Marquette University
e-Publications@Marquette

Accounting Faculty Research and Publications   Accounting, Department of

Using Benford's Law to Detect Fraud in the Insurance Industry

Meredith Maher
Marquette University

Michael D. Akers
Marquette University, michael.akers@marquette.edu

Using Benford's Law To Detect Fraud In The Insurance Industry

Meredith Maher, (E-mail: meredith.maher@marquette.edu), Marquette University
Michael Akers, (E-mail: michael.akers@marquette.edu), Marquette University

Abstract

Benford's Law is the mathematical phenomena that states that the first digits or left most digits in a list of numbers will occur with an expected logarithmic frequency. While this method has been used in industries such as oil and gas and manufacturing to identify fraudulent activity, it has not been applied to the health insurance industry. Since health insurance companies process a large number of claims each year and these claims are susceptible to fraud, the use of this method in this industry is appropriate. This paper examines the application of Benford's Law to four health insurance companies located in the Midwest. For each company, analysis was performed on the first digit distribution, the first two-digit distribution, and providers with high volumes of claims. The results show that the populations are similar to the frequencies predicted by Benford's Law. The findings also suggested possible fraudulent activity by specific providers, however, the companies determined that these results occurred due to abnormal billing practices and were not fraudulent. The insurance companies that participated in this study will continue to use this method to further detect fraudulent claims.

Introduction

he U.S. General Accounting Office has estimated that fraud accounts for up to 10% of the annual expenditure on health care or $100 billion in the United States (GAO 1997). With the improvements in technology over the last decade, it has become increasingly possible to analyze large amounts of data to detect trends in claims. Hence, the use of Benford's Law has now become feasible. In the health insurance industry, there is a large amount of claims data submitted by thousands of health care providers. Benford's Law can be used to detect abnormalities in the data. In this paper, data from four insurance companies was examined and compared with the expected results from Benford's Law.

The first section of the paper provides an overview of Benford's Law. Next, the applicability of Benford's Law and the relevance to the insurance industry is discussed. The third and fourth sections describe the research methodology and results. The last section provides concluding comments. Note: The company and provider names are not disclosed to protect the interests of these corporations.

Benford's Law

Benford's Law is the mathematical phenomena that states that the first digits or left most digits in a list of naturally occurring numbers will occur with an expected logarithmic frequency as illustrated in Table 1. The equation for the curve is given by P $(D_1=d_1) = \log 10(1+1/d_1)$ for d_1 in $\{1,2,3,4,5,6,7,8,9\}$ (Nigrini 1999).

GE physicist Frank Benford discovered an odd pattern that eventually became known as Benford's Law in the 1930s. He noted that the frequency of low numbers such as 1, 2, or 3 was much higher than numbers such as 7, 8, or 9 (Nigrini 1999). The Law applies to populations greater than 10,000 with an extremely high degree of accuracy. However, it can be applied to populations with fewer occurrences with less accuracy. Benford's Law

---

Readers with comments or questions area encouraged to contact the authors via email.

Table 1
Benford's Law Frequencies

First Digit	Frequency
1	0.30103
2	0.17609
3	0.12494
4	0.09691
5	0.07918
6	0.06695
7	0.05799
8	0.05115
9	0.04576

applies to a set of numbers regardless of their units (i.e., it is scale invariant). Thus, it will work on any natural population whether it is in dollars, yen, or square feet (Matthews 2000). Benford's Law assumes that numbers represent the relative sizes of similar objects, such as net incomes, trading volumes, or city populations.

Benford's Law does not apply to numbers that are influenced by human interaction, such as ATM withdrawals. The data is assumed not to have an arbitrary cutoff, which might eliminate some data values and make the data analysis invalid. Also, preassigned numbers such as phone numbers or personal identification numbers will not follow Benford's Law (Nigrini 1997).

By using z-statistics, one can determine if the dataset is within an acceptable boundary of reasonableness between expected and actual values (Nigrini 1997). The goodness of fit test is based on the formula Chi-Squared = Sum $((o_i - e_i)^2 / e_i)$ for each i from 1 to 9. The chi-squared variable is very close to the chi-squared distribution with 8 degrees of freedom. If the observed frequencies are close to the corresponding expected frequency, the chi-squared will be small, which indicates a good fit. When the chi-squared value is large, it indicates a poor fit with the expected distribution. As the population becomes large, the observed and expected frequencies are expected to become nearly the same.

Applicability and Relevance

As stated in Nigrini & Mittermaier (1997), with the cost of computer processing declining and the speed of processors improving, it is now feasible to use data analysis tools, such as Benford's Law, in analytical audit procedures (Nigrini 1997). When utilizing analytical procedures in auditing, the results indicate potential areas of overstatement or understatement in account balances or system data (Nigrini 1997). These potential discrepancies indicate areas where an auditor should perform further testing to ensure that fraud or misstatements have not occurred. Prior research has used Benford's Law to evaluate the authenticity of data, while conformity does imply authenticity; nonconformity raises the level of suspicion.

The theory has been applied to business data populations including sales orders, cancelled checks, inventory purchases, and disbursement records (Nigrini 1999). Data that does not conform to Benford's Law raises the level of suspicion that errors are present in the dataset. Christian and Gupta (1993) used taxpayer data to examine tax evasion for cutoff points on the tax tables by applying Benford's Law for distribution of 5th and 6th digits as a basis for its findings. In 1996, Nigrini used tax information to understand possible tax fraud. Nigrini separated taxpayers into two classes: high and low likelihood of tax evasion. Using the figures from the taxpayers 1040 and supporting forms, Nigrini found that the low group had a better conformity to Benford's Law than did the high likelihood group (Nigrini 1997).

In 1994, Nigrini applied Benford's Law to a dataset of payroll information over 10 years. Nigrini found that the data followed the expected pattern over the first five years, but deviated from the expected values for the second five years. This was caused by a repetitive fraud committed by the payroll clerk. Crowder (1997) noted that emerging trends such as Benford's Law would aid in the detection of fake checks or payroll fraud in the general manufacturing industry. Coderre and Warner (1999) used Benford's Law to determine if the frequency of checks signed just below the threshold for a second signature indicated possible fraud. Their study noted an unusual frequency of numbers below the signing levels, which did not conform to the predicted results.

The use of these data analysis tests has a practical application in the field of auditing large sets of data. All datasets have some variability to them; however, the dataset must closely conform to the requirements of Benford's Law in order to be an applicable test (Nigrini 1997). The use of the first two digits makes the auditability of large

datasets more feasible, because it reduces the population into smaller, more concentrated areas of risk that are reasonable to audit. Benford's Law will also further indicate abnormal patterns in number occurrences (Coderre and Warner 1999). The continuous auditing of datasets using Benford's Law will immediately notify management that errors or fraud maybe occurring (Kogan, Sudit, and Vasarhelyi 1999).

Currently, the insurance industry is concerned with the concept of fraudulent claims due to the estimated $100 billion loss each year due to fraud (GAO 1997). Detective controls, such as significant investments in departments to investigate potential frauds and fraud hotlines, are used to uncover fraud. Health insurance companies also perform random audits of claims to ensure data accuracy and validity.

An analysis of actual results as compared to Benford's Law expectations will demonstrate that the claims received as a whole are not likely to be fictitious. By applying Benford's Law to each company's claims, trends can be detected that may indicate possible errors or frauds. Insurance companies can isolate high-volume providers that maybe submitting either fraudulent or errant claims using Benford's Law. Thus, corporate resources can be more effectively used to research those suspect companies, instead of randomly auditing from a pool of all providers.

Research Methodology

Data was extracted from four insurance companies for all health insurance claims received over a 10-month period between 1/1/2001 and 10/31/2001. The four insurance companies are nationally based and headquartered in the Midwest. Their membership ranges from approximately 100,000 to 1,000,000 member lives. The health insurance underwritten by these companies is a mixture of individual medical and group health insurance.

The data obtained from the corporations was examined and credits received from providers were removed, because the test is based on charges received from providers, not credits issued by providers for past billing errors. Charges of less than one dollar were also removed from the population for the first digit analysis. For the first two-digit (FTD) Benford's Law analysis, charges less than ten dollars were removed from the population. This was necessary to eliminate amounts without a single or double digit, respectively. The resulting number of claims analyzed for the single digit analysis was 11,622,602. The total number of claims analyzed for the double-digit analysis was 10,893,012.

First, using the first digit, the entire population (all four companies) and each individual company were compared to the expected results using Benford's Law. Second, the first two digits distribution was analyzed to determine if any trends exist for the entire population and each individual company. Healthcare providers with greater than 10,000 claims submitted for a particular company were selected and compared to Benford's Law predictions and the overall trend of the respective company. The cutoff (10,000 claims) was selected in order to have a sufficient sample size for analysis.

The data was loaded into Microsoft Access. A query was written to group the claims that began with the same digits together and count them. This query was then further refined to stratify the population by company and by provider. The Microsoft Access command 'LEFT (variable, position)' was used to isolate the first digit. A second query was written to group the claims that began with the same first two digits and count their frequency.

The expected results should indicate that the data for the entire population, corporate and high-volume provider levels would generally conform to Benford's Law for both the first and first two digits distributions. The data was statistically analyzed by used of the goodness of fit test and the Mean Absolute Value test. The goodness of fit test is based on the formula the sum of $(o_i - e_i)^2 / e_i$ for each i from 1 to 9. The chi-squared variable is very close to the chi-squared distribution with 8 degrees of freedom. The goodness of fit test takes into account the overall size of the population. The Mean Absolute Value is used to determine the absolute difference between the expected and actual values. This statistic is calculated by adding the absolute value of the differences between expected and actual proportions and dividing by 9. It does not take into account the size of the population.

Results

*Entire Population Analysis -- First Digit.*

In total, 11.6 million claims were analyzed. Table 2 indicates the results of the overall analysis. Chart 1 represents the results in a pictorial format.

Table 2
Entire Population Analysis

First Digit	Actual Count	Actual Percentage	Expected Percentage	Z-Statistic
1	3,192,305	0.274664	0.30103	26,840.0
2	1,825,197	0.157039	0.17609	23,355.4
3	1,398,525	0.120328	0.12494	1,978.7
4	1,143,879	0.098418	0.09691	272.7
5	1,131,544	0.097357	0.07918	48,498.9
6	1,007,062	0.086647	0.06695	67,352.4
7	772,474	0.066463	0.05799	14,388.8
8	635,128	0.054646	0.05115	2,777.2
9	516,488	0.044438	0.04576	443.9
Total	11,622,602	1.000000	1.00000	* 186,508.0

* The actual and expected ratios are statistically different at a 95% confidence interval

Chart 1
Entire Population Analysis
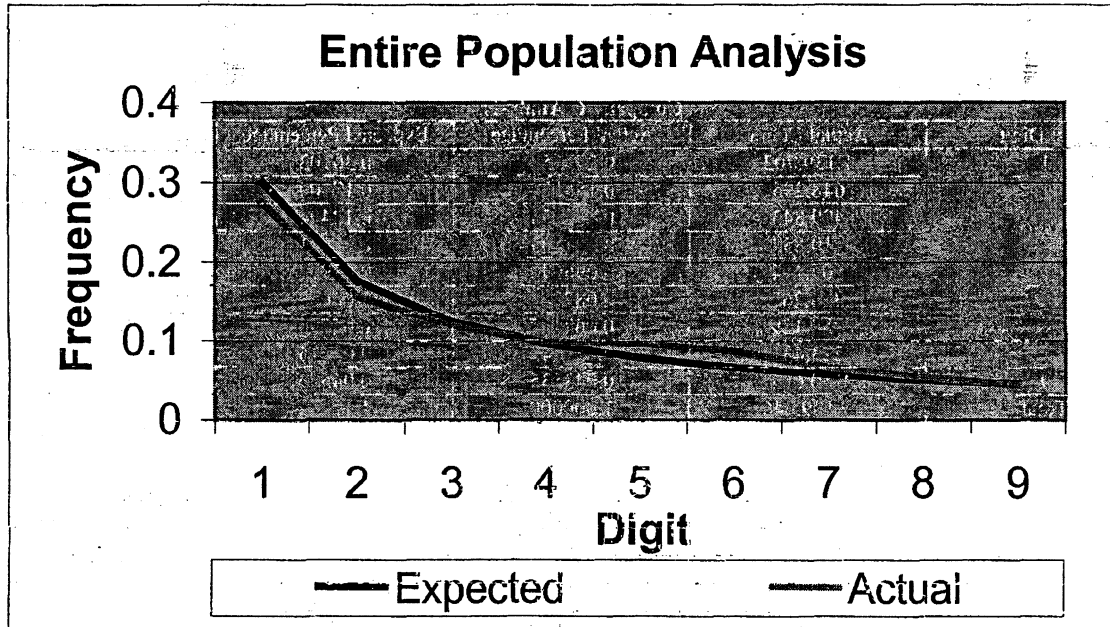
Although the actual percentages appear close to Benford's Law, they are statistically different at the 95% level. The z-statistics, Table 2, were used to determine whether the differences between the actual and expected proportions are significant. The Mean Absolute Value was calculated at 3.74 percent. Given the visual similarity between the two curves and the relatively low MAV, the first digit distribution of claims received is determined to be similar, but not statistically the same as predicted by Benford's Law. Due to the large number of claims used in the analysis, it was not cost justified to investigate the differences between the expected and actual results.

Table 3
Company A Analysis

First Digit	Actual Count	Actual Percentage	Expected Percentage	Z-Statistic
1	1,267,523	0.277760	0.30103	8,208.6
2	717,001	0.157121	0.17609	9,324.8
3	538,013	0.117898	0.12494	1,811.2
4	443,808	0.097254	0.09691	5.6
5	442,711	0.097014	0.07918	18,330.2
6	391,995	0.085900	0.06695	24,476.8
7	303,640	0.066538	0.05799	5,749.9
8	252,967	0.055434	0.05115	1,637.3
9	205,717	0.045080	0.04576	46.1
Total	4,563,375	1.000000	1.00000	* 69,590.6

Table 4
Company B Analysis

First Digit	Actual Count	Actual Percentage	Expected Percentage	Z-Statistic
1	483,421	0.273698	0.30103	4,383.2
2	278,553	0.157708	0.17609	3,389.3
3	218,473	0.123693	0.12494	22.0
4	173,737	0.098365	0.09691	38.6
5	170,407	0.096479	0.07918	6,675.4
6	151,857	0.085977	0.06695	9,550.9
7	113,198	0.064089	0.05799	1,133.0
8	96,635	0.054712	0.05115	438.1
9	79,974	0.045279	0.04576	8.9
Total	1,766,255	1.000000	1.00000	* 25,639.3

Table 5
Company C Analysis

First Digit	Actual Count	Actual Percentage	Expected Percentage	Z-Statistic
1	1,139,702	0.274650	0.30103	9,592.9
2	648,295	0.156229	0.17609	9,295.6
3	499,843	0.120454	0.12494	668.4
4	410,852	0.099009	0.09691	188.7
5	402,474	0.096990	0.07918	16,623.5
6	364,267	0.087783	0.06695	26,900.7
7	280,129	0.067507	0.05799	6,481.2
8	223,431	0.053843	0.05115	588.4
9	180,653	0.043535	0.04576	448.9
Total	4,149,646	1.000000	1.00000	* 70,788.4

Table 6
Company D Analysis

First Digit	Actual Count	Actual Percentage	Expected Percentage	Z-Statistic
1	301,659	0.263843	0.30103	5,252.2
2	181,348	0.158614	0.17609	1,983.0
3	142,196	0.124370	0.12494	3.0
4	115,482	0.101005	0.09691	197.8
5	115,952	0.101416	0.07918	7,139.5
6	98,943	0.086540	0.06695	6,553.7
7	75,507	0.066042	0.05799	1,278.3
8	62,095	0.054311	0.05115	223.3
9	50,144	0.043858	0.04576	90.4
Total	1,143,326	1.000000	1.00000	* 22,721.2

* The actual and expected ratios are statistically different at a 95% confidence interval

Analysis of the Four Companies – First Digit.

When each company was isolated, their digit frequencies were very similar to that of the entire population analysis. Table 3, 4, 5, and 6 show the analysis of each company. The results are statistically different at the 95% level according to the goodness of fit test. However, as stated above, the first digit distribution of claims received is similar to Benford's Law, and thus can be used to analyze claims submitted by high-volume providers.

Entire Population Analysis – First Two Digits.

The two-digit analysis, represented in chart 2, demonstrated a particularly surprising trend. The frequency of second digits that ended in a 5 or 0 was much higher than the expected frequency. After obtaining procedures regarding pricing strategies at various large hospitals based in the Midwest, it was noted that hospitals target specific percentage increases for each procedure each year. This specific percentage is applied across the board; however, small adjustments can be made to these prices due to the impact of Medicare pricing. Medicare will pay a predetermined service rate regardless of the charged price, thus as hospitals and physicians try to increase their prices to cover their costs, the costs are manually adjusted based on the amount of usage of Medicare patients. This explains the manual rounding that occurs at the 0 and 5 second-digit points. For example, if the provider is targeting an 8% increase and the new target price is $24.41, the provider may elect to manually round up to $25.00 to adjust for Medicare. However, as discussed in Section One, a key assumption in Benford's Law is that human interaction with the data will result in a distribution that will not conform to Benford's Law. This is both visually apparent in Chart 2 and statistically verified with the z-statistic calculations in Table 7. Thus, there are significant statistical differences between the entire population and Benford's Law for the first two digits analysis.

Chart 2
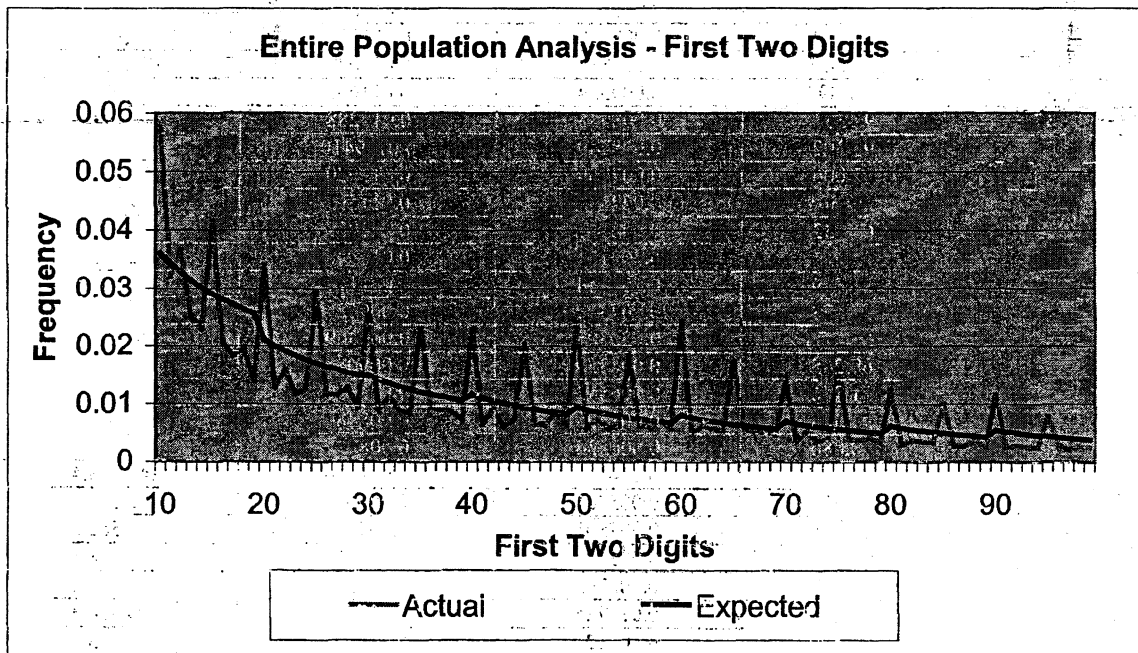**Entire Population Analysis – First Two Digits**

Table 7
Entire Population Analysis - First Two Digits

First Two Digits	Actual Count	Actual Percentage	Expected Percentage	Z-Statistic
10	627,704	0.057624	0.036027	141,029.3
11	349,469	0.032082	0.034284	1,541.1
12	397,992	0.036536	0.032758	4,747.1
13	272,253	0.024993	0.031406	14,264.9
14	249,531	0.022907	0.030196	19,165.4
15	452,955	0.041582	0.029104	58,281.3
16	224,223	0.020584	0.028107	21,934.2
17	198,428	0.018216	0.027198	32,311.4
18	214,403	0.019683	0.026361	18,431.2
19	151,917	0.013946	0.025588	57,692.8
20	370,214	0.033986	0.021074	86,173.1
21	140,231	0.012873	0.020055	28,012.3
22	173,485	0.015926	0.019162	5,952.3
23	126,741	0.011635	0.018371	26,906.7
24	137,055	0.012582	0.017664	15,925.2
25	319,168	0.029300	0.017024	96,422.8
26	127,651	0.011719	0.016442	14,778.4
27	125,956	0.011563	0.015910	12,936.3
28	139,487	0.012805	0.015420	4,830.7
29	102,110	0.009374	0.014968	22,772.1
30	294,816	0.027065	0.014953	106,867.6
31	96,594	0.008868	0.014229	22,009.0
32	123,150	0.011305	0.013596	4,203.6
33	97,142	0.008918	0.013035	14,165.6
34	91,389	0.008390	0.012533	14,919.1
35	259,455	0.023818	0.012079	124,277.2
36	101,890	0.009354	0.011666	4,991.1
37	97,333	0.008935	0.011288	5,342.6
38	95,302	0.008749	0.0109410	4,784.2
39	77,006	0.007069	0.010620	12,931.0
40	249,364	0.022892	0.011598	119,797.0
41	69,532	0.006383	0.011037	21,376.2
42	99,299	0.009116	0.010546	2,112.0
43	64,749	0.005944	0.010111	18,703.5
44	79,978	0.007342	0.009721	6,341.5
45	221,178	0.020305	0.009369	139,028.7
46	69,265	0.006359	0.009048	8,710.1
47	71,276	0.006543	0.008756	6,090.3
48	90,388	0.008298	0.008486	45.7
49	70,086	0.006434	0.008237	4,300.4
50	263,187	0.024161	0.009476	247,883.4
51	62,757	0.005761	0.009018	12,810.8
52	81,748	0.007505	0.008616	1,562.6
53	62,858	0.005770	0.008261	8,178.1
54	67,994	0.006242	0.007943	3,966.2
55	203,974	0.018725	0.007655	174,380.3
56	67,800	0.006224	0.007393	2,013.1
57	63,837	0.005860	0.007154	2,547.9
58	76,048	0.006981	0.006934	3.6
59	62,867	0.005771	0.006730	1,488.5
60	264,639	0.024294	0.008013	360,396.1
61	53,130	0.004877	0.007625	10,784.2
First Two Digits	Actual Count	Actual Percentage	Expected Percentage	Z-Statistic

62	67,255	0.006174	0.007285	1,846.7
63	62,316	0.005721	0.006985	2,492.3
64	57,924	0.005318	0.006716	3,171.1
65	187,080	0.017174	0.006473	192,733.0
66	63,154	0.005798	0.006251	358.3
67	53,488	0.004910	0.006049	2,334.7
68	62,833	0.005768	0.005863	16.6
69	50,917	0.004674	0.005691	1,977.7
70	155,475	0.014273	0.006940	84,391.0
71	39,975	0.003670	0.006604	14,204.9
72	59,556	0.005467	0.006310	1,227.1
73	38,245	0.003511	0.006050	11,608.0
74	43,646	0.004007	0.005817	6,136.2
75	178,098	0.016350	0.005606	224,249.1
76	41,744	0.003832	0.005415	5,037.2
77	40,163	0.003687	0.005239	5,010.1
78	44,974	0.004129	0.005078	1,933.8
79	36,069	0.003311	0.004929	5,785.0
80	136,175	0.012501	0.006122	72,419.1
81	32,127	0.002949	0.005825	15,468.3
82	40,144	0.003685	0.005566	6,923.1
83	38,663	0.003549	0.005336	6,519.5
84	36,858	0.003384	0.005131	6,481.2
85	109,559	0.010058	0.004945	57,575.8
86	29,750	0.002731	0.004776	9,536.4
87	30,322	0.002784	0.004621	7,960.9
88	40,924	0.003757	0.004479	1,268.8
89	29,243	0.002685	0.004348	6,930.5
90	127,095	0.011668	0.005477	76,236.5
91	27,877	0.002559	0.005212	14,705.2
92	30,400	0.002791	0.004980	10,480.4
93	27,090	0.002487	0.004774	11,936.4
94	26,932	0.002472	0.004590	10,643.4
95	84,667	0.007773	0.004424	27,607.7
96	30,667	0.002815	0.004273	5,414.6
97	23,969	0.002200	0.004134	9,855.0
98	29,233	0.002684	0.004007	4,762.0
99	27,412	0.002516	0.003890	5,280.4
TOTAL	10,893,023	1.000000	1.000000	* 3,083,637.7

* The actual and expected ratios are statistically different at a 95% confidence interval

Analysis of the Four Companies – First Two Digits.

Each of the four companies was analyzed to determine if the distribution was similar to Benford's Law for the first two digits. In every case, the results were similar to the trends noted in the Entire Population Analysis – First Two Digits discussion. The actual percentages and z-statistics have not been included in this section due to their similarity with the above results.

Analysis of High-Volume Providers for the Four Companies – First Digit.

Five providers were selected from Company A based on the criteria that the provider submitted more than 10,000 claims during the 10-month period under review, as depicted below in Table 8. Visually, one may notice by examining the z-Statistics that all providers, except for provider #47312, have a z-statistic that is lower than 6,000. Provider #47312 has a significant deviation from the expected results for the digits one and two. After further analysis of this provider, it was determined that this provider was submitting an unusually high volume of $2 claims

due to the types of services that it was required to provide. The insurance company found that these claims resulted from valid medical procedures.

Table 8
Company A High Volume Providers (Z-Statistics)

Digit	Provider #83838	Provider #39101	Provider #93199	Provider #47281	Provider #47312
1	152.2	32.7	192.5	883.0	** 5,820.6
2	1.3	854.1	389.3	577.5	** 13,542.5
3	33.8	2,010.9	879.3	1,449.7	57.6
4	285.9	8.2	40.0	217.7	216.9
5	449.9	0.3	407.5	120.9	311.3
6	0.6	1317.1	83.7	82.8	57.1
7	15.1	356.4	238.7	128.7	73.5
8	424.7	80.2	153.1	12.5	802.9
9	367.8	348.8	1.4	20.0	897.8
TOTAL	* 1,731.2	* 5,008.6	* 2,385.6	* 3,492.7	** 21,780.1

* The actual and expected ratios are statistically different at a 95% confidence interval
** The actual and expected ratios are statistically different at a 99.9% confidence interval

Three providers from Company B submitted more than 10,000 claims in the 10-month period. The z-statistics for these three providers are represented in Table 9. All results were statistically different at the 95% level except for those indicated with a "**", which were statistically different at the 99.9% level. Visually, the z-statistic for provider #12275 digit 6 appears much higher than expected. The insurance company determined that this provider was submitting valid procedures and charges.

Table 9
Company B High Volume Providers (Z-Statistics)

Digit	Provider #12275	Provider #23112	Provider #84912
1	20.6	8.4	383.9
2	332.2	294.6	423.9
3	545.5	373.1	822.2
4	33.1	15.7	67.5
5	2.1	148.8	3.7
6	** 872.3	46.4	38.3
7	286.5	78.7	118.4
8	16.1	21.4	2.1
9	192.0	17.7	112.4
TOTAL	* 2300.2	* 1004.8	* 1972.4

* The actual and expected ratios are statistically different at a 95% confidence interval
** The actual and expected ratios are statistically different at a 99.9% confidence interval

There are seven providers whom submitted more than 10,000 claims to Company C. These seven providers were analyzed as indicated in the research methodology section and the results can be found in Table 10. Upon visual inspection, the high frequency of the digit 4 for provider #12931, the high frequency of the digit 3 for provider #91882 and provider #92981, and the high frequency of the digit 2 for provider #18291 are suspect since they are statistically different at the 99.9% level. All other results in Table 10 are statistically different at the 95% level. The insurance company noted that these providers submitted valid claims.

Table 10
Company C High Volume Providers (Z-Statistics)

Digit	Prov 91822	Prov 38129	Prov 09132	Prov 48218	Prov 18291	Prov 12931	Prov 92981
1	120.2	733.8	54.1	165.6	103.8	2256.8	702.2
2	290.0	247.8	559.4	0.4	** 2201.1	88.0	342.0
3	** 1281.9	154.4	620.7	603.7	186.4	489.8	** 1844.7
4	69.1	5.4	35.4	420.2	151.4	** 3065.3	267.5
5	13.7	502.7	481.7	21.0	160.3	286.3	4.5
6	689.3	331.1	169.4	951.5	294.7	38.1	101.5
7	374.6	315.7	304.5	267.5	387.0	54.5	5.3
8	0.1	606.9	60.7	86.7	73.0	616.9	0.5
9	229.8	357.0	0.2	1.9	1.3	92.9	3.7
TOTAL	* 3068.7	* 3254.9	* 2286.2	* 2518.4	* 3559.0	* 6988.8	* 3271.9

* The actual and expected ratios are statistically different at a 95% confidence interval
** The actual and expected ratios are statistically different at a 99.9% confidence interval

Company D did not have any providers who submitted more than 10,000 claims and thus no providers were analyzed for this company.

Conclusion

Benford's Law is an excellent tool to predict the distribution of the first digit or first two digits in a large population of data, given that the data has not been interfered with human interaction.   Given conformity to Benford's Law, one can use this tool as a method of detecting possible fraudulent or errant claims received on behalf of a health insurance company. This study reviewed the overall conformity of the entire population to Benford's Law at the first digit and first two digits level.

This study detected several possible irregularities that required further investigation.   There were six providers that were isolated as high-risk for possible errant claims. With a reasonable sample of high-risk claims, the company could determine if the claims were either valid or invalid. Without the Benford's Law test, the company would not have known that certain providers are charging for large numbers of selected services. The review of the two-digit distribution indicated that hospitals and physicians are rounding their prices to maximize their profits while still in compliance with governmental regulations. Thus, the expectation from the two-digit analysis of Benford's Law does not provide results that are comparable to the claims received from health care providers. The insurance companies whom participated in this study have implemented a periodic review to detect if there are any abnormal trends in the data received from providers.

Suggestions for Future Research

Other possible applications of Benford's Law in the insurance industry include the review of property and casualty claims such as automobile or property claims. At the hospital and physician level, its use would facilitate the monitoring of medical suppliers' bills that could be fraudulent and would eventually be paid by insurance companies. Future research could be performed to ensure that the curve depicted in this article is truly similar to Benford's Law and not inadvertently similar to Benford's Law.   📖

References

1. "Numerology for accountants," *Journal of Accounting*, Vol. 186, Issue 5, November, pp 15, 1998.
2. "What are the odds?" *CFO*, Vol. 17, Iss. 2, February, pp 17-22, 2001.
3. Coderre, David and Warner, Paul D, "Computer-Assisted Techniques for Fraud Detection," *CPA Journal*, Vol. 69, Iss. 8, August, pp 57-60, 1999.
4. Coderre, David G, "Computer Assisted Fraud Detection," *Internal Auditor*, Vol. 57, Iss. 4, August, pp 25-28, 2000.
5. Crowder, Nita, "Fraud detection techniques," *Internal Auditor*, Vol. 54, Iss. 2, April, pp 17-21, 1997.
6. Ettredge, Michael and Srivastava, Rajendra, "Using Digital Analysis to Enhance Data Integrity," *Issues in Accounting Education*, Vol. 14, Iss. 4, November, pp 675-691, 1999.
7. Fogarty, Timothy J, "Everything You Know About Auditing is Probably Wrong: Practice Realities in the '00s," *Ohio CPA Journal*, Vol. 59, Iss. 3, July-September, pp 45-50, 2000.
8. Irmay, Shragga, "The relationship between Zipf's law and the distribution of first digits," *Journal of Applied Statistics*, Vol. 24, Iss. 4, August, pp 383-394, 1997.
9. Kogan, Alexander, Sudit, Ephraim, and Vasarhelyi, Miklos, "Continuous Online Auditing: A Program of Research," *Journal of Information Systems*, Vol. 13, Iss. 2, Fall, pp 87-104, 1999.
10. Matthews, Robert, "Benford bend," *World Link*, May/June, pp 10-12, 2000.
11. McConville, Daniel J, "Benford's law traps check fraud perps," *Corporate Cashflow*, Vol. 16, Iss. 9, September, pp 12, 1995.
12. Nigrini, Mark J, "A taxpayers compliance application of Benford's Law," *The Journal of the American Taxation Association*, Vol. 18, Issue 1, Spring, pp 72-92, 1996.
13. Nigrini, Mark J, "Adding value with digital analysis," *The Internal Auditor*, Vol. 56, Iss. 1, February, pp 21-24, 1999.
14. Nigrini, Mark J, "I've got your number," *Journal of Accountancy*, Vol. 187, Iss. 5, May, pp 79-84, 1999.
15. Nigrini, Mark J, "The use of Benford's Law as an aid in analytical procedures," *Auditing*, Vol. 16, Issue 2, Fall, pp 52-68, 1997.
16. Tapp, Darren J, "Using technology to detect fraud," *Pennsylvania CPA Journal*, Vol. 71, Iss. 4, Winter, pp 20-24, 2000.
17. Tutton, Michael, "Lover of the law of numbers," *CA Magazine*, Vol. 128, Iss. 4, May, pp 7, 1995.
18. US General Accounting Office, "High Risk Program: Information on Selected High-Risk Areas," *GAO* report HR-97-30, May 16, 1997.
19. York, David, "Benford's Law," *Accountancy*, Vol. 126, Iss. 1283, July, pp 126, 2000.