly below a certain psychological threshold (for example, earnings of 19.9 million are rounded off to 20 million). Should such rounding occur, assumed Carslaw, the number 9 as the second digit in a list of company earnings would occur rather rarely, whereas the number 0 as the second digit would occur relatively frequently. In this case, Carslaw used the frequencies calculated by Benford as a benchmark for the results of his analyses. They resulted in the fact that in a list of company earnings the number 0 as the second digit would occur relatively often and the number 9 relatively seldom (compared to Benford's Law).

A few more things had to be established before Benford‘s Law could be of any use to those on the finance side of things. In 1961 Roger Pinkham proved that Benford‘s Law held true no matter what the unit of measurement. That means that it does not matter whether you measure items in yen, dollars, feet, miles, or meters. In 1972, Hal Varian found that you could use it to detect fraud in socioeconomic data. In the 1980s it was used to detect the reasonableness of round numbers and it was also found that invented numbers do not conform. Carslaw, in 1988, found that companies in New Zealand were not completely honest in their annual financial

In 1993 Christian and Gupta discovered another interesting phenomenon with reference to the practice. They analyzed data of tax figures in order to discover signs for tax evasion. They assumed taxpayers intended to force their taxable income into the next highest graduated tax rate. Thus, the values of the graduated tax rates created the threshold values which were to be supervised in the income tax charts. Any reduction of the taxable income by a couple of US Dollars to below a certain graduated tax rate could possibly lead to substantial tax savings. According to analyses of income tax filings, comparably more taxpayers have an income which ends with the digits 40-49 and 90-99 than an income which ends with the digits 50 59 and 00-09. This indicates that US taxpayers intended (and surely still intend) to force their income below the next highest graduated tax rate of the US income tax charts.

reports. (In the succeeding years, this finding was verified by others and for other countries). Also, in 1988, Ted Hill found that people cannot create numbers and still conform to Benford's Law. The real breakthrough for auditors came in 1994 when Mark Nigrini, a South African chartered accountant, codified a practical use. His 1992 thesis showed that accounting data conforms to Benford's Law. In 1994 he assisted tax agencies find suspect returns. From there, he worked with companies to find fraud and continued his research to expand the applications of Benford's Law. Currently, Benford's Law analysis is admissible in U.S. courts at all levels and has been used internationally as well.

*In 1993, in State of Arizona v. Wayne James Nelson (CV92-18841), the accused was found guilty of trying to defraud the state of nearly $2 million. Nelson, a manager in the office of the Arizona State Treasurer, argued that he had diverted funds to a bogus vendor to demonstrate the absence of safeguards in a new computer system. The amounts of the 23 checks issued are shown in **exhibit.***

Exhibit

The table lists the checks that a manager in the office of the Arizona State Treasurer wrote to divert funds for his own use. The vendors to whom the checks were issued were fictitious.

Date of Check	Amount
October 9, 1992	$ 1,927.48
	27,902.31
October 14, 1992	86,241.90
	72,117.46
	81,321.75
	97,473.96
October 19, 1992	93,249.11
	89,658.17
	87,776.89
	92,105.83
	79,949.16
	87,602.93
	96,879.27
	91,806.47
	84,991.67
	90,831.83
	93,766.67
	88,338.72
	94,639.49
	83,709.28
	96,412.21
	88,432.86
	71,552.16
TOTAL	$ 1,878,687.58

Because human choices are not random, invented numbers are unlikely to follow Benford's law. Here are some divergent signs that Benford's law would have drawn attention to:

As is often the case in fraud, the embezzler started small and then increased dollar amounts.

Most of the amounts were just below $100,000. It's possible that higher dollar amounts received additional scrutiny or that checks above that amount required human signatures instead of automated check writing. By keeping the amounts just below an additional control threshold, the manager tried to conceal the fraud.

The digit patterns of the check amounts are almost opposite to those of Benford's law. Over 90% have 7, 8 or 9 as a first digit. Had each vendor been tested against Benford's law, this set of numbers also would have had a low conformity, signaling an irregularity.

The numbers appear to have been chosen to give the appearance of randomness. Benford's law is quite counterintuitive; people do not naturally assume that some digits occur more frequently. None of the check amounts was duplicated; there were no round numbers; and all the amounts included cents. However, subconsciously, the manager repeated some digits and digit combinations. Among the first two digits of the invented amounts, 87, 88, 93 and 96 were all used twice. For the last two digits, 16, 67 and 83 were duplicated. There was a tendency toward the higher digits; note that 7 through 9 were the most frequently used digits, in contrast to Benford's law. A total of 160 digits were used in the 23 numbers. The counts for the ten digits from

0 to 9 were 7, 19, 16, 14, 12, 5, 17, 22, 22, and 26, respectively. An Auditor familiar with Benford's law could have easily spotted the fact that these numbers—invented to seem random by someone ignorant of Benford's law—fall outside expected patterns and thus merit closer examination.

Biases in corporate data*. In one company's accounts payable data, there was a large first-two digit spike (excess of actual over expected) at 24. An analysis showed that the amount $24.50 occurred abnormally often. The audit revealed that these were claims for travel expenses and that the company had a $25 voucher requirement. Employees were apparently biased toward claiming $24.50.*

Ducking authorization levels*. Sometimes managers concentrate their purchases just below their authorization levels so their choices won't be scrutinized. Managers with $3,000 purchasing levels might have a lot of invoices for $2,800 to $2,999, which would show up in data analysis by spikes at 28 and 29.*

During one bank audit, the auditors analyzed the first two digits of credit card balances written off as uncollectible. The graph showed a large spike at 49. An analysis of the related dollar amounts (that is, from $480 to $499 and from $4,800 to $4,999) showed that the spike was caused mainly by amounts between $4,800 and $4,999, and that one officer was responsible for the bulk of these write-offs. The write-off limit for internal personnel was $5,000. It turned out that the officer was operating with a circle of friends who would apply for credit cards. After they ran up balances of just under $5,000, he would write the debts off.

The logic behind Benford's Law.

Benford's Law states that many sets of numbers follow a predictable pattern, no matter what their origin or subject. The digit 1 will be the leading digit approximately 30% of the time. A leading digit is simply the left-most digit in a number. For example, the leading digit of —123 is 1 and the leading digit of —0.0552 is 5 (0 cannot be a leading digit). In fact, digits 1, 2, or 3 lead approximately 60% of the time. 9 as a leading digit appears only 4.5% of the time. Why? Well, think of the following question: You won ₹ 10 lakh in the lottery (congratulations!) and decide to invest in something that gives a guaranteed 10% return. It will take approximately 7.3 years to turn your ₹ 10 lakh into ₹ 20 lakh. When you reach ₹ 50 lakh it takes less than 2 years to reach ₹ 60 lakh. When you reach ₹ 90 lakh, it will take just over 1 year to reach ₹ 1 crore and then the cycle starts over again.

Again, the question why does the digit 1 as the first digit occur more frequently in a natural population than the digits 8 or 9 can be explained with a simple example. If someone starts counting upwards in whole numbers, starting with 1, the 1 will at any point in time occur proportionally (amount of the numbers starting with 1 divided by the amount of numbers starting with 9) more frequently or equally frequently as the 9. Furthermore, the phenomenon can be explained by the fact that in a natural series of numbers (age development of people, population development of a city) there is indeed a larger amount of smaller than bigger values.

Current computing capabilities make Benford‘s Law fairly easy to implement. Research has addressed its use in many different fields and has expanded on its applicability.

WHAT’S THE MATH?

In 1938, the physicist Frank Benford laid the foundation for the empirical law named after him (Benford's Law). He analyzed the distribution of the first digit in a natural population of numbers and discovered that the number 1 as the first digit of every number occurs in 30.6 % of the cases compared to the number 9 as first digit in only 4.5 % of the analyzed cases. Thus, Benford's main statement is that the frequency of the first digit in a population's numbers decreases with the increasing value of the number in the first digit.

In the course of further analysis and with the help of some statistical assumptions, Benford was able to prove empirically that his discovery includes a legality which facilitates, in the form of mathematical formulas, the derivation of the probable frequency of occurrence of any digit or any numerical combination at the beginning of numbers (first digit) from a number series. The formula[1] is:

$$P(d)=\log(1 + 1/d)$$

Where P(d) stands for the probability that a number starts with the digit d.

[1] ***Kindly note:*** *Detailed mathematical aspects are not discussed in this paper. The emphasis is more on how we can apply Benford’s law. Please refer to* ***Appendix 1*** *contains the expected frequencies of the digits in second to 4th position.*

Application of Benford's Law in Audit and detection of Frauds

Benford's Law helps to address the ―expectation gap. As auditors, we all know that we do NOT look for fraud, though if we find it, we will report it. We also know that the public expects us to look for fraud. Audit programs and most audit steps are not geared to find fraud. It would be like finding a needle in a haystack or having the winning lottery number. However, Benford's Law tests 100% of transactions and enable auditors to assess risks of being provided with erroneous, manipulated or fraudulent data. Based on such risk assessment, the auditor may then choose to apply further and specific audit procedures.

A review of various guidance to auditor which could warrant applying Benford Law tests to audit of financial statements or assessment of non-financial reports (especially for data utilized in regularity audit or performance audit) are discussed below.:

- **ISSAI 1315/ISA 315 *Identifying and Assessing the Risks of Material Misstatement through Understanding the Entity and Its Environment***

  The auditor shall perform risk assessment procedures to provide a basis for the identification and assessment of risks of material misstatement at the financial statement and assertion levels (*para 5*). The risk assessment procedures shall include Analytical procedures (*para 6*). The auditor shall identify and assess the risks of material misstatement at the assertion level for classes of transactions, account balances, and disclosures to provide a basis for designing and performing further audit procedures (*para 25*). Analytical procedures performed as risk assessment procedures may identify aspects of the entity of which the auditor was unaware and may assist in assessing the risks of material misstatement in order to provide a basis for designing and implementing responses to the assessed risks. Analytical procedures performed as risk assessment procedures may include both financial and non-financial information (*para A7*). Analytical procedures may help identify the existence of unusual transactions or events, and amounts, ratios, and trends that might indicate matters that have audit implications. Unusual or unexpected relationships that are identified may assist the auditor in identifying risks of material misstatement, especially risks of material misstatement due to fraud (*para A8*). However, when such analytical procedures use data aggregated at a high level (which may be the situation with analytical procedures performed as risk assessment procedures), the results of those analytical procedures only provide a broad initial indication about whether a material misstatement may exist. Accordingly, in such cases, consideration of other information that has been gathered when identifying the risks of material misstatement together with the results of such analytical procedures may assist the auditor in understanding and evaluating the results of the analytical procedures (*para A9*).

- **ISSAI 1330 / ISA 330 *The Auditor's Responses to Assessed Risks***

  The auditor shall design and implement overall responses to address the assessed risks of material misstatement at the financial statement level (*para 5*).

- **ISSAI 1520 / ISA 520 *Analytical Procedures***

  If analytical procedures performed in accordance with this ISA identify fluctuations or relationships that are inconsistent with other relevant information or that differ from expected values by a significant amount, the auditor shall investigate such differences by: (a) Inquiring of management and obtaining appropriate audit evidence relevant to management's responses; and (b) Performing other audit procedures as necessary in the circumstances (*para 7*). The auditor's substantive procedures at the assertion level may be tests of details, substantive analytical procedures, or a combination of both. The decision about which audit procedures to perform, including whether to use substantive analytical procedures, is based on the auditor's judgment about the expected effectiveness and efficiency of the available audit procedures to reduce audit risk at the assertion level to an acceptably low level (*para A4*). The auditor's determination of the amount of difference from the expectation that can be accepted without further investigation is influenced by materiality and the consistency with the desired level of assurance, taking account of the possibility that a misstatement, individually or when aggregated with other misstatements, may cause the financial statements to be materially misstated. ISSAI 1330/ISA 330 requires the auditor to obtain more persuasive audit evidence the higher the auditor's assessment of risk. Accordingly, as the assessed risk increases, the amount of difference considered acceptable without investigation decreases in order to achieve the desired level of persuasive evidence (*para A16*).

- **ISSAI 1530 / ISA 530 *Audit Sampling***

  When designing an audit sample, the auditor shall consider the purpose of the audit procedure and the characteristics of the population from which the sample will be drawn (*para 6*). The auditor shall investigate the nature and cause of any deviations or misstatements identified, and evaluate their possible effect on the purpose of the audit procedure and on other areas of the audit (*para 12*).Further, in analyzing the deviations and misstatements identified, the auditor may observe that many have a common feature, for example, type of transaction, location, product line or period of time. In such circumstances, the auditor may decide to identify all items in the population that possess the common feature, and extend audit procedures to those items. In addition, such deviations or misstatements may be intentional, and may indicate the possibility of fraud (*para A17*).

CAG's Standing Order on Role of Audit in Relation to Cases of Fraud and Corruption

Examination of system for detection and prevention of fraud and corruption will be an integral part of a regularity audits and also of performance audits, when it forms one of the audit (sub) objectives (*para 1.1*). Fraud examination is a part of the normal auditing procedures and includes being alert for situations, control weaknesses, inadequacies in record keeping, errors and unusual transactions or results, which could be indicative of fraud, corruption, improper expenditure or lack of probity; and focusing audit strategy on areas and operations prone to fraud and corruption by developing effective high risk indicators for fraud (*para 2*).

Fraud may involve manipulation, falsification or alteration of records or documents; misappropriation/ misapplication of assets; suppression or omission of the effects of transactions from records or documents; recording of transaction without substances; and misapplication of accounting policies (*para 3.6*). The mandate of Government Audit is broader than solely that of financial statement auditor and includes responsibility for verification of regularity and performance. Hence, the auditor should be aware of the possibility of fraud not only in the preparation and presentation of financial statements but other areas covered by regularity (compliance) and performance audits as well (*para 3.7*).

Professional skepticism is an attitude that includes a questioning mind and a critical assessment of audit evidence. Professional skepticism requires an ongoing questioning of whether the information and audit evidence obtained suggests the existence of fraud having a material effect on audit findings/ opinion (*para 12.1)*. The field offices should carry out an independent risk assessment and prioritize their audit planning accordingly. This should include consideration of any information received from the public or media on suspected cases of fraud and corruption. The audit plans should focus on high risk areas. Such high risk areas include Revenue receipts, cash management, expenditure on AC bills, financial statements, operating information etc. (*para 13.1and 13.8)*.

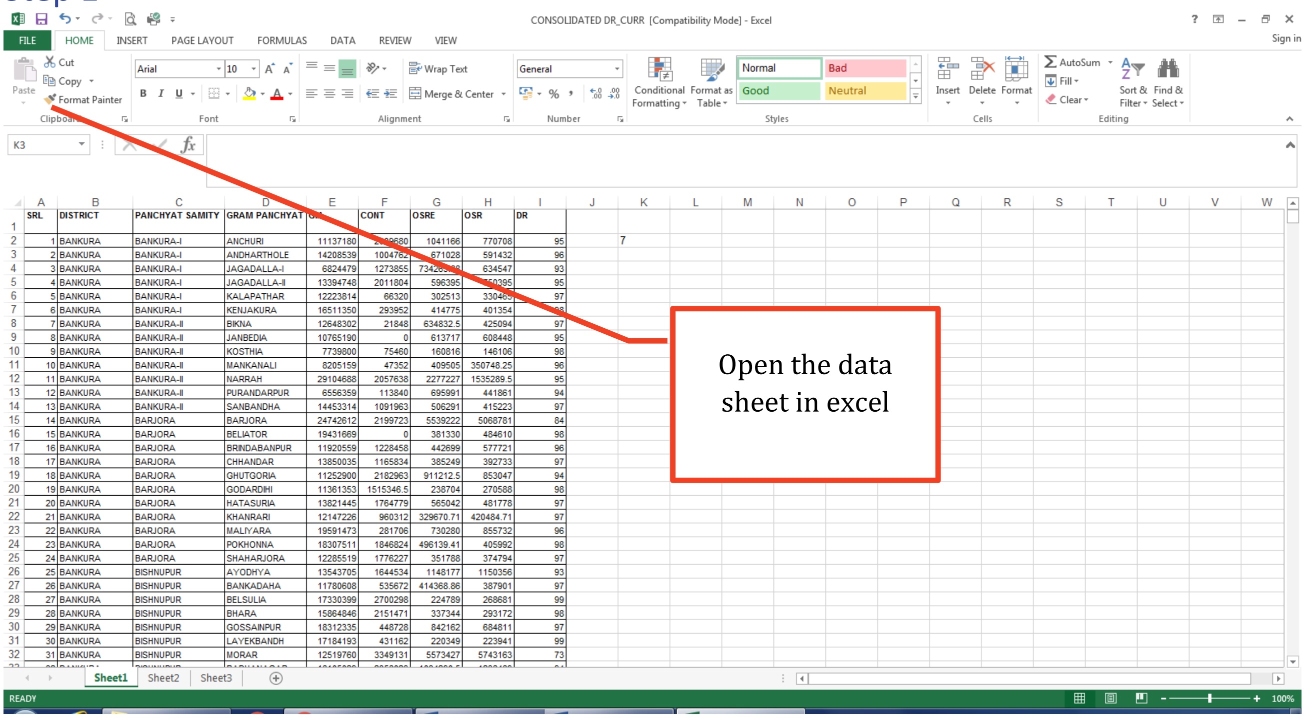
Though audit cannot insure against frauds, the possibility of their occurrence should be kept in mind while preparing for and conducting audit, by maintaining an attitude of professional skepticism (*para 13.2).* The auditor may keep in view that when a fraud is conducted there is a deliberate effort to conceal the facts and distract the auditor (*para 14.3)*. Analytical procedures are helpful in identifying the existence of unusual transactions or events, and amounts, ratios, and trends that might indicate matters having audit implications. When performing analytical procedures at the planning stage or during the course of audit, unusual or unexpected relationships may indicate risk of fraud. Fraud detection measures need to be built in the audit procedures, so that during the audit, the auditor can highlight a transaction for a possible fraud or identify such consistent system failures, which can lead to a fraud (*para 15.2and 15.3)*.

Benford law helps in identification of possible manipulation or misreporting of financial data. And therefore helps the auditor with possible leads for further more elaborate examination of transactions to reveal possible instances of fraud. Best of all, Benford's Law does not require specialized (expensive) software.

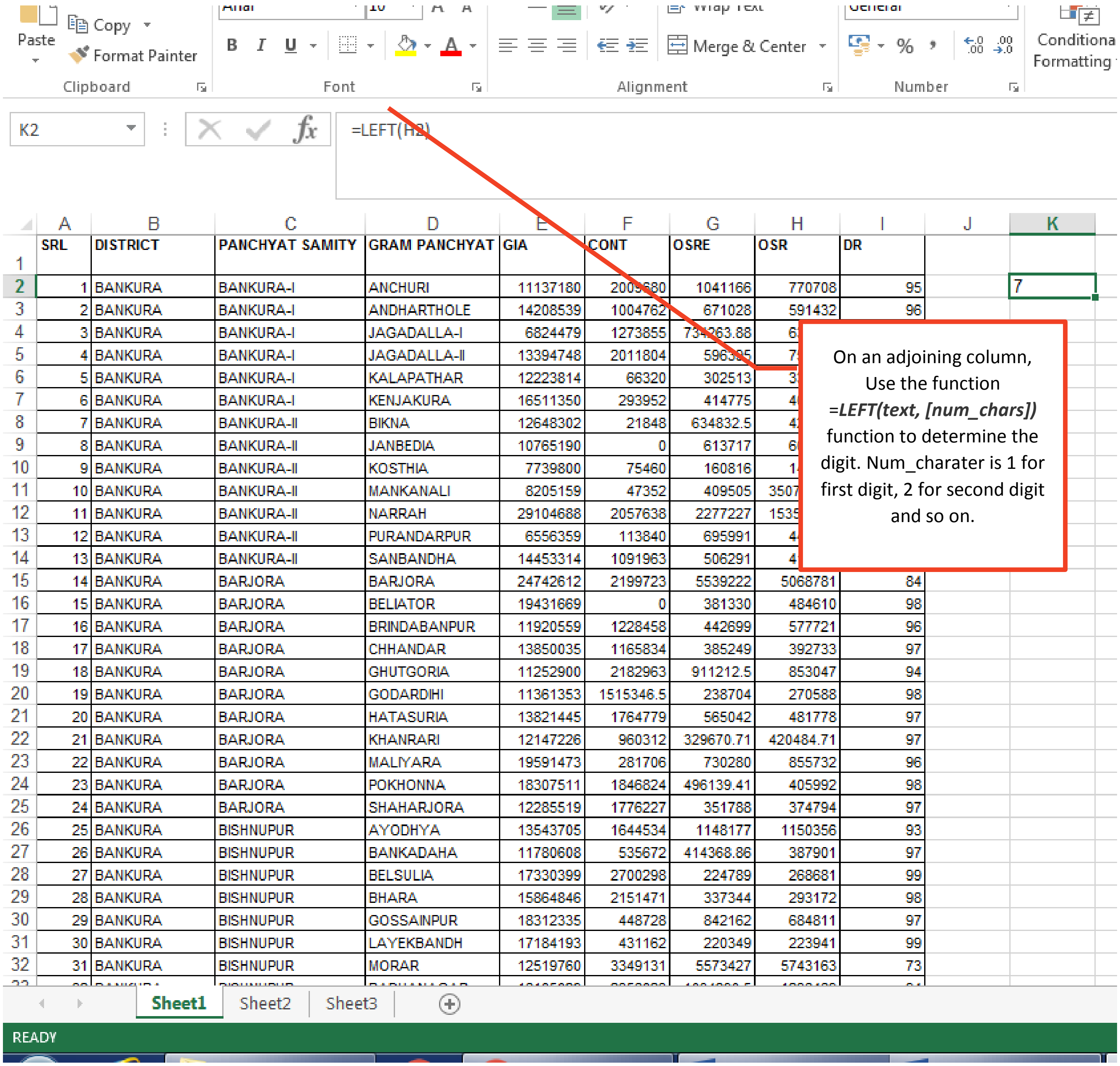
In the next two sections we examine how Microsoft Excel or IDEA software could be used to perform compliance of data set to Benford law.

Using Benford's Law: EXCEL

Step 1



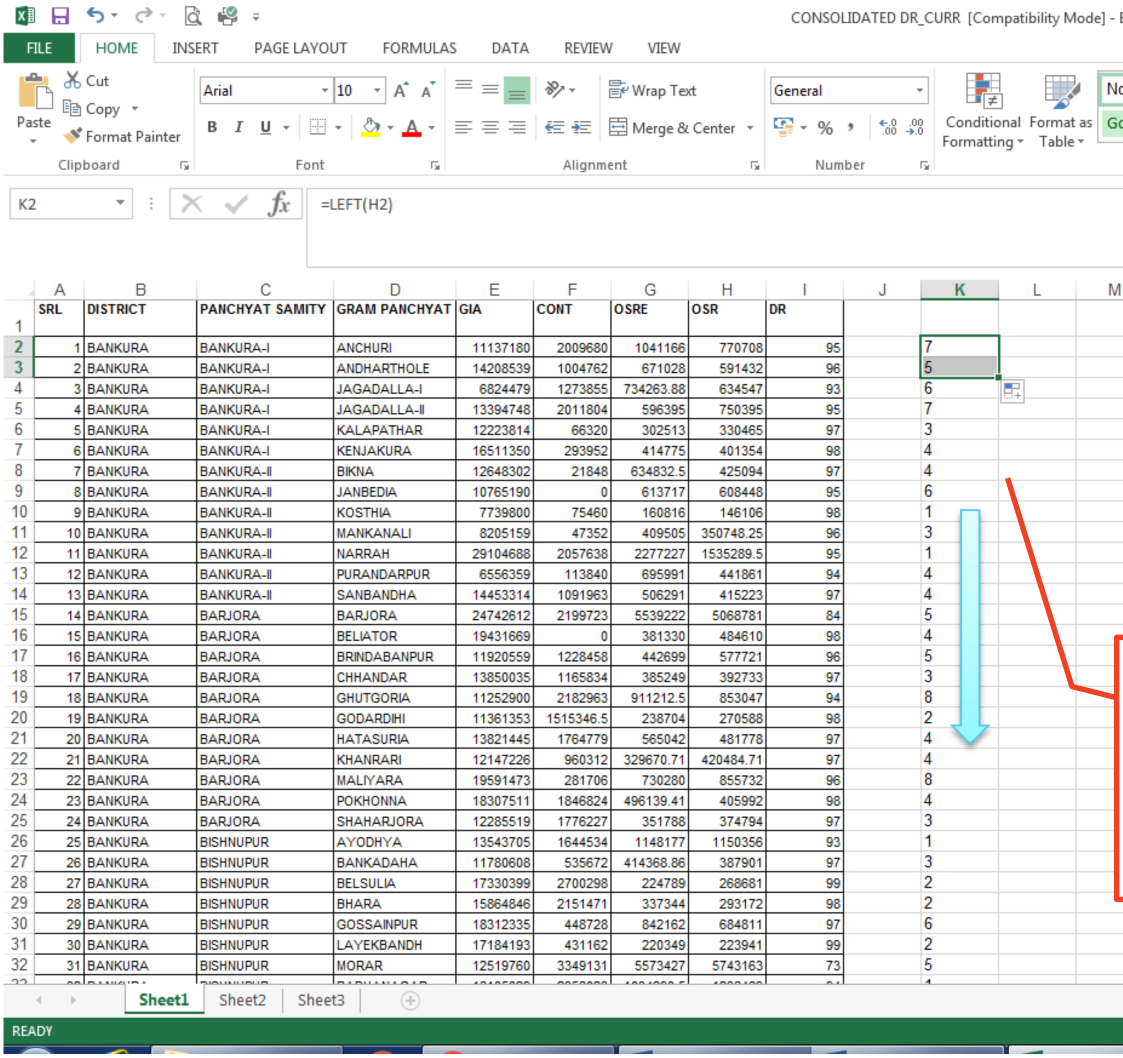
Step 2



SRL	DISTRICT	PANCHYAT SAMITY	GRAM PANCHYAT	GIA	CONT	OSRE	OSR	DR		
1	BANKURA	BANKURA-I	ANCHURI	11137180	200[illegible]680	1041166	770708	95		7
2	BANKURA	BANKURA-I	ANDHARTHOLE	14208539	1004762	671028	591432	96		
3	BANKURA	BANKURA-I	JAGADALLA-I	6824479	1273855	[illegible]	[illegible]	[illegible]		
4	BANKURA	BANKURA-I	JAGADALLA-II	13394748	2011804	[illegible]	[illegible]	[illegible]		
5	BANKURA	BANKURA-I	KALAPATHAR	12223814	66320	302513	[illegible]	[illegible]		
6	BANKURA	BANKURA-I	KENJAKURA	16511350	293952	414775	[illegible]	[illegible]		
7	BANKURA	BANKURA-II	BIKNA	12648302	21848	634832.5	[illegible]	[illegible]		
8	BANKURA	BANKURA-II	JANBEDIA	10765190	0	613717	[illegible]	[illegible]		
9	BANKURA	BANKURA-II	KOSTHIA	7739800	75460	160816	[illegible]	[illegible]		
10	BANKURA	BANKURA-II	MANKANALI	8205159	47352	409505	[illegible]	[illegible]		
11	BANKURA	BANKURA-II	NARRAH	29104688	2057638	2277227	[illegible]	[illegible]		
12	BANKURA	BANKURA-II	PURANDARPUR	6556359	113840	695991	[illegible]	[illegible]		
13	BANKURA	BANKURA-II	SANBANDHA	14453314	1091963	506291	[illegible]	[illegible]		
14	BANKURA	BARJORA	BARJORA	24742612	2199723	5539222	5068781	84		
15	BANKURA	BARJORA	BELIATOR	19431669	0	381330	484610	98		
16	BANKURA	BARJORA	BRINDABANPUR	11920559	1228458	442699	577721	96		
17	BANKURA	BARJORA	CHHANDAR	13850035	1165834	385249	392733	97		
18	BANKURA	BARJORA	GHUTGORIA	11252900	2182963	911212.5	853047	94		
19	BANKURA	BARJORA	GODARDIHI	11361353	1515346.5	238704	270588	98		
20	BANKURA	BARJORA	HATASURIA	13821445	1764779	565042	481778	97		
21	BANKURA	BARJORA	KHANRARI	12147226	960312	329670.71	420484.71	97		
22	BANKURA	BARJORA	MALIYARA	19591473	281706	730280	855732	96		
23	BANKURA	BARJORA	POKHONNA	18307511	1846824	496139.41	405992	98		
24	BANKURA	BARJORA	SHAHARJORA	12285519	1776227	351788	374794	97		
25	BANKURA	BISHNUPUR	AYODHYA	13543705	1644534	1148177	1150356	93		
26	BANKURA	BISHNUPUR	BANKADAHA	11780608	535672	414368.86	387901	97		
27	BANKURA	BISHNUPUR	BELSULIA	17330399	2700298	224789	268681	99		
28	BANKURA	BISHNUPUR	BHARA	15864846	2151471	337344	293172	98		
29	BANKURA	BISHNUPUR	GOSSAINPUR	18312335	448728	842162	684811	97		
30	BANKURA	BISHNUPUR	LAYEKBANDH	17184193	431162	220349	223941	99		
31	BANKURA	BISHNUPUR	MORAR	12519760	3349131	5573427	5743163	73		

Please note that if there are numbers that are less than 0, the same could be converted to non-decimal number before applying Benford's law. The Benford's law does not distinguish between a less than one decimal number or a natural number. For instance, 0.3690 could be converted as 3690 and the Benford's law would still hold.

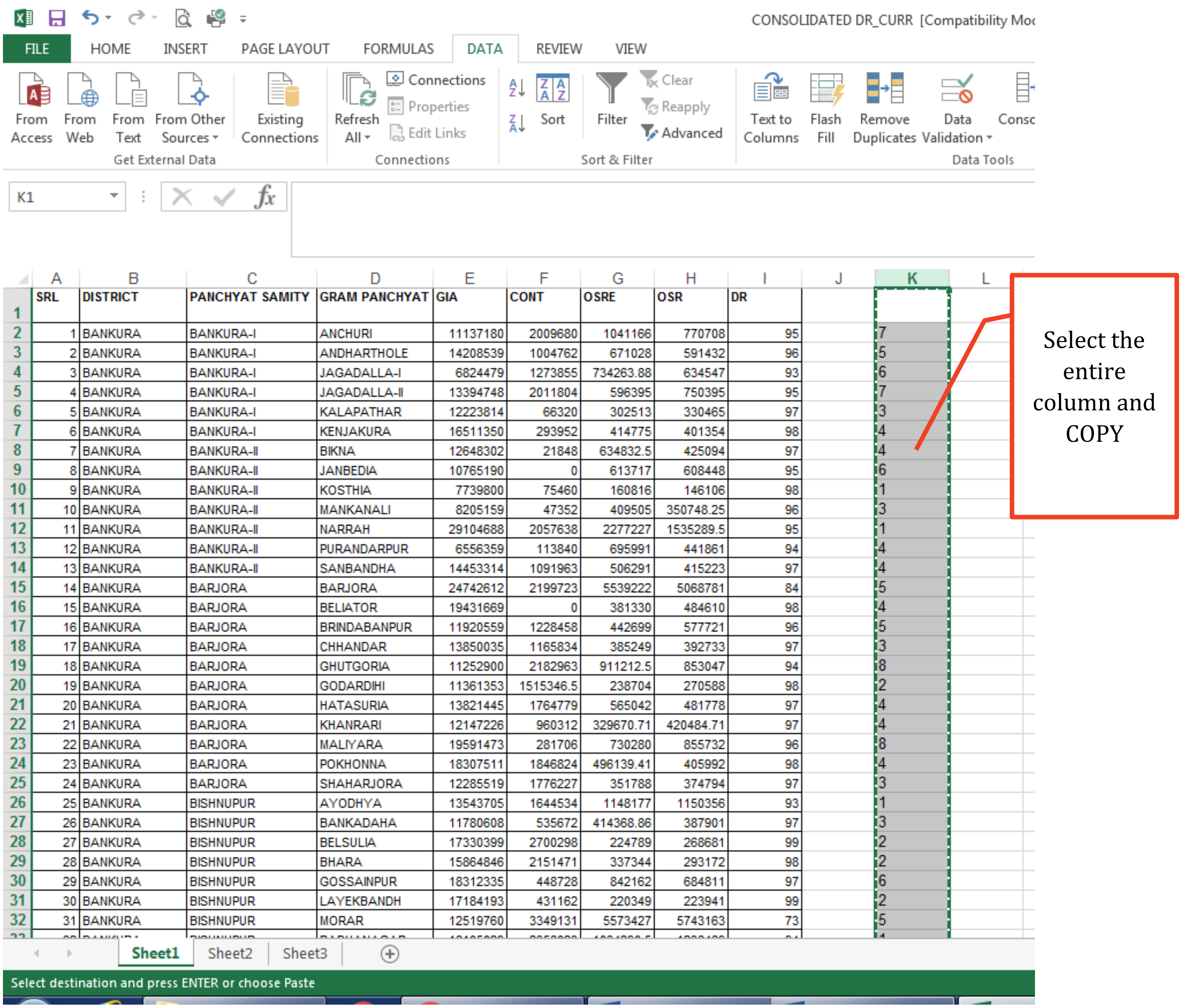
Step 3



	A	B	C	D	E	F	G	H	I	J	K
1	SRL	DISTRICT	PANCHYAT SAMITY	GRAM PANCHYAT	GIA	CONT	OSRE	OSR	DR		
2	1	BANKURA	BANKURA-I	ANCHURI	11137180	2009680	1041166	770708	95		7
3	2	BANKURA	BANKURA-I	ANDHARTHOLE	14208539	1004762	671028	591432	96		5
4	3	BANKURA	BANKURA-I	JAGADALLA-I	6824479	1273855	734263.88	634547	93		6
5	4	BANKURA	BANKURA-I	JAGADALLA-II	13394748	2011804	596395	750395	95		7
6	5	BANKURA	BANKURA-I	KALAPATHAR	12223814	66320	302513	330465	97		3
7	6	BANKURA	BANKURA-I	KENJAKURA	16511350	293952	414775	401354	98		4
8	7	BANKURA	BANKURA-II	BIKNA	12648302	21848	634832.5	425094	97		4
9	8	BANKURA	BANKURA-II	JANBEDIA	10765190	0	613717	608448	95		6
10	9	BANKURA	BANKURA-II	KOSTHIA	7739800	75460	160816	146106	98		1
11	10	BANKURA	BANKURA-II	MANKANALI	8205159	47352	409505	350748.25	96		3
12	11	BANKURA	BANKURA-II	NARRAH	29104688	2057638	2277227	1535289.5	95		1
13	12	BANKURA	BANKURA-II	PURANDARPUR	6556359	113840	695991	441861	94		4
14	13	BANKURA	BANKURA-II	SANBANDHA	14453314	1091963	506291	415223	97		4
15	14	BANKURA	BARJORA	BARJORA	24742612	2199723	5539222	5068781	84		5
16	15	BANKURA	BARJORA	BELIATOR	19431669	0	381330	484610	98		4
17	16	BANKURA	BARJORA	BRINDABANPUR	11920559	1228458	442699	577721	96		5
18	17	BANKURA	BARJORA	CHHANDAR	13850035	1165834	385249	392733	97		3
19	18	BANKURA	BARJORA	GHUTGORIA	11252900	2182963	911212.5	853047	94		8
20	19	BANKURA	BARJORA	GODARDIHI	11361353	1515346.5	238704	270588	98		2
21	20	BANKURA	BARJORA	HATASURIA	13821445	1764779	565042	481778	97		4
22	21	BANKURA	BARJORA	KHANRARI	12147226	960312	329670.71	420484.71	97		4
23	22	BANKURA	BARJORA	MALIYARA	19591473	281706	730280	855732	96		8
24	23	BANKURA	BARJORA	POKHONNA	18307511	1846824	496139.41	405992	98		4
25	24	BANKURA	BARJORA	SHAHARJORA	12285519	1776227	351788	374794	97		3
26	25	BANKURA	BISHNUPUR	AYODHYA	13543705	1644534	1148177	1150356	93		1
27	26	BANKURA	BISHNUPUR	BANKADAHA	11780608	535672	414368.86	387901	97		3
28	27	BANKURA	BISHNUPUR	BELSULIA	17330399	2700298	224789	268681	99		2
29	28	BANKURA	BISHNUPUR	BHARA	15864846	2151471	337344	293172	98		2
30	29	BANKURA	BISHNUPUR	GOSSAINPUR	18312335	448728	842162	684811	97		6
31	30	BANKURA	BISHNUPUR	LAYEKBANDH	17184193	431162	220349	223941	99		2
32	31	BANKURA	BISHNUPUR	MORAR	12519760	3349131	5573427	5743163	73		5

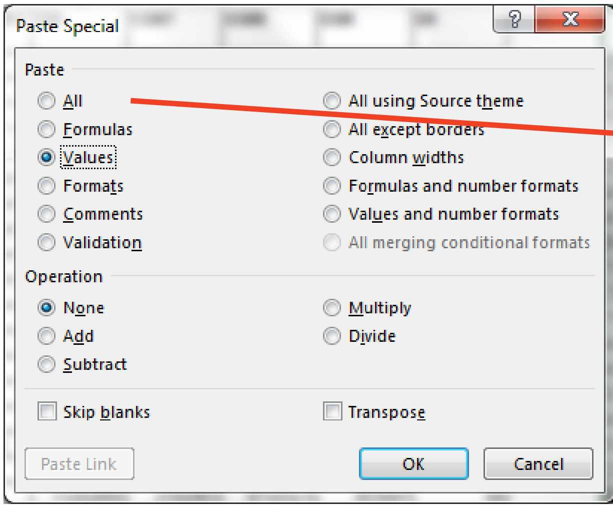
Drag and drop across the column to fill every cell.

Step 4



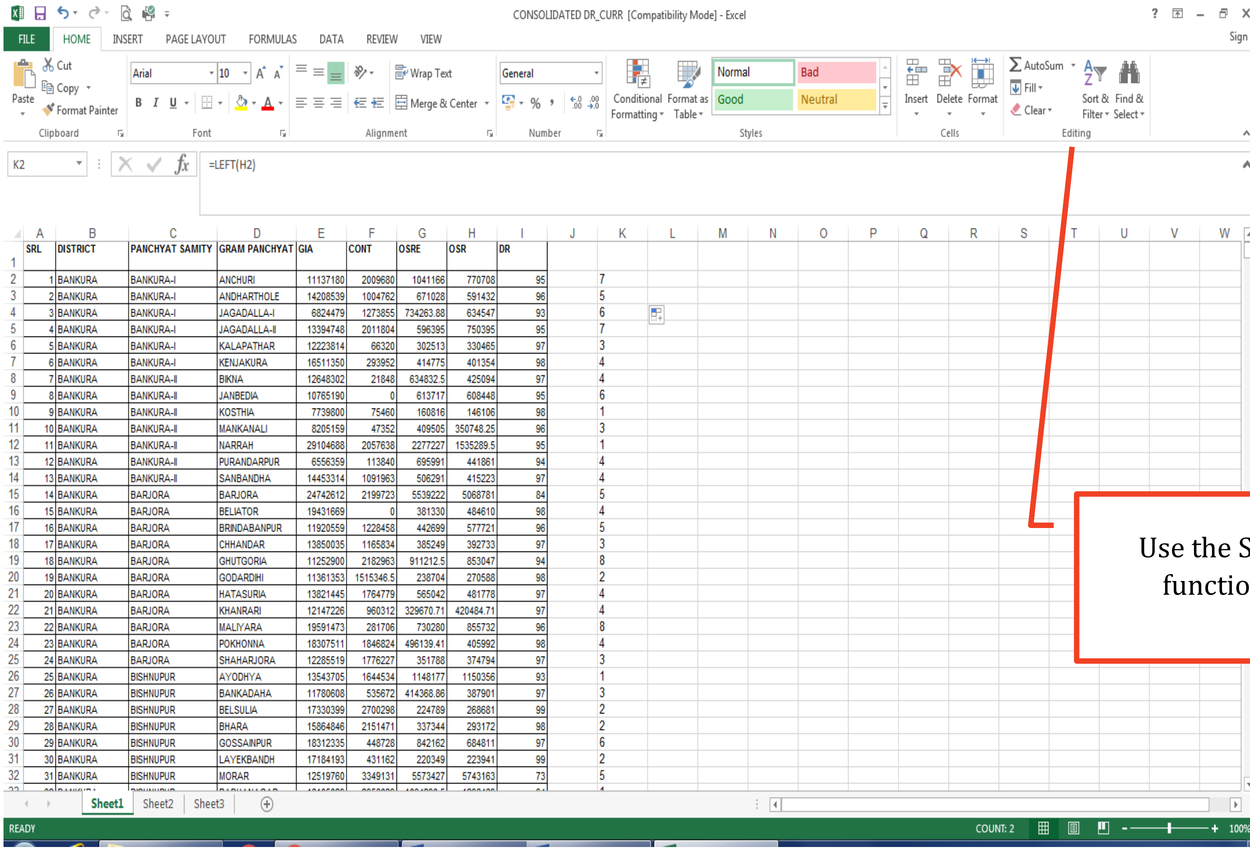
SRL	DISTRICT	PANCHYAT SAMITY	GRAM PANCHYAT	GIA	CONT	OSRE	OSR	DR	J	K
1	BANKURA	BANKURA-I	ANCHURI	11137180	2009680	1041166	770708	95		7
2	BANKURA	BANKURA-I	ANDHARTHOLE	14208539	1004762	671028	591432	96		5
3	BANKURA	BANKURA-I	JAGADALLA-I	6824479	1273855	734263.88	634547	93		6
4	BANKURA	BANKURA-I	JAGADALLA-II	13394748	2011804	596395	750395	95		7
5	BANKURA	BANKURA-I	KALAPATHAR	12223814	66320	302513	330465	97		3
6	BANKURA	BANKURA-I	KENJAKURA	16511350	293952	414775	401354	98		4
7	BANKURA	BANKURA-II	BIKNA	12648302	21848	634832.5	425094	97		4
8	BANKURA	BANKURA-II	JANBEDIA	10765190	0	613717	608448	95		6
9	BANKURA	BANKURA-II	KOSTHIA	7739800	75460	160816	146106	98		1
10	BANKURA	BANKURA-II	MANKANALI	8205159	47352	409505	350748.25	96		3
11	BANKURA	BANKURA-II	NARRAH	29104688	2057638	2277227	1535289.5	95		1
12	BANKURA	BANKURA-II	PURANDARPUR	6556359	113840	695991	441861	94		4
13	BANKURA	BANKURA-II	SANBANDHA	14453314	1091963	506291	415223	97		4
14	BANKURA	BARJORA	BARJORA	24742612	2199723	5539222	5068781	84		5
15	BANKURA	BARJORA	BELIATOR	19431669	0	381330	484610	98		4
16	BANKURA	BARJORA	BRINDABANPUR	11920559	1228458	442699	577721	96		5
17	BANKURA	BARJORA	CHHANDAR	13850035	1165834	385249	392733	97		3
18	BANKURA	BARJORA	GHUTGORIA	11252900	2182963	911212.5	853047	94		8
19	BANKURA	BARJORA	GODARDIHI	11361353	1515346.5	238704	270588	98		2
20	BANKURA	BARJORA	HATASURIA	13821445	1764779	565042	481778	97		4
21	BANKURA	BARJORA	KHANRARI	12147226	960312	329670.71	420484.71	97		4
22	BANKURA	BARJORA	MALIYARA	19591473	281706	730280	855732	96		8
23	BANKURA	BARJORA	POKHONNA	18307511	1846824	496139.41	405992	98		4
24	BANKURA	BARJORA	SHAHARJORA	12285519	1776227	351788	374794	97		3
25	BANKURA	BISHNUPUR	AYODHYA	13543705	1644534	1148177	1150356	93		1
26	BANKURA	BISHNUPUR	BANKADAHA	11780608	535672	414368.86	387901	97		3
27	BANKURA	BISHNUPUR	BELSULIA	17330399	2700298	224789	268681	99		2
28	BANKURA	BISHNUPUR	BHARA	15864846	2151471	337344	293172	98		2
29	BANKURA	BISHNUPUR	GOSSAINPUR	18312335	448728	842162	684811	97		6
30	BANKURA	BISHNUPUR	LAYEKBANDH	17184193	431162	220349	223941	99		2
31	BANKURA	BISHNUPUR	MORAR	12519760	3349131	5573427	5743163	73		5

Step 5



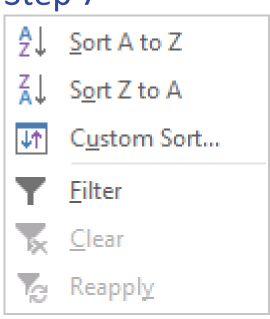
On the same column, paste special only the values. This will replace the entire column with value instead of existing function ie left(char, num).

Step 6



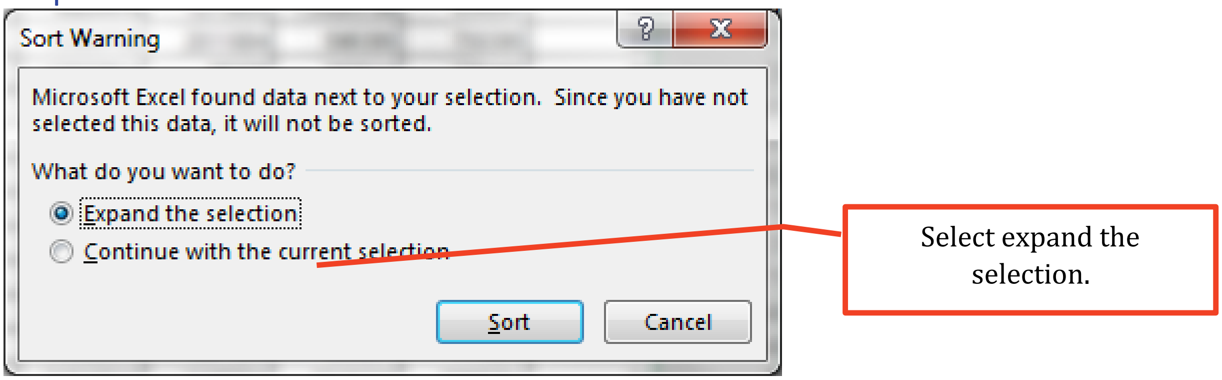
Use the Sort function

Step 7

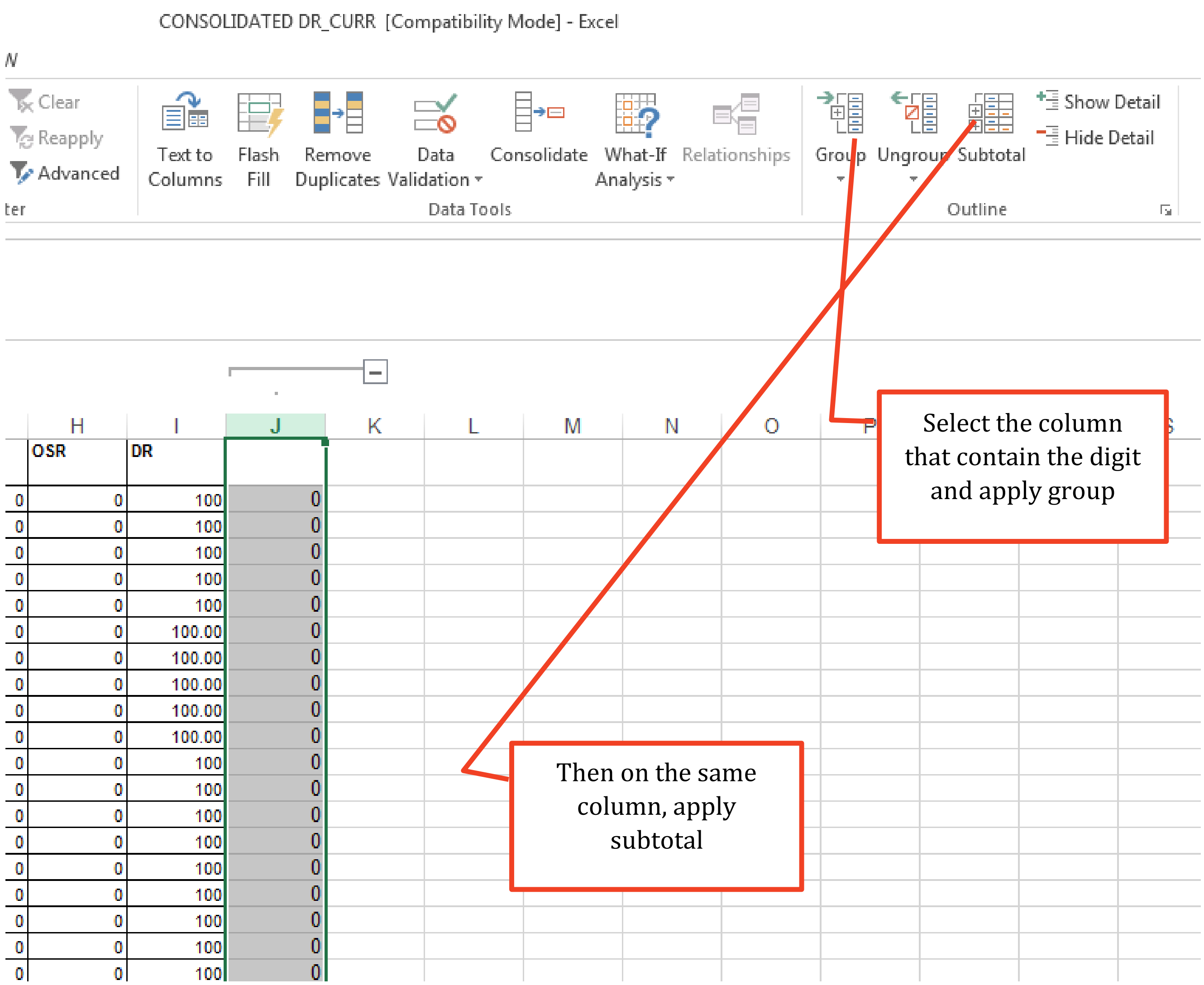


Now Sort from small to large

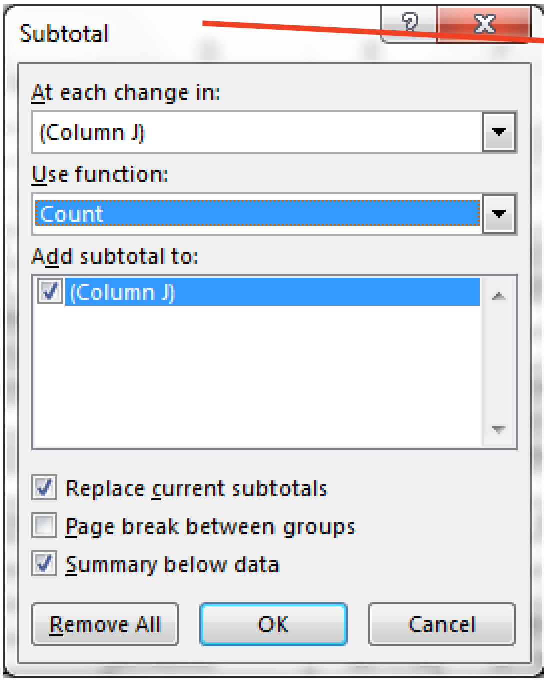
Step 8


Sort Warning
Microsoft Excel found data next to your selection. Since you have not selected this data, it will not be sorted.
What do you want to do?
Expand the selection
Continue with the current selection
Sort
Cancel
Select expand the selection.

Step 9

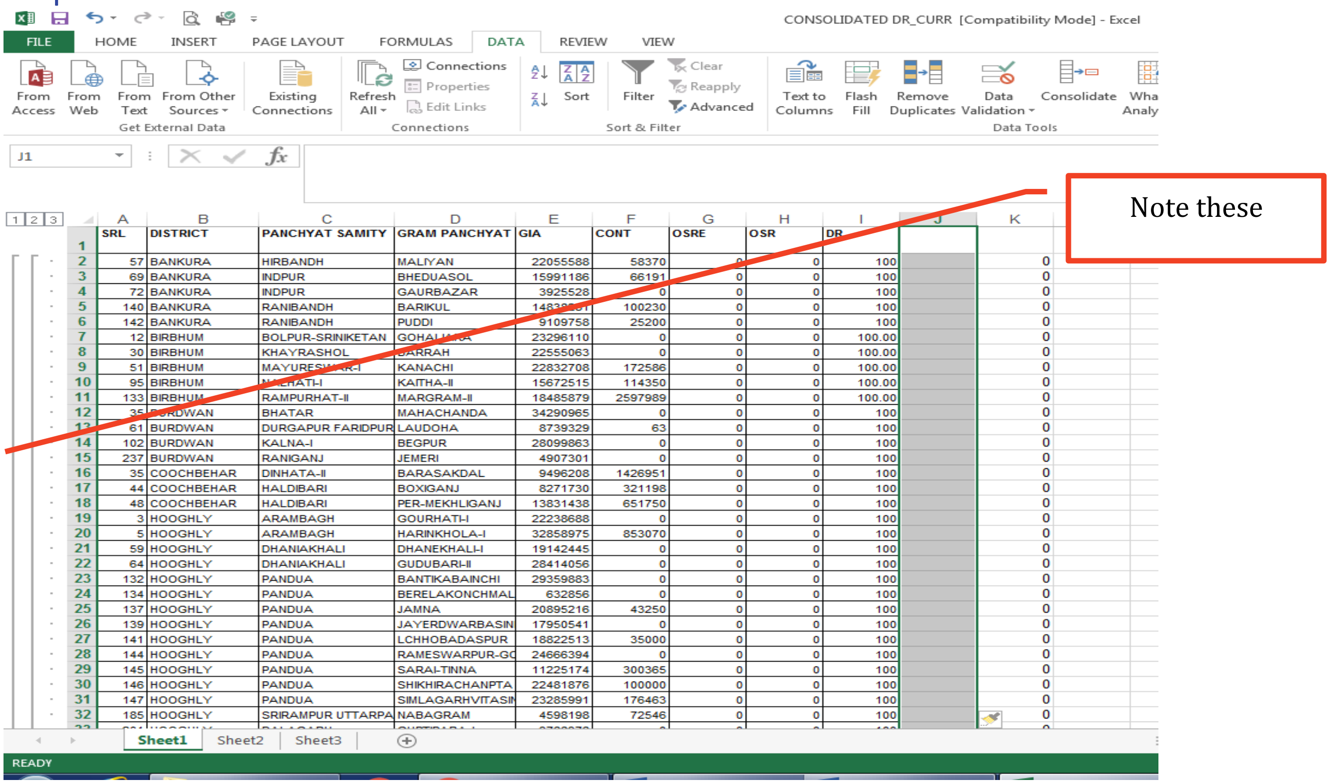

CONSOLIDATED DR_CURR [Compatibility Mode] - Excel
Clear
Reapply
Advanced
Text to Columns
Flash Fill
Remove Duplicates
Data Validation
Consolidate
What-If Analysis
Relationships
Data Tools
Group
Ungroup
Subtotal
Show Detail
Hide Detail
Outline
OSR
DR
Select the column that contain the digit and apply group
Then on the same column, apply subtotal

Step 10



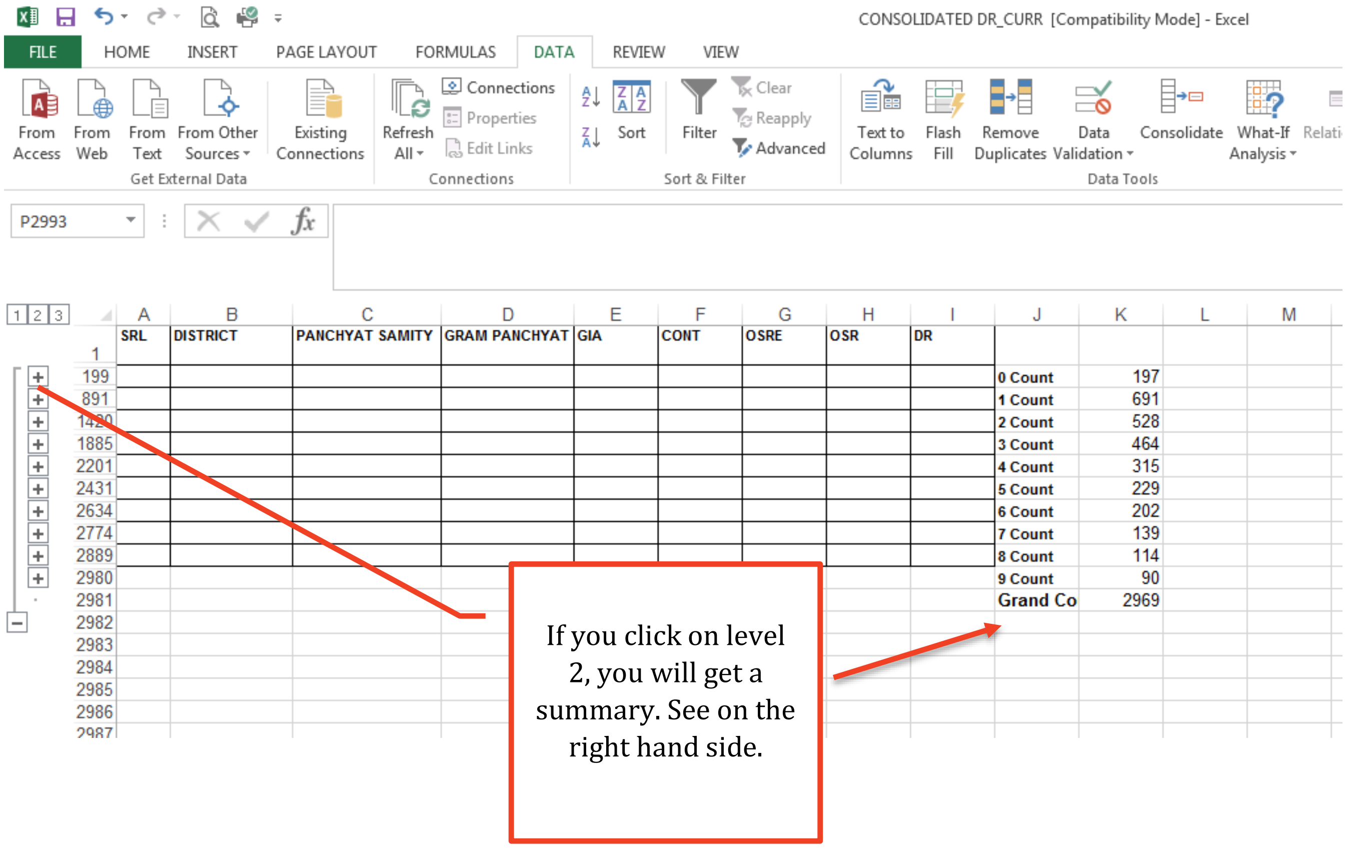
In the dialogue box select **COUNT** and don't change anything that comes in default. Press OK

Step 11

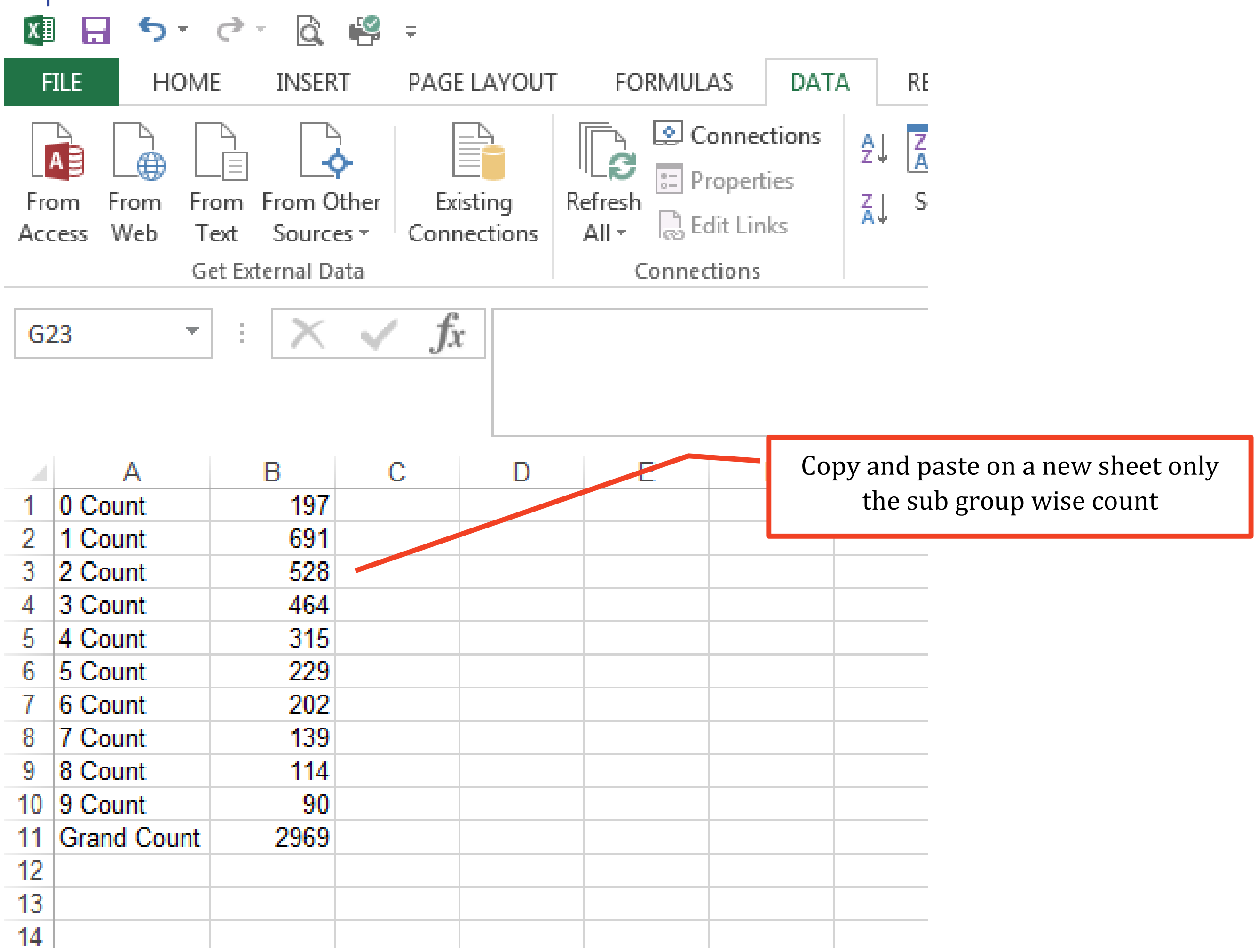


	SRL	DISTRICT	PANCHYAT SAMITY	GRAM PANCHYAT	GIA	CONT	OSRE	OSR	DR		
2	57	BANKURA	HIRBANDH	MALIYAN	22055588	58370	0	0	100		0
3	69	BANKURA	INDPUR	BHEDUASOL	15991186	66191	0	0	100		0
4	72	BANKURA	INDPUR	GAURBAZAR	3925528	0	0	0	100		0
5	140	BANKURA	RANIBANDH	BARIKUL	[illegible]	100230	0	0	100		0
6	142	BANKURA	RANIBANDH	PUDDI	9109758	25200	0	0	100		0
7	12	BIRBHUM	BOLPUR-SRINIKETAN	[illegible]	23296110	0	0	0	100.00		0
8	30	BIRBHUM	KHAYRASHOL	[illegible]	22555063	0	0	0	100.00		0
9	51	BIRBHUM	[illegible]	KANACHI	22832708	172586	0	0	100.00		0
10	95	BIRBHUM	[illegible]	KAITHA-II	15672515	114350	0	0	100.00		0
11	133	BIRBHUM	RAMPURHAT-II	MARGRAM-II	18485879	2597989	0	0	100.00		0
12	35	[illegible]	BHATAR	MAHACHANDA	34290965	0	0	0	100		0
13	61	BURDWAN	DURGAPUR FARIDPUR	LAUDOHA	8739329	63	0	0	100		0
14	102	BURDWAN	KALNA-I	BEGPUR	28099863	0	0	0	100		0
15	237	BURDWAN	RANIGANJ	JEMERI	4907301	0	0	0	100		0
16	35	COOCHBEHAR	DINHATA-II	BARASAKDAL	9496208	1426951	0	0	100		0
17	44	COOCHBEHAR	HALDIBARI	BOXIGANJ	8271730	321198	0	0	100		0
18	48	COOCHBEHAR	HALDIBARI	PER-MEKHLIGANJ	13831438	651750	0	0	100		0
19	3	HOOGHLY	ARAMBAGH	GOURHATI-I	22238688	0	0	0	100		0
20	5	HOOGHLY	ARAMBAGH	HARINKHOLA-I	32858975	853070	0	0	100		0
21	59	HOOGHLY	DHANIAKHALI	DHANEKHALI-I	19142445	0	0	0	100		0
22	64	HOOGHLY	DHANIAKHALI	GUDUBARI-II	28414056	0	0	0	100		0
23	132	HOOGHLY	PANDUA	BANTIKABAINCHI	29359883	0	0	0	100		0
24	134	HOOGHLY	PANDUA	BERELAKONCHMAL	632856	0	0	0	100		0
25	137	HOOGHLY	PANDUA	JAMNA	20895216	43250	0	0	100		0
26	139	HOOGHLY	PANDUA	JAYERDWARBASIN	17950541	0	0	0	100		0
27	141	HOOGHLY	PANDUA	LCHHOBADASPUR	18822513	35000	0	0	100		0
28	144	HOOGHLY	PANDUA	RAMESWARPUR-GO	24666394	0	0	0	100		0
29	145	HOOGHLY	PANDUA	SARAI-TINNA	11225174	300365	0	0	100		0
30	146	HOOGHLY	PANDUA	SHIKHIRACHANPTA	22481876	100000	0	0	100		0
31	147	HOOGHLY	PANDUA	SIMLAGARHVITASIN	23285991	176463	0	0	100		0
32	185	HOOGHLY	SRIRAMPUR UTTARPA	NABAGRAM	4598198	72546	0	0	100		0

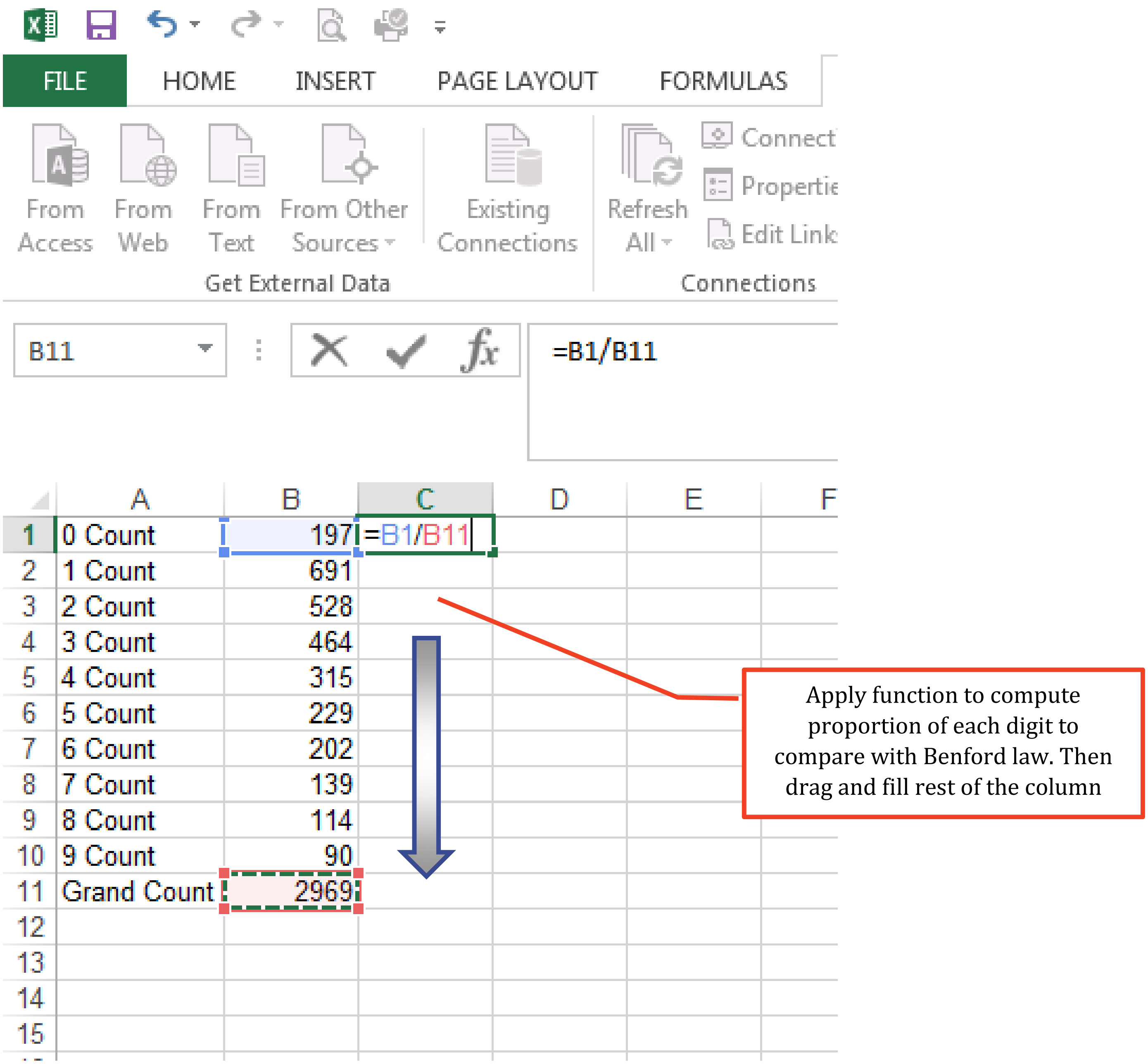
Step 12


CONSOLIDATED DR_CURR [Compatibility Mode] - Excel
FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW
From Access From Web From Text From Other Sources Existing Connections Refresh All Connections Properties Edit Links Sort Filter Clear Reapply Advanced Text to Columns Flash Fill Remove Duplicates Data Validation Consolidate What-If Analysis
Get External Data Connections Sort & Filter Data Tools
P2993
SRL DISTRICT PANCHYAT SAMITY GRAM PANCHYAT GIA CONT OSRE OSR DR
0 Count 197
1 Count 691
2 Count 528
3 Count 464
4 Count 315
5 Count 229
6 Count 202
7 Count 139
8 Count 114
9 Count 90
Grand Co 2969
If you click on level 2, you will get a summary. See on the right hand side.

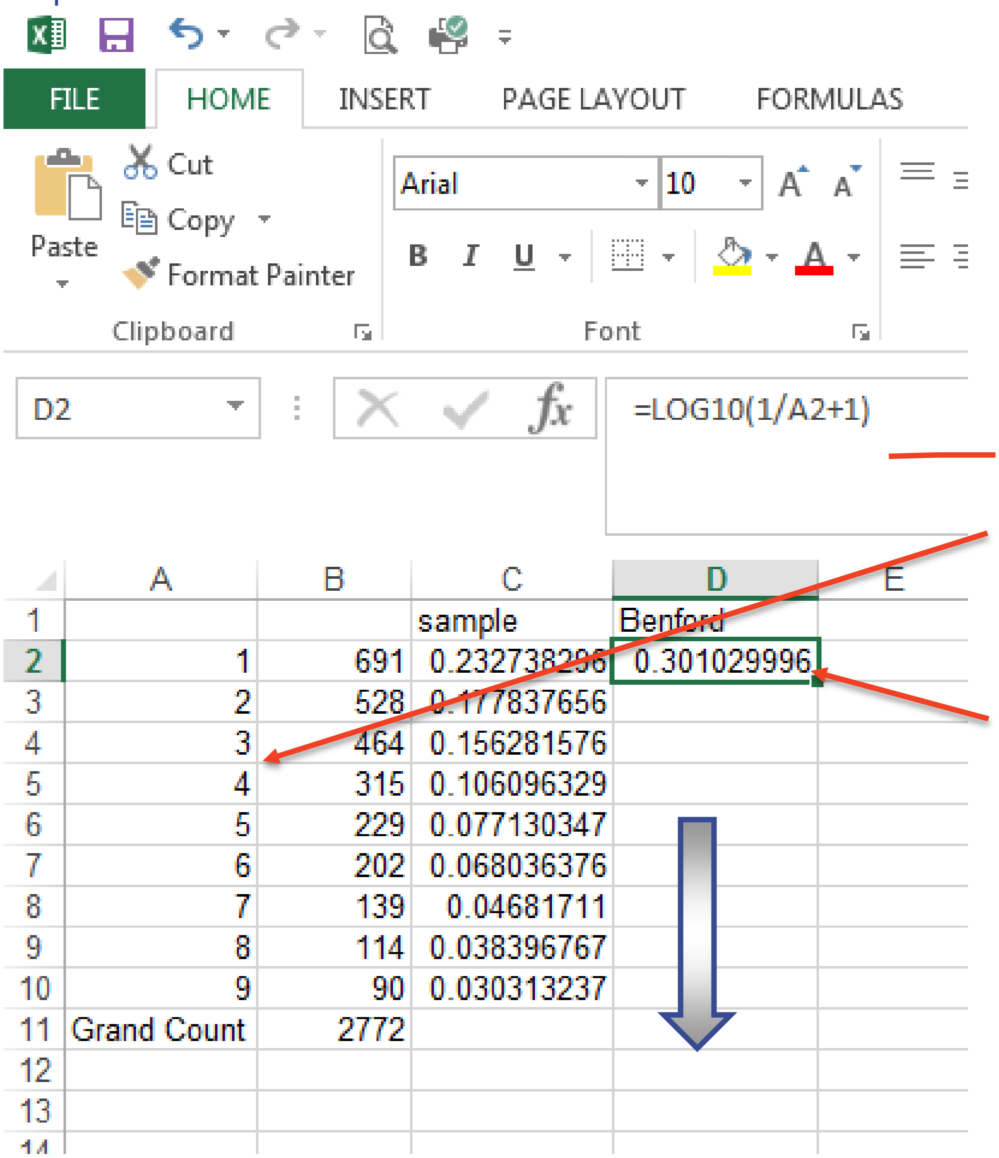
Step 13


FILE
HOME
INSERT
PAGE LAYOUT
FORMULAS
DATA
From Access
From Web
From Text
From Other Sources
Existing Connections
Get External Data
Refresh All
Connections
Properties
Edit Links
Connections
G23
A
B
C
D
E
0 Count 197
1 Count 691
2 Count 528
3 Count 464
4 Count 315
5 Count 229
6 Count 202
7 Count 139
8 Count 114
9 Count 90
Grand Count 2969
Copy and paste on a new sheet only the sub group wise count

Step 14


FILE
HOME
INSERT
PAGE LAYOUT
FORMULAS
From Access
From Web
From Text
From Other Sources
Existing Connections
Refresh All
Connect
Propertie
Edit Link
Get External Data
Connections
B11
=B1/B11
A
B
C
D
E
F
1 0 Count 197 =B1/B11
2 1 Count 691
3 2 Count 528
4 3 Count 464
5 4 Count 315
6 5 Count 229
7 6 Count 202
8 7 Count 139
9 8 Count 114
10 9 Count 90
11 Grand Count 2969
12
13
14
15
Apply function to compute proportion of each digit to compare with Benford law. Then drag and fill rest of the column

Step 15



	A	B	C	D
1			sample	Benford
2	1	691	0.232738296	0.301029996
3	2	528	0.177837656	
4	3	464	0.156281576	
5	4	315	0.106096329	
6	5	229	0.077130347	
7	6	202	0.068036376	
8	7	139	0.04681711	
9	8	114	0.038396767	
10	9	90	0.030313237	
11	Grand Count	2772		

Apply the Benford formula and drag and fill rest of the column.

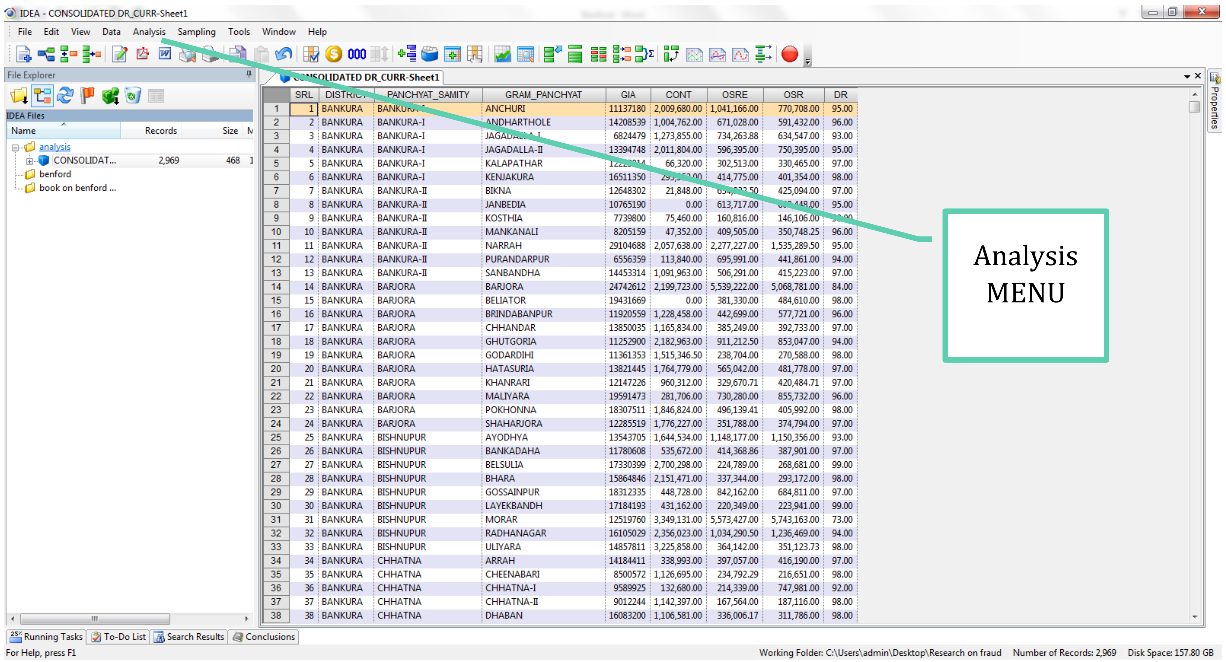
Also note some changes made in column A for easier calculation in the formula.

Further, the sample and Benford column have been identified.

Zero and the corresponding counts have been eliminated because in first digit zero has no value.

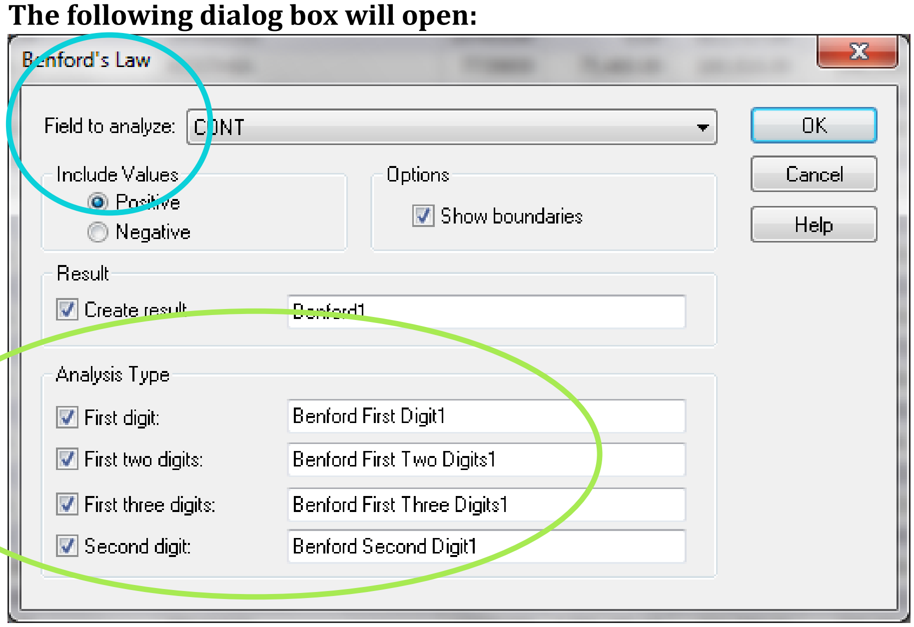
Using Benford’s Law: IDEA

1. Open the data table in IDEA

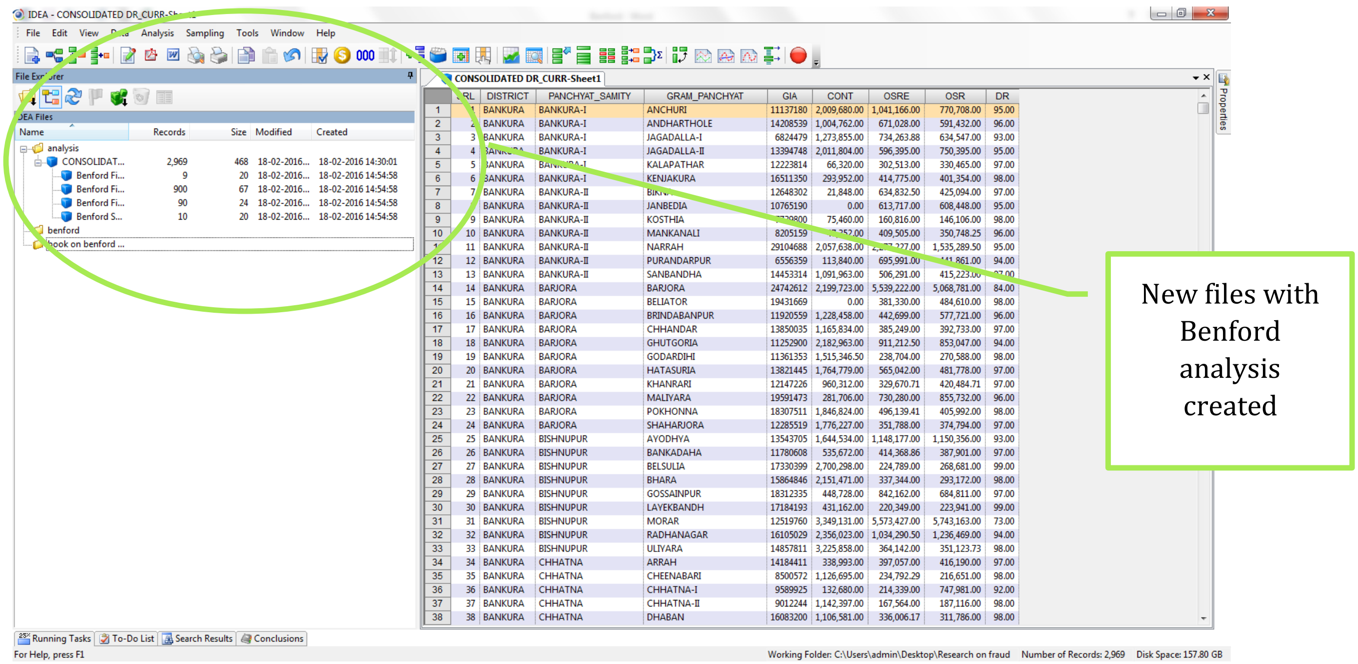


2. Go to menu ANALYSIS and then Benford Analysis

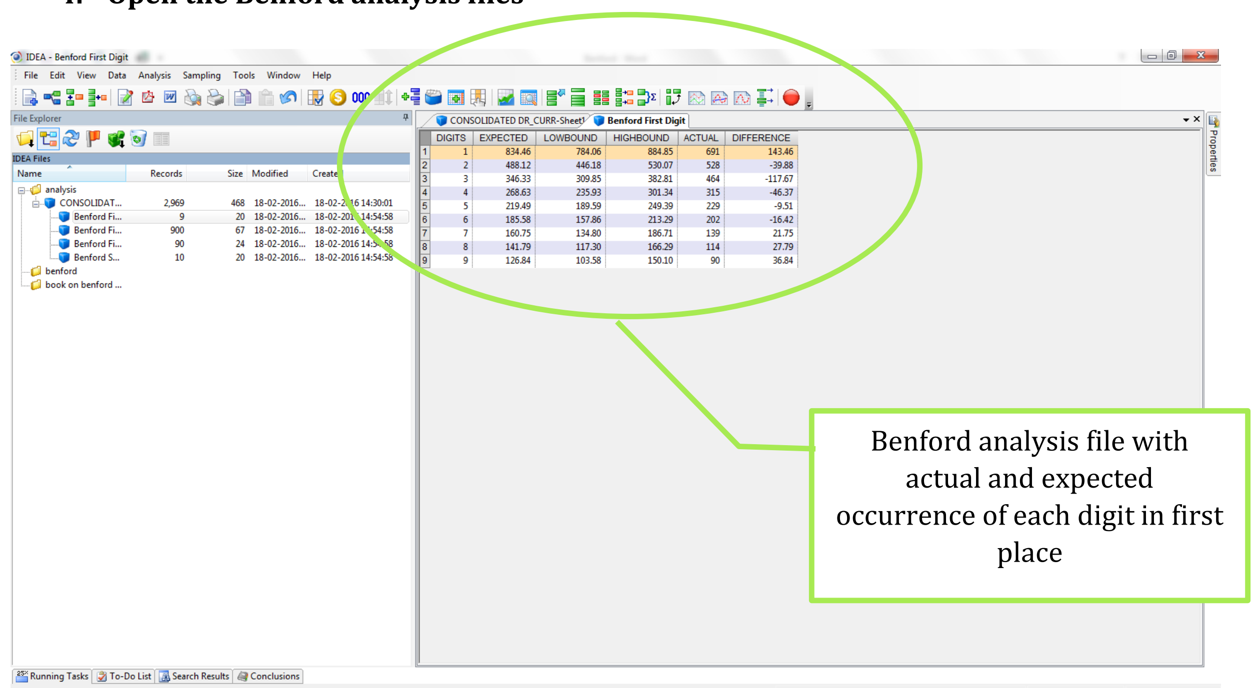
The following dialog box will open:



3. **Choose the FIELD TO ANALYSE, choose the ANALYSIS TYPE the click OK**
   The additional file with analysis will automatically open



4. **Open the Benford analysis files**



Limitations to application of Benford's Law

The limitations to application of Benford's law is twofold:

A. There are certain conditions that must be fulfilled for Benford law to be applicable. These are:
   a) The data must describe the same object. They could be different data set for say share prices in stock exchange, list of debts written off, volume of earth excavated, population of cities etc.
   b) There should not be any stipulation as to what could be the lower or upper limits. Although Benford's law tests can still be applied with suitable modification particularly if there are voluminous transactions in between and the upper and lower limit are far apart.
   c) The numbers should not be defined numbers for instance mobile phone numbers have a stipulation that they should begin with 9 or 8. Also, Benford law will not apply to voucher numbers are if there is a stipulation that a particular class of voucher should begin with a particular code eg. Sales vouchers will begin with 2, etc.

B. The second limitation of Benford's law is that while it may indicate that frauds may have occurred, it does not always pin point what is the nature of fraud or how they had occurred. On most occasion, Benford's law is most suitable for risk assessment prior to or during audit. However, reporting of possible fraud will require subsequent investigation, analysis, evidence collection and reporting.

Dataset used in this research

In this research paper we are looking into three different data set.

1. The first data set contains 2,969 records. It is an compilation of data from form 27 of Gram Panchayats Management System (GMPS) from which data pertaining to funds received as grants-in-aid, other funds received from State Government (mainly to meet administrative costs), the Gram panchayat's own sourced funds (reported by the gram panchayat) and expenditure from own sourced funds by the gram panchayat was collated into a database.
2. The second data set contains 1,14,245 records. It is from registrar of assurance and contains 28 fields that includes amongst others, data pertaining to valuation of property as reported in the sale deeds and a valuation of the property determined by the Government based on its rules.
3. The third and final data set contains 15,46,425 records. This data is from returns filed with commercial tax directorate by registered dealers. The data contains 12 fields and contains sales volume reported by the registered dealers against each type of commodity.

Analysis of the Data (using IDEA)

A BRIEF NOTE

In IDEA, the expected value is valid for the range, which is calculated on the basis of a pre-determined confidence level (for example, 95 %). This range is presented as an upper and lower bound. In case the actual frequencies are above the upper bound or below the lower bound, the individual actual frequency deviates in statistical terms significantly from the expected frequency. This is an indication of non-compliance of the data-points with Benford's law.

Appendix 2 *contains a brief discussion on various analysis of Benford's law.*

1. First Data set

i. Analysis of "Own Revenue Reported by the Gram Panchayats"

TABLE 1

DIGITS	EXPECTED	HIGHBOUND	ACTUAL	DIFFERENCE
1	834.46	884.85	691	143.46
2	488.12	530.07	528	-39.88
3	346.33	382.81	464	-117.67
4	268.63	301.34	315	-46.37
5	219.49	249.39	229	-9.51
6	185.58	213.29	202	-16.42
7	160.75	186.71	139	21.75
8	141.79	166.29	114	27.79
9	126.84	150.10	90	36.84

The **First digit analysis** shows significant deviation from expected values for digits 1 and 3 as is evident from ***Table 1***. This was followed up with **first two digit analysis.**

An extract of the second digit analysis of the data (***Table 2)*** revealed that 21 digits had repeated themselves in the data for more than expected number of time (entire 90 two digit is not included in the ***Table 2***). Additionally, the first two digit tests show that out of 21 digits that deviated most from their expected frequency of occurrence as per

TABLE 2

DIGITS	EXPECTED	LOWBOUND	HIGHBOUND	ACTUAL	DIFFERENCE
10	114.74	92.54	136.94	87	27.74
32	37.04	24.01	50.08	64	-26.96
12	96.36	75.90	116.82	76	20.36
11	104.75	83.48	126.02	85	19.75
27	43.78	29.68	57.88	63	-19.22
16	72.98	55.03	90.94	56	16.98
14	83.06	63.98	102.14	68	15.06
21	56.00	40.16	71.85	70	-14.00
18	65.09	48.08	82.10	52	13.09
25	47.22	32.60	61.83	60	-12.78
30	39.47	26.05	52.90	52	-12.53
39	30.48	18.59	42.37	43	-12.52
13	89.22	69.48	108.95	77	12.22
67	17.84	8.59	27.08	30	-12.16
19	61.75	45.16	78.34	50	11.75
38	31.27	19.24	43.30	43	-11.73
46	25.89	14.88	36.90	37	-11.11
33	35.94	23.09	48.79	47	-11.06
37	32.10	19.92	44.29	43	-10.90
31	38.22	25.00	51.45	49	-10.78
35	33.91	21.41	46.41	44	-10.09

Benford's law, eight digits began with one (38%) and eight digits began with three (38%) which is in congruence with the first digit test above.

TABLE 3					
DIGITS	EXPECTED	LOWBOUND	HIGHBOUND	ACTUAL	DIFFERENCE
1	831.75	781.51	881.98	701	130.75
2	486.54	444.73	528.35	533	-46.46
3	345.21	308.84	381.57	418	-72.79
4	267.76	235.17	300.36	332	-64.24
5	218.78	188.98	248.58	239	-20.22
6	184.97	157.35	212.6	178	6.97
7	160.23	134.36	186.1	147	13.23
8	141.33	116.92	165.75	124	17.33
9	126.43	103.24	149.61	91	35.43

ii. *Analysis of "Expenditure out of Own Revenue Reported by the Gram Panchayats"*

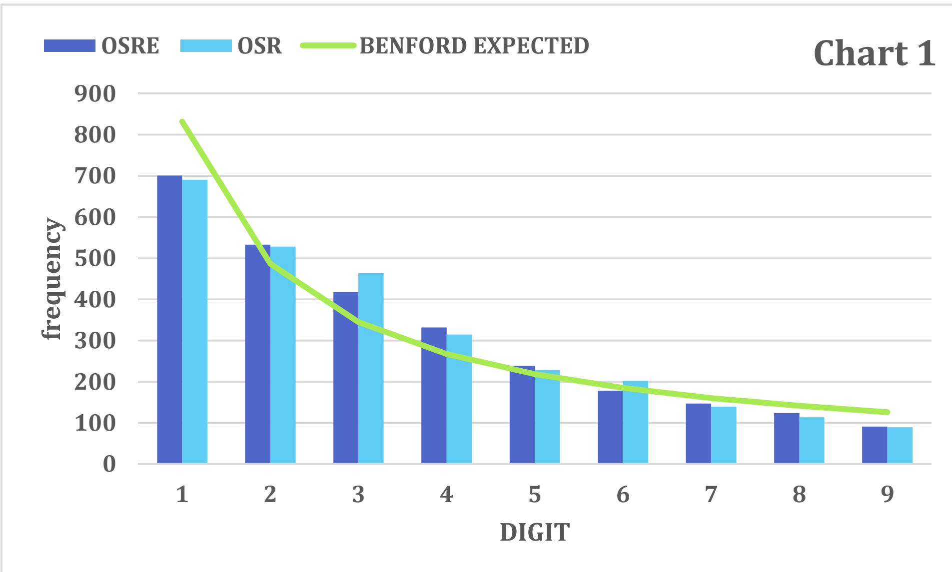
An analysis of the conformity of Benford's law to ***first digits*** is given at ***Table 3*** below. Further, ***two digit test*** for compliance with Benford's law in ***Table 4*** revealed that out of 16 digits that showed most deviation from the Benford's law, six digits have one as their first digit.

Possible Explanation for the Deviations from Benford Law

TABLE 4					
DIGITS	EXPECTED	LOWBOUND	HIGHBOUND	ACTUAL	DIFFERENCE
27	43.64	29.58	57.7	72	-28.36
13	88.93	69.26	108.59	61	27.93
10	114.37	92.24	136.5	87	27.37
17	68.59	51.18	85.99	48	20.59
35	33.8	21.34	46.26	50	-16.2
12	96.05	75.66	116.44	81	15.05
22	53.34	37.88	68.8	67	-13.66
38	31.17	19.18	43.16	44	-12.83
40	29.63	17.92	41.34	42	-12.37
15	77.44	59.01	95.87	66	11.44
11	104.41	83.2	125.62	93	11.41
34	34.78	22.16	47.41	46	-11.22
41	28.92	17.34	40.49	40	-11.08
65	18.32	8.98	27.66	29	-10.68

One of the possible explanation for high numbers (i.e. above expectation) of lower digit could be because the panchayats are reluctant to report higher revenue from own source out of fear from losing out on Government support. It is also possible that own-source revenues which constitute local taxes are not properly collected or collected but remain out of books.

The ***chart 1*** shows that deviations from the Benford's Law was in similar direction for both own source revenue and own source expenditure for the years.

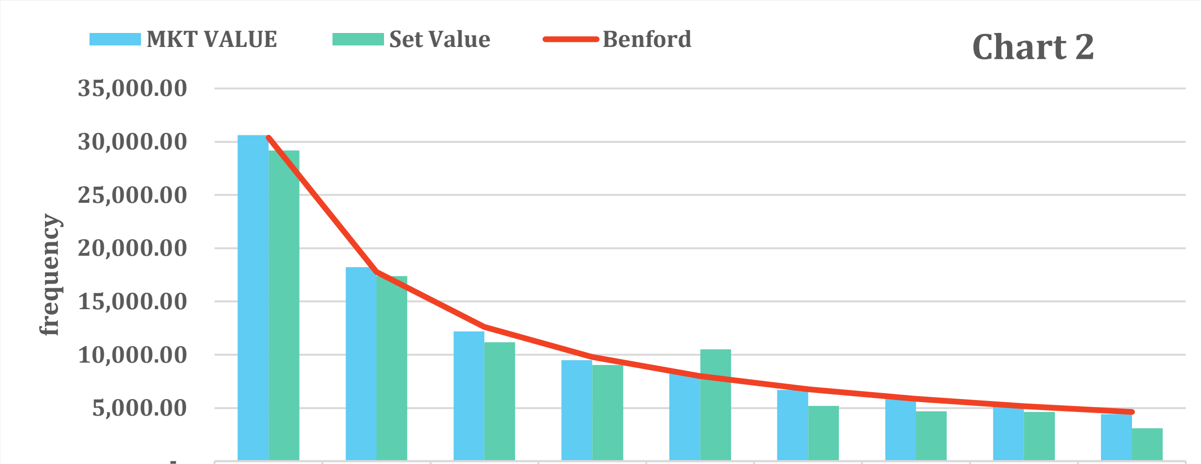


Although examining the records in field was not possible for the purpose of this research, an additional analysis of the database threw up an interesting fact. Out of 2,969 records, 1,566 records showed that the Gram Panchayats have reported that they had spent more money from their own sources compared to the sums that they have collected as their own sourced revenue for the year. These 1,566 Gram Panchayats have aggregated ₹ 105.26 crore as their expenditure from own sourced revenue while earning ₹ 82.97 crore as their own sourced revenue.

2. Second Data set

The Second set of data analysed consists of two series as is shown in ***Chart 2*** (*please see in next page*). The first is the market value (**MKT VALUE)** of a property that was determined by the Government as per rules for the purpose of determining stamp duty receivable by the Government. It is based on wide variety of factors including nature of the property, its intended use and location of the property in the State. This data set was found to be in conformity with the Benford law.

However, the values declared by the property owners (**Set Value)** as the value of the transaction



	1	2	3	4	5	6	7	8	9
MKT VALUE	30,610	18,221	12,192	9,501.	8,254.	6,693.	5,821.	5,198.	4,436.
Set Value	29,168	17,386	11,178	9,048.	10,511	5,214.	4,699.	4,639.	3,111.
Benford	30,384.	17,773.	12,610.	9,781.6	7,992.1	6,757.2	5,853.4	5,163.0	4,618.5

for transfer of property did not conform to Benford's Law.

Possible Explanation for the Deviations from Benford Law

The stamp duty is collected by the Government based on the higher of the values between the transaction value declared by the owner of the property or the market value as determined by the Government. Government determines the value of the transaction by examining the transfer of property deed where purchase consideration is mentioned. Property owners are naturally inclined to reduce the stamp duty payable to the Government. This also means that with the intention of not being asked to pay duty beyond what is minimum to be paid, the limiting factor of the declared value of transaction is the market value determined by the Government. This fact is precisely the case.

An analysis of the dataset [2] revealed that out of 1,01,806 records containing non-null market value/transaction value, 81,622 records (80.17%) had declared transaction value less

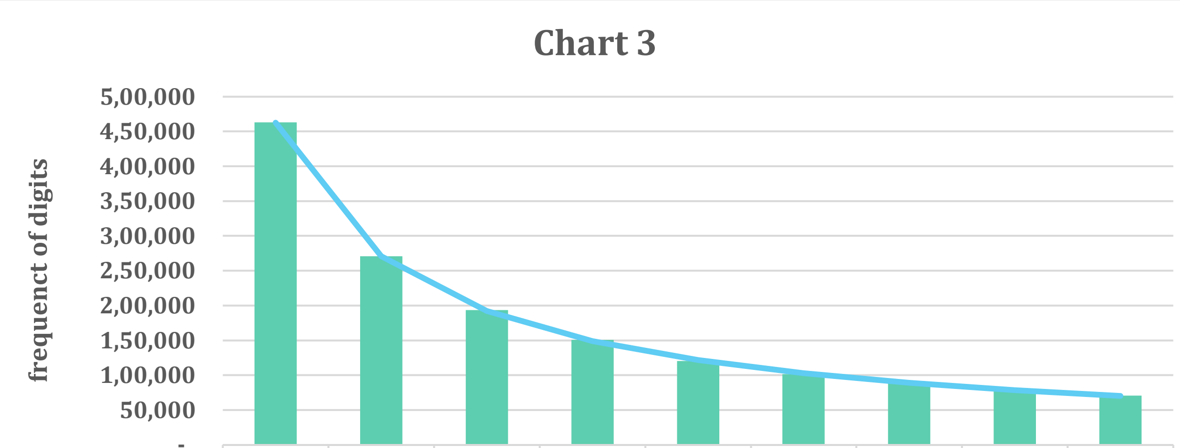
[2] Out of 114,245 cases recorded in the dataset, 12,439 cases recorded zero as both the market value and the transaction value and are therefore excluded from the Benford's law analysis.

than the market value and 19,984 records (19.63%) had transaction value exactly equal to market value determined by the Government. Only 200 deeds were registered where the transaction value was declared as more than the market value determined by the Government.

Thus, we may conclude that because the transaction values declared were intentionally manipulated to ensure that transaction values declared do not exceed the market value determined by the Government for that property, the transaction values did not conform to Benford's Law.

3. Third Data set

As already explained above, the third data set comprises of declared value of aggregate transactions against each type of commodity filed in their return by registered dealers to the directorate of commercial taxes. The Benford law compliance of this dataset is plotted into ***chart 3*** below.



	1	2	3	4	5	6	7	8	9
ACTUAL	4,62,8	2,70,9	1,93,4	1,50,6	1,20,1	1,01,6	88,660	78,022	70,670
EXPECTED	4,62,74	2,70,68	1,92,05	1,48,96	1,21,71	1,02,91	89,145	78,631	70,338

It is evident from ***chart 3*** that this data set is in conformity with Benford's law. This may indicate that the dealers are filing returns that reflect the facts of the transactions and there is less manipulation of data for volume of sale. This could be attributed to the record trail that the dealer has to maintain to obtain necessary credits for input tax; and are therefore difficult to manipulate.

Conclusion

In our analysis, we had utilized three different data set that are created by different Government agencies and are utilized for very different purpose. It is evident that when data are reported as they naturally occur, conformity with Benford's law could be expected. When data is adjusted to drive home a specific utility, such data do not conform to Benford's law.

However, while Benford's law can indicate that data filed might have been managed with a particular objective, it is seldom sufficient to explain a cause-effect relationship. Such explanations will continue to require a vigilant auditor and his skills of scrutiny and evidence collection. In this regard, Benford's Law compliance tests are an important arsenal in auditor's hand but it is not the end in itself.

Further Reading

1. I've Got Your Number: How a mathematical phenomenon can help CPAs uncover fraud and other irregularities. BY MARK J. NIGRINI MAY 1999 Journal of Accountancy

2. Forensic Analytics - Methods and Techniques for Forensic Accounting Investigations. MARK J. NIGRINI, B.COM.(HONS), MBA, PH.D. John Wiley & Sons, Inc.

3. The First Digit Phenomenon - A century-old observation about an unexpected pattern in many numerical tables applies to the stock market, census statistics and accounting data. Author: Theodore P Hill. American Scientist Volume B6

4. The Use of Benford's Law as an aid in Analytical Procedures: Mark J. Nigrini and Linda J Mittermaier. Auditing a Journal of Practice and Theory Volume 16 no. 2 Fall 1997

5. Fraud and Forensic Accounting in a Digital Environment: White Paper for The Institute for Fraud Prevention. Conan C. Albrecht, Marriott School of Management, Brigham Young University.

6. Cheating behavior and the Benford's Law. Professor Dominique Geyer, Nantes Graduate School of Management, France

7. A Statistical Derivation of the Significant Digit Law. Theodore P Hill. Statistical Science, Volume 10 No, 4, 1995

8. Fraud risk management: a guide to good practice CIMA, 2009

9. Annual Report to the Nation on Occupational Fraud and Abuse from Association of Certified Fraud Examiners

10. Using Digital Analysis to Enhance Data Integrity. Michael L. Ettredge Rajendra P. Srivastava University of Kansas, June, 1998

11. The effective use of Benford's Law to assist in detecting Fraud in Accounting Data, Cindy Durtschi, William Hillison and Carl Pacini. Journal of Forensic Accounting, Vol. 5, 2004

12. A Universal Mathematical Law Criterion for Algorithmic Validity. Dickinson, John R. University of Windsor. Developments in Business Simulation and Experiential Learning, Volume 29, 2002

Appendix 1: Formula and Expected frequency.

(Source: Cindy Durtschi, William Hillison and Carl Pacini, 2004)

The formulas for the frequencies are shown below with D_1 representing the first digit and D_2 the second digit of a number. A two digit number is therefore written as D_1D_2.

USING BASE 10 LOGARITHMS THE FORMULAS ARE:

First digit of number	$P(D_1=d_1)= \log\{1+(1/d_1)\}$ where $d_1 \in \{1,2,3,....9\}$
Second digit of number	$P(D_2=d_2)= \sum_{d_1=1}^{9} \log\{1+(1/d_1d_2)\}$ where $d_1 =1$ and $d_2 \in \{1,2,3,....9,0\}$
For two digit combination	$P(D_1D_2=d_1d_2)= \log\{1+(1/d_1d_2)\}$ $P(D_2 = d_2 \mid D_1 = d_1)= \log\{1+(1/d_1d_2)\} / \log\{1+(1/d_1)\}$ Where D_1 represents the first digit of a number, D_2 represents the second digit of a number, etc.

Expected Frequencies Based on Benford's Law

Digit	First Place	Second Place	Third Place	Fourth Place
0		0.11968	0.10178	0.10018
1	0.30103	0.11389	0.10138	0.10014
2	0.17609	0.19882	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

Source: Nigrini, 1996

Appendix 2: Brief note on various Benford's Law Analysis and their implication.

FIRST DIGIT AND SECOND DIGIT ANALYSES

In these first tests, the individual first or second digit of numbers .n a data series will be analyzed. As a result of the analysis, the frequency of the digits 1-9 (when the first digit) or 0-9 (when the second digit) is presented in graphic and table form Thus, it is a comparison of the reference value and the actual value to evaluate the plausibility of the underlying data material (actual value) according to the expected distribution (reference value) expected by Benford.

The expected output serves as rough check of the actual numerical distribution in the population. Statistically significant deviations may be questioned. The justifications can result from value limits in the data (for example, maximum amounts of payment) or numeric systems (for example, circles of numbers) or individual reasons leading to the explainable increase in the frequency of certain digits.

FIRST TWO DIGITS ANALYSIS

This test examines the frequency of the numerical combinations 10 to 99 in the first two digits of a series of numbers. The test is presented in a graphic form that shows the expected frequencies according to Benford (reference value) and the actual frequencies of the analyzed data (actual value) on an abscissa divided into 10-99. The x axis includes the expected and actual frequencies per numerical combination. Numerical combinations, which occur with a frequency exceeding the confidence interval, are marked as anomalies. In addition to the presentation in graphic form the data, which the graphs are based on, are presented in table form.

In particular, the output serves for the analysis of avoided threshold values. Thus, these tests help to clearly visualize when order or permission limits have been systematically avoided

FIRST THREE DIGITS ANALYSIS

This test examines the frequency of the numerical combinations 100 to 999 in the first three digits of a series of numbers. The test is presented in a graphic form in which the expected frequencies of the analyzed data (actual value) are illustrated in an abscissa divided into 100 to 999. The x axis includes the expected and actual frequencies per numerical combination. Numerical combinations, which occur with frequencies exceeding the confidence interval, are marked as anomalies. In addition to the presentation in graphic form, the data, which the graphs are based on, are also presented in table form.

In particular, the output serves for analysis after conspicuous rounding off operations. In general, this analysis will include many deviations because in order to receive a comparative distribution with Benford, there must be a large amount of observation values. The reason is that at least 899 observation values are needed so that every numerical combination occurs at least once (100-999). Therefore, this analysis usually does not lead to a meaningful result until it is based on a population of over 10,000 observation values. It seems advantageous that the degree of exactitude is higher and the business events to be questioned per numerical combination tend to be lower in this test than in the others.

ROUNDED BY ANALYSIS

This test is used to analyze the relative increasing frequency of rounded numbers. The determination comprises the frequency of the numbers that are divisible by 10, 25, 100, and 1,000 (as well as any user-defined value of whole numbers) without remainders.

The empirically observed frequency of the analyses conducted by Nigrini is used as measure of the reference value. According to this, values that are divisible by 10 are expected in a range of 10% of the observation values and values divisible by 25, 100 and 1,000 in a range of 4%, 1% and 0.1% as reference value. Here it is important that the decimal places of a number are considered as well. If they were included, the number 100.50 would no longer be a multiple of 25. In the opposite case, the *places after the decimal separator are simply 'cut off', i.e., in our example the number 100.50 is a multiple of 25 as it is interpreted as a 100. Thus, values are treated like whole numbers.*

DUPLICATES ANALYSIS

The analysis of multiple duplicates includes all number values in the entire database that occur more than once. This test helps the user to recognize all existing duplicates in the data supply whereas the result table presents the duplicates sorted according to the descending frequency. The aim of the test is to identify certain numbers that occur more than once (for example, possible double payments). The difference from the other tests is that this test does not analyze any numerical combinations but the pure value of a number.

(Source: IDEA manual on Benford's Law Analysis)