Benfords Law and Fraud Detection

It is often the simplest tools that are the most effective in fraud detection. Often times we get caught up in the technology that we some times miss the forest because of the trees. Such is the case with Benford's Law. A simple tool that can be used to determine anomalies in the data that may lead an investigator to look further for inappropriate activity.

The first digit of a number can be in the range of 1-9. So it would follow logically that the probability of that number being a 9 for example would be 1 in 9 or 11.1 percent. In actual fact, Dr. Frank Benford, a physicist with General Electric determined that the probability was actually 4.58 percent. He calculated this by performing a mathematical analysis of 20,229 sets of numbers, including such wildly disparate categories as the areas of rivers, baseball statistics, numbers in magazine articles and the street addresses of the first 342 people listed in the book "American Men of Science." All these seemingly unrelated sets of numbers followed the same first-digit probability pattern. In all cases, the number 1 turned up as the first digit about 30 percent of the time, more often than any other. He discovered that the expected rate of occurrence for the first digit followed the following formula

Probability p of any given first digit is log to the base 10 of (1 + 1/p).

The resulting table then predicts the frequency of the first digits

Digit	Percentage
1	30.10
2	17.61
3	12.49
4	9.69
5	7.92
6	6.69
7	5.80
8	5.12
9	4.58

Let us now apply this knowledge to a fictitious engagement we have been asked to perform. We have been asked to investigate a small company that specializes in the manufacture of widgets. It should also be noted that any invoice under $1000 does not require a purchase order. We will take a sampling of 30 invoice amounts which are listed below. In real life of course, your statistical sample will be much larger

1258.86	955.56	823.49	922.34	998.79
125.33	544.62	945.11	1845.77	887.34
478.51	625.82	925.28	782.65	1587.65
2025.58	936.47	583.55	5425.36	854.25
996.57	2858.36	847.36	688.36	2588.25
1617.88	936.36	453.55	772.77	33.85

The following table shows our analysis of the data

Digit	Number of Occurrences	Percentage Occurrence	Benford's Law
1	5	16.66	30.10
2	3	10.00	17.61
3	1	3.33	12.49
4	2	6.66	9.69
5	3	10.00	7.92
6	2	6.66	6.69
7	2	6.66	5.80
8	4	13.33	5.12
9	8	26.66	4.58

It is clear that the percentage occurrence of the numbers do not match those of Benford's law. Even more importantly we see that the digit 9 is completely out of line with what we would have expected. This is a red flag and should be investigated. You will have noted that I mentioned that invoices under $1000 did not require a purchase order. Given the fact that many of the invoice amounts were between $900 and $1000 this looks like a ripe area to look for fraudulent activity. Each of the invoices in the $900-$1000 range should now be pulled to look for bogus invoices

Any deviation greater than 2 percent should be considered a red flag and investigated provided of course that your sample is statistically large enough. As a rule of thumb you should take a minimum of 10% of the sample set to perform your analysis. More of course is better. It is also important to note that the digital frequency remains extremely consistent from one period to the next. Sampling the data for the previous period, or even better, over multiple periods is an excellent way to tell if there are any areas that should be investigated.

Let us take a practical example. If you suspect a cashier is putting in bogus transactions then an analysis of the first digits will prove to be very helpful. First, run an analysis of each of the cashiers transactions for multiple periods of time. (i.e. monthly). The numbers should be remarkably similar especially when compared against the same period of time. If one cashier shows a different pattern than the norm, clearly there are problems that need to be further investigated. This type of analysis is extremely well suited to

businesses with high transaction volumes. This would include restaurants, bars and retail outlets, however it is equally as effective in analyzing businesses (large or small) when comparing the numbers against an established norm.