CORNELL CENTER FOR HOSPITALITY RESEARCH

A Quick and Easy Approach to

Financial Fraud Detection

by Pamela C. Moulton and Fang Liu

EXECUTIVE SUMMARY

Financial fraud is a significant cost in the hospitality industry. According to the Report to the Nations on Occupational Fraud and Abuse, the typical organization loses 5 percent of its annual revenues to fraud.[1] Hotels in particular are estimated to lose 5 to 6 percent of revenues to fraud on average, while the National Restaurant Association estimates that restaurants on average lose 4 percent of revenues to fraud. These are losses as a percentage of top-line revenues, not profits, meaning that their magnitudes represent a significant risk to hospitality firms, given the industry's relatively thin net margins. This study presents a simple methodology for detecting financial irregularities that may signal fraud based on a mathematical principle known as Benford's Law. The analysis presented here can be applied by hospitality industry managers at all levels, from individual units or departments to entire regions or companies. The Cornell Hospitality Tool accompanying this report provides an easy-to-use spreadsheet-based application that can be used to quickly analyze any set of financial values (for example, guest checks, receivables, payables, or reimbursements) to quickly detect suspicious activities.

[1] Association of Certified Fraud Examiners, Report to the Nations on Occupational Fraud and Abuse, 2016 Global Fraud Study.

ABOUT THE AUTHORS

Pamela Moulton, PhD, is an associate professor of finance and area coordinator for finance at the School of Hotel Administration, Cornell SC Johnson College of Business. Her teaching and research interests include financial markets and market microstructure, with a special interest in the role of investors. Her current research focuses on the impact of high-frequency trading on stock performance, the role of designated and voluntary market makers in stock liquidity, and detecting fraud in financial statements. Moulton's research has been published in several of the leading finance and accounting journals, including the *Journal of Finance*, the *Journal of Financial Economics*, the *Journal of Accounting and Economics*, and the *Journal of Financial and Quantitative Analysis*. Prior to her academic career, Moulton worked in fixed-income research for more than a dozen years at various Wall Street investment banks, including Deutsche Bank, where she was a managing director and global co-head of relative value research. From 2003 to 2006, she was a managing director and senior economist at the New York Stock Exchange, where she focused on equity market microstructure research.

Fang Liu, PhD, is an assistant professor of finance at the Cornell SC Johnson College of Business, School of Hotel Administration. She earned a PhD in business administration (finance) from Washington University in St. Louis, a master's degree in financial mathematics from Stanford University, and a bachelor's degree in finance from Nanjing University (China). Liu's research interests are in the areas of theoretical and empirical asset pricing, options, volatility and risk, and jumps and disasters. Her recent work focuses on extracting forward-looking information from option prices and using such information for the prediction of stock returns and jumps. Liu's research has been published in the *Journal of Financial Economics, the American Economic Journal: Microeconomics*, and the *Journal of Economic Theory*.

*The authors give special thanks and recognition to **Zahra Abdulhussein**, School of Hotel Administration Class of 2019, for her outstanding work refining the Benford digit analysis tool and dramatically expanding its capabilities.*

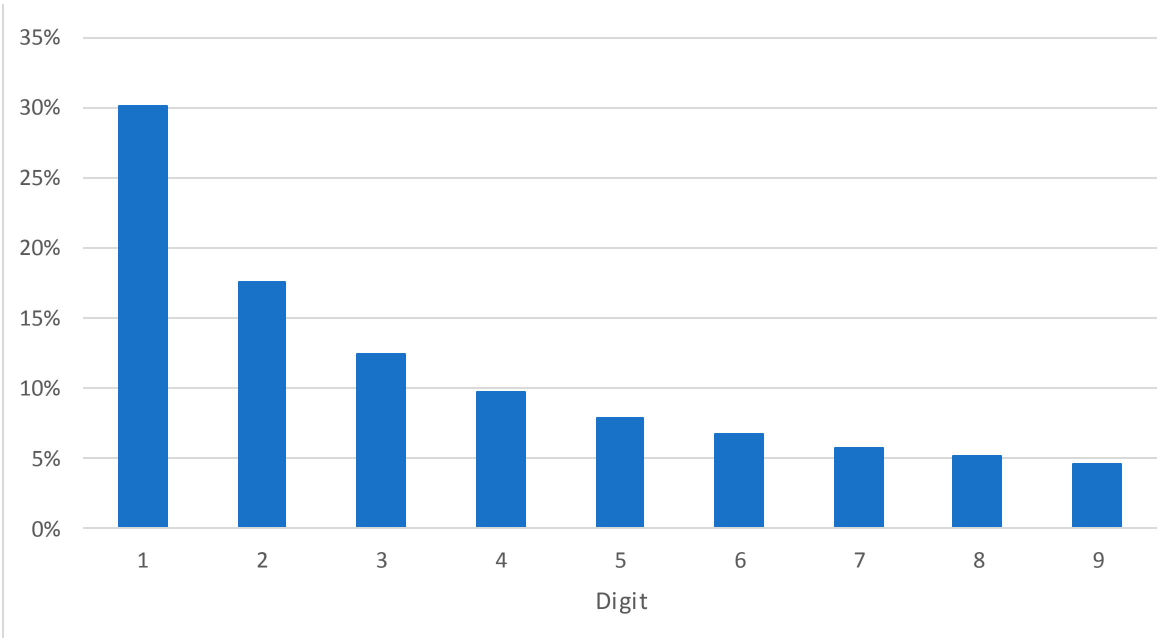
A Quick and Easy Approach to *Financial Fraud Detection*

by Pamela C. Moulton and Fang Liu

The nature of the hospitality industry—with its many disparate transactions and potential failure of internal control—exposes its firms to financial fraud, despite operators' best efforts. Ideally, a good hotel manager, for example, wants to spend his or her time greeting guests and creating memorable experiences for them, rather than spending lengthy time going through guest checks, payables, and receivables with a fine-tooth comb. Likewise, restaurant managers are moving at considerable speed during their busy times.

Exhibit 1

Frequency distribution for first digits under Benford's Law



At the same time, managers are well aware that if internal controls are not in place or are being subverted, then by the time auditors review the books after year-end, significant sums may already be lost.[2] In the largest available study of financial fraud cases, the Association of Certified Fraud Examiners found a median loss due to fraud of $150,000 in 2016, with 23 percent of cases resulting in losses of more than $1 million. Interestingly, the median loss in dollar terms was the same for small organizations (those with fewer than 100 employees) as for the largest organizations (those with more than 10,000 employees). One can easily infer that a dollar loss of a particular magnitude is likely to have a much greater impact on smaller organizations. The same study found that the size of losses and duration of fraud schemes were significantly lower when frauds were discovered through active detection methods rather than in annual audits.

Given that hospitality operators do want to check regularly for potential fraud, our goal in this paper is to introduce a simple way for managers to quickly detect irregularities in financial numbers. We build our tool on Benford's Law, a mathematical principle that predicts the frequency with which any of the numbers 1 through 9 should appear as the first digit (or 0 through 9 as a second or subsequent digit) in naturally occurring sets of numbers, as opposed to numbers that have been "cooked up" or manipulated by humans engaging in fraud. In this report, we first examine the origins of Benford's Law, and then we apply it to three samples of data from hospitality operations to see how well they conform to the predicted distributions (and how to draw appropriate conclusions when they do not do so). Finally, we introduce a Cornell Hospitality Tool for hospitality managers to apply Benford's Law to identify potential irregularities in their own data.

Benford's Law and Fraud Detection

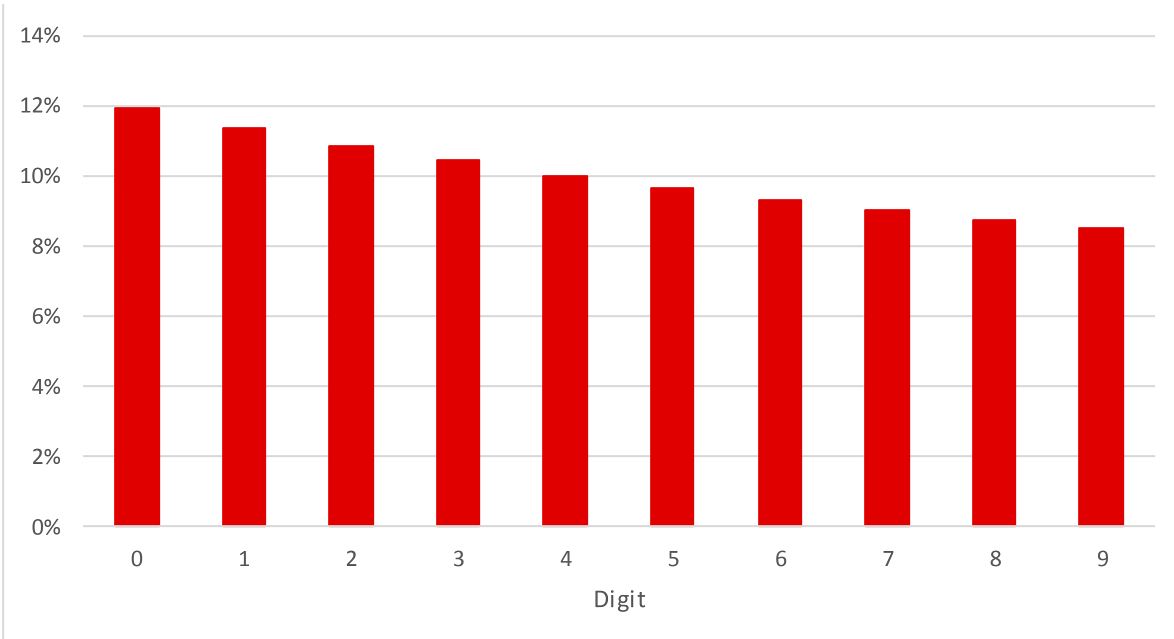
Benford's Law is named for the physicist Frank Benford, who in 1938 published an article predicting the frequency with which a particular number would appear in a given position (for example, first or second digit) in a set of naturally occurring numbers (that is, numbers not chosen by humans). At first thought, one might expect that each of the nine possible digits has

---

2 For a straightforward presentation of internal control mechanisms, see: A. Neal Geller, *Internal Control: A Fraud Prevention Handbook for Hotel and Restaurant Managers* (Ithaca, NY: Cornell University School of Hotel Administration, 1992).

Exhibit 2

Frequency distribution for second digits under Benford's Law



an equal probability of occurring in the first-digit position of a number (for example, the 2 in 2186). Thus, in a large set of numbers one might expect that approximately 1/9th, or about 11 percent, would start with each digit. But Benford observed that the early pages in books of logarithm tables were far more worn than the later pages, leading him to conclude that people must look up numbers beginning with 1 more often than numbers beginning with 2, and so on. This in turn led him to wonder whether numbers starting with 1 occur more frequently in nature than those beginning with 9.

Based on his analysis of natural numbers, Benford proposed that the frequency of first digits followed a logarithmic distribution, as follows. He proposed that a randomly selected number should begin with the digit 1 about $\log_{10}(1+\frac{1}{1})$, or 30.1 percent of the time, the frequency of numbers with leading digit 2 should be $\log_{10}(1+\frac{1}{2})$, or about 17.6 percent, and so on until the frequency of 9s should be $\log_{10}(1+\frac{1}{9})$, or about 4.6 percent. Benford found support for his proposition in over 20,000 numbers from 20 different sources, including data on river surface areas, populations, specific heats of chemical compounds, and American League baseball statistics.

An intuitive explanation for Benford's Law arises from the observation that a smaller percentage increase is required to move from one first digit to the next at higher first digits than at lower first digits.[3] For example, if a company has sales of $1 million, it must double its sales to move to $2 million, but to increase from $2 million to $3 million it needs to increase sales by only 50 percent. Thus, we can think of it as "harder" to move from a first digit of 1 to 2 in a series than to move from 2 to 3, which suggests that more observations would have a first digit of 1 than 2, and so on. Exhibit 1 presents the frequency distribution for first digits predicted by Benford's Law.

Using this logic, Benford also derived a rule for the frequency with which different digits should appear in the second position in a number. There are ten possible digits (0 through 9) in the second position, and the predicted distribution of second digits indeed decreases from 0 to 9, but much less sharply than the decrease for first digits, with the percentages dropping from about 12 percent for 0 to 8.5 percent for 9. Exhibit 2 presents the frequency distribution for second digits predicted by Benford's Law.

There are three main conditions for Benford's Law to hold. First, the numbers being examined should

[3] For the formal mathematical proof of Benford's Law, see: Thomas Hill, "The Significant-digit Law," *Statistical Science*, Vol. 10 (1995), pp. 354–363.

be "naturally occurring," meaning that they are not chosen directly by humans, since humans are known to have biases in selecting numbers. For example, prices would not be expected to follow Benford's Law, because they are often chosen strategically to influence consumer choices. On the other hand, total sales (sums of price times quantity) would be more likely to adhere to Benford's Law. Second, to permit first-digit and second-digit analysis, the numbers should have at least three significant digits (for example, 123 or 1.23, but not 12 or 1.2). Third, the set of numbers needs to be large enough for the frequency distribution to become apparent. Larger datasets are better, and datasets with fewer than 100 observations will almost certainly be too noisy to reveal anything of interest.

Benford's Law has been used in a wide variety of situations to detect when numbers have been manipulated. For example, political scientists have used Benford's Law to detect voting irregularities, economists have found that corrected corporate financial statements conform more closely to Benford's Law than the original manipulated versions, and auditors have found Benford's Law to be a useful tool in detecting accounting irregularities.[4] Such studies use the difference between the actual digit distribution in a set of numbers and the distribution predicted by Benford's Law as a signal that the numbers may have been manipulated and therefore merit further investigation. In one documented case, for instance, a manager noticed that a department suddenly was issuing far more reimbursement checks with a first digit of 1 than predicted by Benford's Law. The employee in charge of reimbursements explained that checks were being issued only after the totals reached $1,000 instead of for smaller individual amounts. While this was a reasonable explanation, it was not the whole story. The auditor subsequently discovered that a large percentage of those reimbursements over $1,000 were being paid to shell companies set up by that employee in charge of reimbursements.

This and other examples of Benford's Law highlighting suspicious patterns in financial data suggest that despite Benford's Law being publicly revealed 80 years ago, manipulators of financial data are either unaware that it can be used to detect their activities or unable to make their fraudulent behavior conform to Benford's Law.

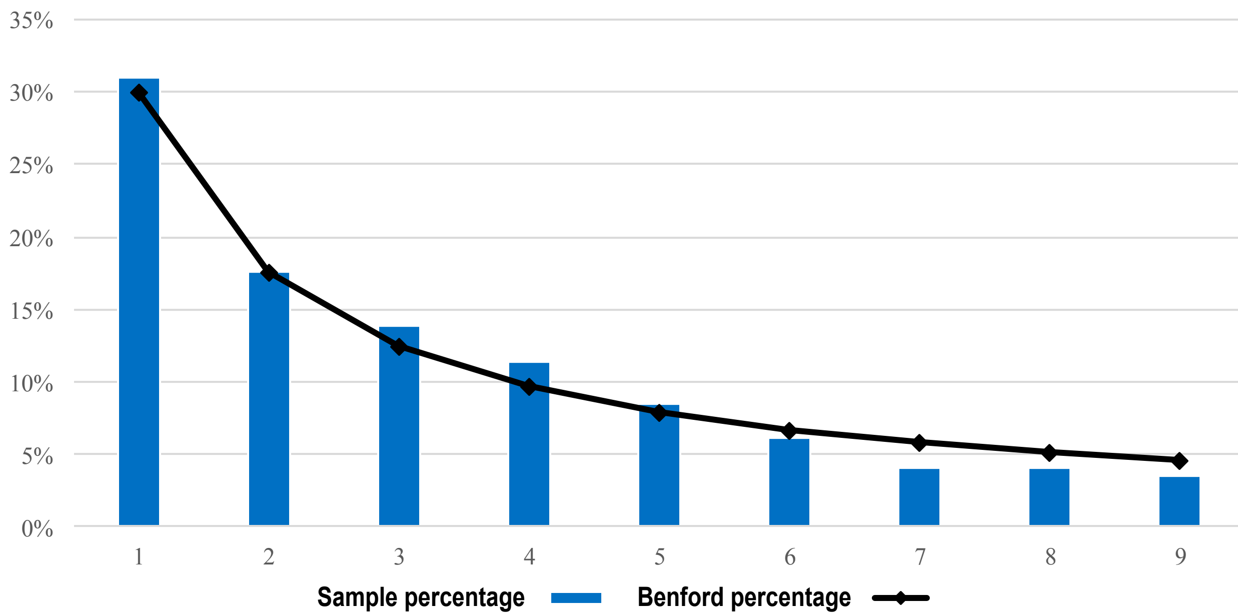
Benford's Law in Hospitality Data

On the following pages, we examine three sets of real data from hospitality operations to illustrate how the data may conform to Benford's Law (or not!) and to get a sense for how this analysis may be used in a hospitality context to detect financial fraud.

[4] See: Walter Mebane, "Election Forensics: Vote counts and Benford's Law," Political Methodology Society, University of California, Davis, (2006); Dan Amiram, Zahn Bozanic, and Ethan Rouen, "Financial statement errors: evidence from the distributional properties of financial statement numbers," *Review of Accounting Studies,* Vol. 20 (2015), pp. 1540-1593; and Cindy Durtschi, William Hillison, and Carl Pacini, "The effective use of Benford's law to assist in detecting fraud in accounting data," *Journal of Forensic Accounting,* Vol. 5 (2004), pp. 17-34.

Exhibit 3

First-digit frequency distribution for fine-dining checks vs. Benford's Law



	First Digit Summary								
	1	2	3	4	5	6	7	8	9
Digit counts	756	429	339	279	208	149	99	98	86
Sample percentage	30.95%	17.56%	13.88%	11.42%	8.51%	6.10%	4.05%	4.01%	3.52%
Benford predicted percentage	30.10%	17.61%	12.49%	9.69%	7.92%	6.69%	5.80%	5.12%	4.58%
Absolute deviation	0.84%	0.05%	*1.38%*	*1.73%*	0.60%	0.60%	***1.75%***	***1.10%***	***1.06%***
Mean absolute deviation	*1.01%*								

Note: *Italics* indicates significance at 5%; ***bold italics*** indicates significance at 1%.

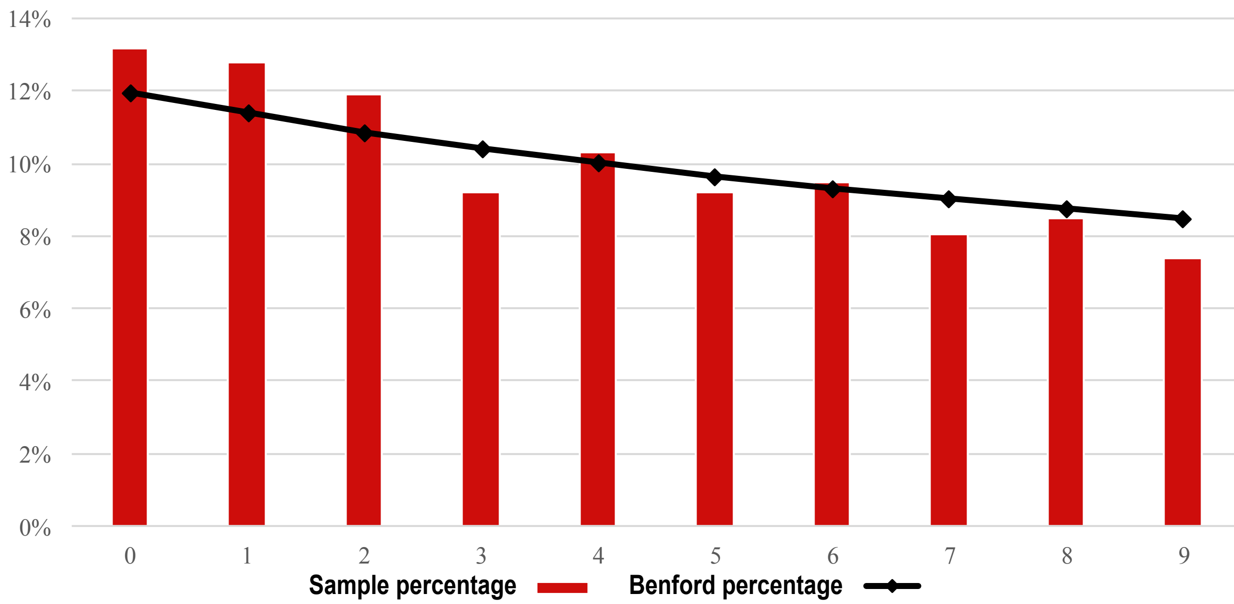
Fine-dining Restaurant Test

Dataset 1: The first dataset comprises 2,443 dinner checks from a hotel's fine-dining restaurant. Exhibit 3 compares the distribution of first digits in this dataset to the distribution predicted by Benford's Law. The top graph shows that the distribution of first digits (represented by the blue columns) is close to the Benford prediction (the black markers on the line). The table below the graph reports the sample percentages for each digit, subtracts the Benford percentage to compute the absolute deviation, and then calculates the mean absolute deviation, which provides a summary statistic for how well the data conform to Benford's Law. The frequencies of some of the individual digits deviate significantly, indicated by differences that are in *italics* (significant at the 5% level) and ***bold italics*** (significant at the 1% level).[5] Although the mean absolute deviation (1.01 percent) is statistically significant, practitioners generally view mean absolute deviations that are around 1 percent or less as not particularly suspicious. This analysis confirms the graphic evidence that the first digits fit Benford's Law reasonably closely. Thus, we would not suspect that

[5] We assess the statistical significance of individual digit absolute deviations using z-statistics and test their joint significance using a chi-square test.

Exhibit 4

Second-digit frequency distribution for fine-dining checks vs. Benford's Law



	Second Digit Summary									
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Digit counts	322	313	291	225	252	225	231	197	207	180
Sample percentage	13.18%	12.81%	11.91%	9.21%	10.32%	9.21%	9.46%	8.06%	8.47%	7.37%
Benford predicted percentage	11.97%	11.39%	10.88%	10.43%	10.03%	9.67%	9.34%	9.04%	8.76%	8.50%
Absolute deviation	1.21%	*1.42%*	1.03%	1.22%	0.28%	0.46%	0.12%	0.97%	0.28%	*1.13%*
Mean absolute deviation	*0.81%*									

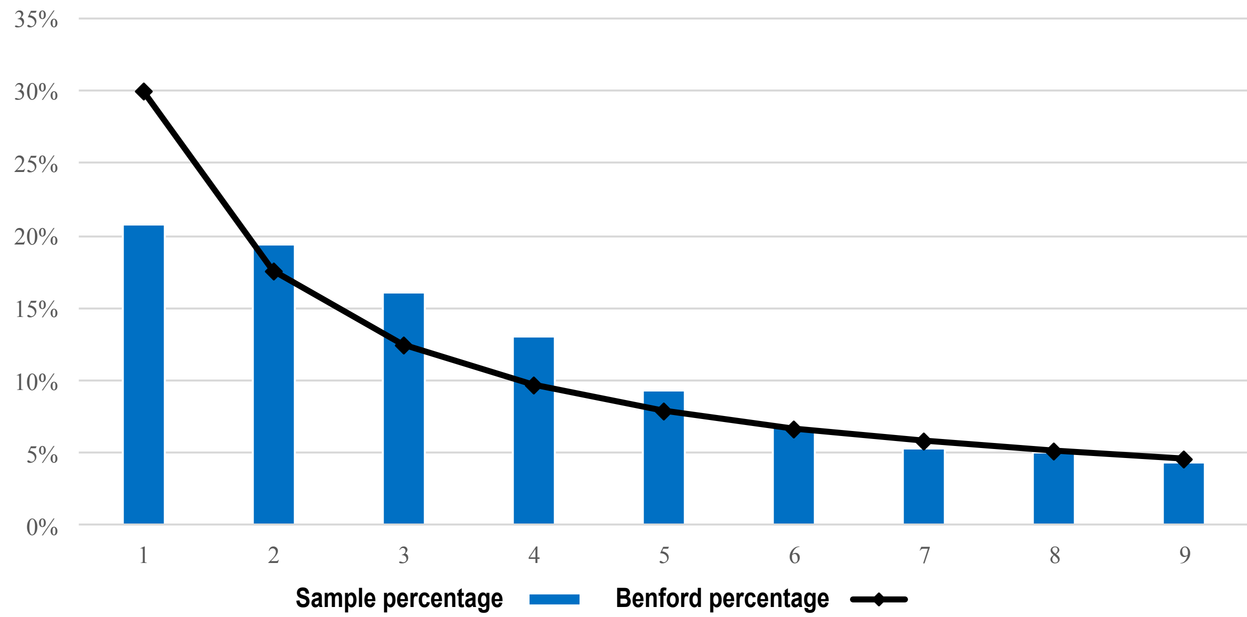
Note: *Italics* indicates significance at 5%; ***bold italics*** indicates significance at 1%.

these restaurant checks have been manipulated overall, because they seem to fit the expected natural distribution of leading digits.

We next look at the distribution of second digits in the same series of fine-dining restaurant checks, comparing the distribution to the predictions of Benford's Law for second-digit distributions, as shown in Exhibit 4. Although the frequencies for two of the individual digits deviate significantly from Benford's Law, overall the second digit distribution is close to that of Benford's Law, with a mean absolute deviation of only 0.81 percent (once again statistically significant, but below the 1-percent level used in practice). This confirms our assessment based on the first digits that there is nothing particularly suspicious in this dataset. Together these tests also confirm the applicability of Benford's Law to the fine-dining restaurant checks.

Exhibit 5

First-digit frequency distribution for hotel guest folios vs. Benford's Law



	First Digit Summary								
	1	2	3	4	5	6	7	8	9
Digit counts	3,285	3,067	2,535	2,068	1,467	1,083	833	797	678
Sample percentage	20.77%	19.40%	16.03%	13.08%	9.28%	6.85%	5.27%	5.04%	4.29%
Benford predicted percentage	30.10%	17.61%	12.49%	9.69%	7.92%	6.69%	5.80%	5.12%	4.58%
Absolute deviation	***9.33%***	***1.79%***	***3.54%***	***3.39%***	***1.36%***	0.15%	***0.53%***	0.08%	0.29%
Mean absolute deviation	*2.27%*								

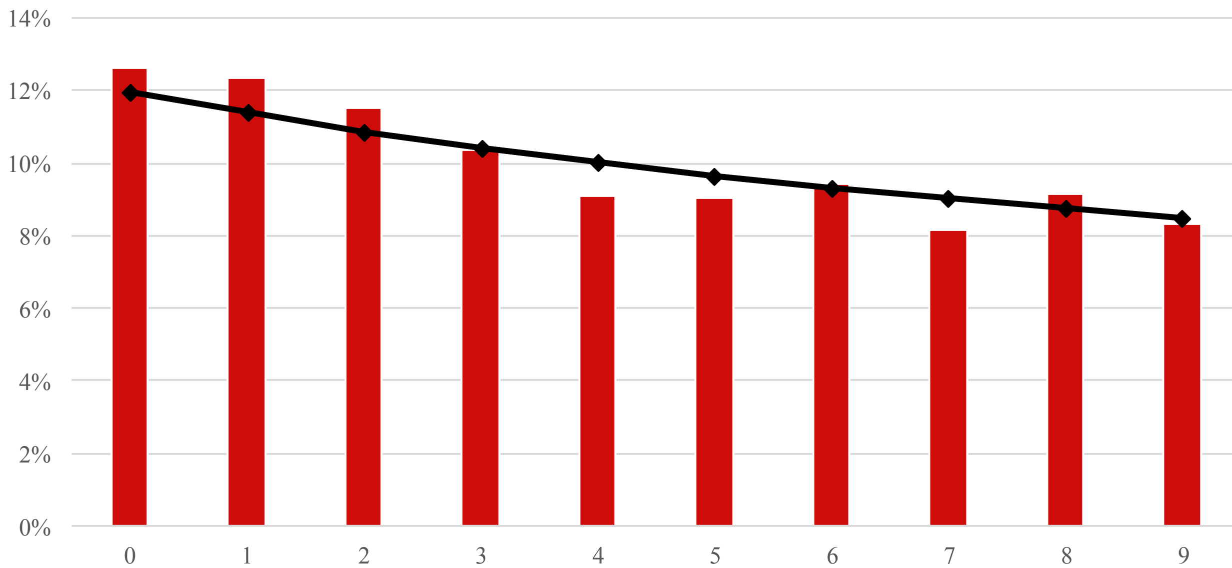
Note: *Italics* indicates significance at 5%; ***bold italics*** indicates significance at 1%.

Upscale Hotel Test

Dataset 2: This dataset includes 15,813 guest-stay observations for guest folios from an upscale hotel. Exhibit 5 presents the analysis of the first-digit distribution versus Benford's Law, showing that several digits deviate significantly from the Benford prediction. For example, only about 21 percent of the guest totals begin with 1, compared to a prediction of 30.1 percent. This produces a significant absolute deviation of 9.33 percent. This divergence accounts for most of the 2.27-percent mean absolute deviation, which is both statistically significant and is above the 1-percent practical threshold. The key question then becomes whether the lower frequency of 1s (and therefore higher frequency of other digits) in this dataset is reasonable or suspicious. Given the room rates in this upscale hotel, guests are unlikely to have total bills under $200. They are also unlikely to have total bills

Exhibit 6

Second-digit frequency distribution for hotel guest folios vs. Benford's Law



	Second Digit Summary									
	0	1	2	3	4	5	6	7	8	9
Digit counts	1,995	1,950	1,822	1,636	1,434	1,432	1,489	1,293	1,447	1,315
Sample percentage	12.62%	12.33%	11.52%	10.35%	9.07%	9.06%	9.42%	8.18%	9.15%	8.32%
Benford predicted percentage	11.97%	11.39%	10.88%	10.43%	10.03%	9.67%	9.34%	9.04%	8.76%	8.50%
Absolute deviation	*0.65%*	***0.94%***	*0.64%*	0.09%	***0.96%***	***0.61%***	0.08%	***0.86%***	0.39%	0.18%
Mean absolute deviation	***0.54%***									

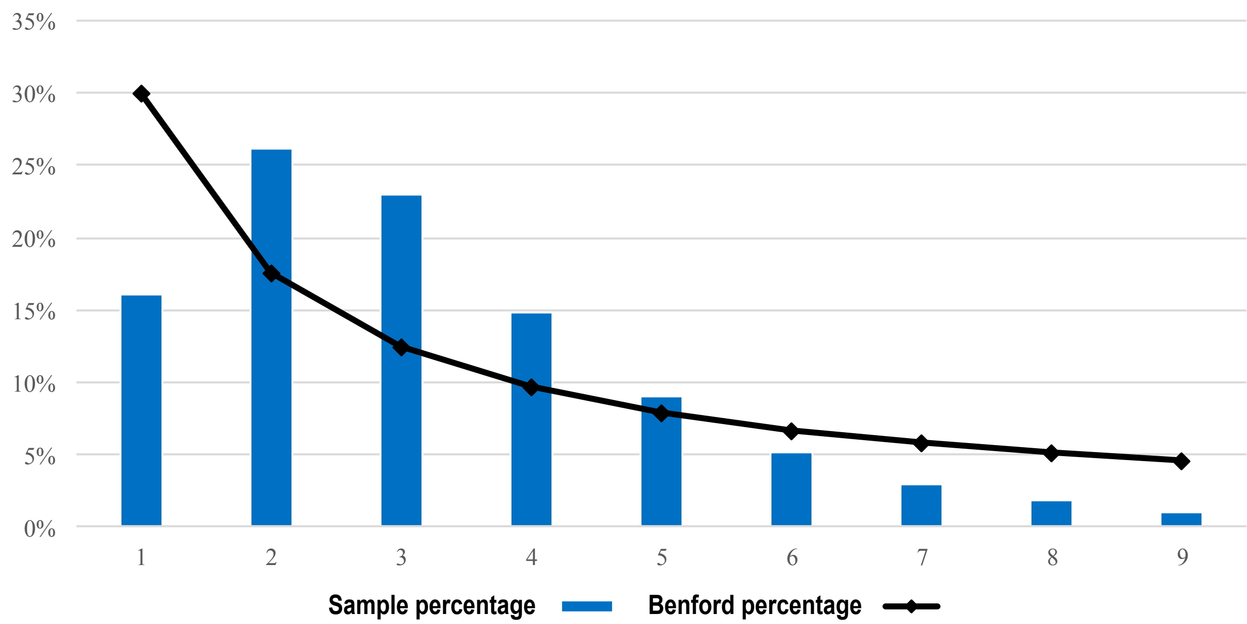
Note: *Italics* indicates significance at 5%; ***bold italics*** indicates significance at 1%.

of $1,000 or more, since most stays last for one or two nights. In this context, the relative dearth of records beginning with 1 should not be surprising (together with more that begin with 3 or 4). So, a mean absolute deviation of 2.27 percent for the first digit does not appear to signal suspicious behavior.

In a dataset such as this, where there is a natural reason for leading digits not to follow Benford's Law, it is more useful to focus on the second-digit distribution, an analysis which is presented in Exhibit 6. We would not expect to see any bias in the second-digit distribution for hotel guest folios, and indeed the second-digit distribution overall is close to that predicted by Benford's Law. The mean absolute deviation of only 0.54 percent falls below our practical threshold of 1 percent.

EXHIBIT 7

First-digit frequency distribution for chain restaurant checks vs. Benford's Law



	First Digit Summary								
	1	2	3	4	5	6	7	8	9
Digit counts	39,650	64,648	56,836	36,714	22,159	12,616	7,401	4,394	2,557
Sample percentage	16.05%	26.18%	23.01%	14.87%	8.97%	5.11%	3.00%	1.78%	1.04%
Benford predicted percentage	30.10%	17.61%	12.49%	9.69%	7.92%	6.69%	5.80%	5.12%	4.58%
Absolute deviation	***14.05%***	***8.57%***	***10.52%***	***5.17%***	***1.05%***	***1.59%***	***2.80%***	***3.34%***	***3.54%***
Mean absolute deviation	***5.63%***								

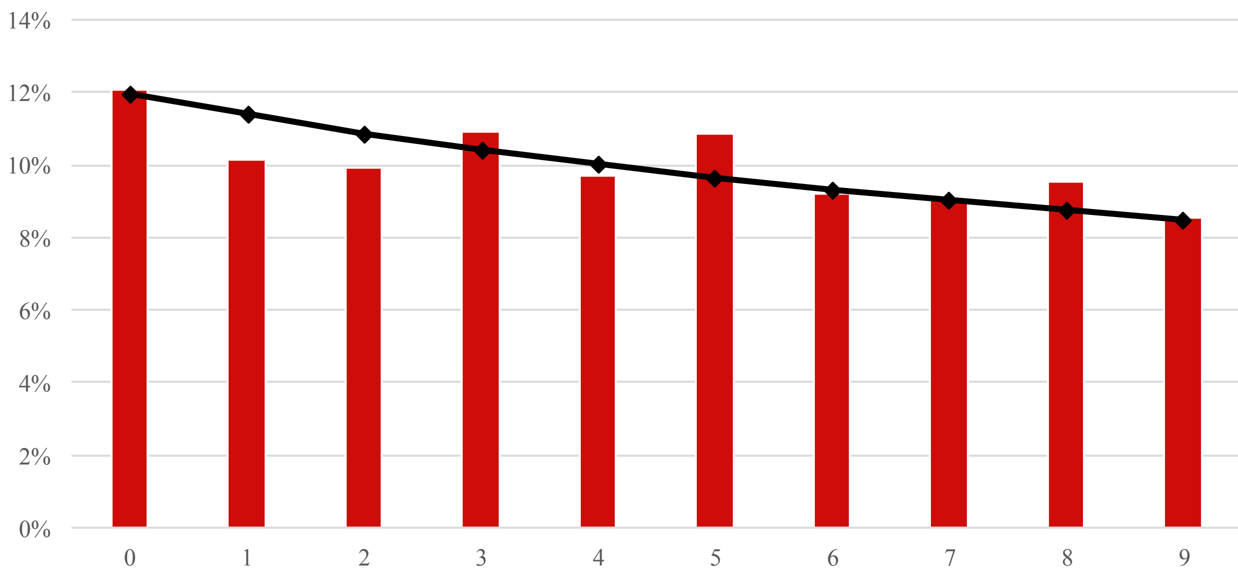
Note: *Italics* indicates significance at 5%; ***bold italics*** indicates significance at 1%.

Franchise Restaurant Test

Dataset 3: For this test, the dataset comprises 246,975 observations of restaurant checks from 45 chain restaurants operated by a single franchisee. Exhibit 7 shows that these restaurant checks have a first-digit distribution that deviates dramatically from the predicted Benford distribution, with far fewer first digits of 1 and far more checks beginning with 2, 3, or 4. In fact, the frequencies of all of the first digits demonstrate significant absolute deviations from Benford's distribution, and the overall mean absolute deviation is a statistically significant 5.63 percent, well above our 1-percent threshold. However, once again it is important to think about the source of the data, as we did in the case of the upscale hotel guest folios. At this restaurant chain, it is not surprising that relatively few checks begin with the digit 1 (that is, between $10 and

Exhibit 8

Second-digit frequency distribution for chain restaurant checks vs. Benford's Law



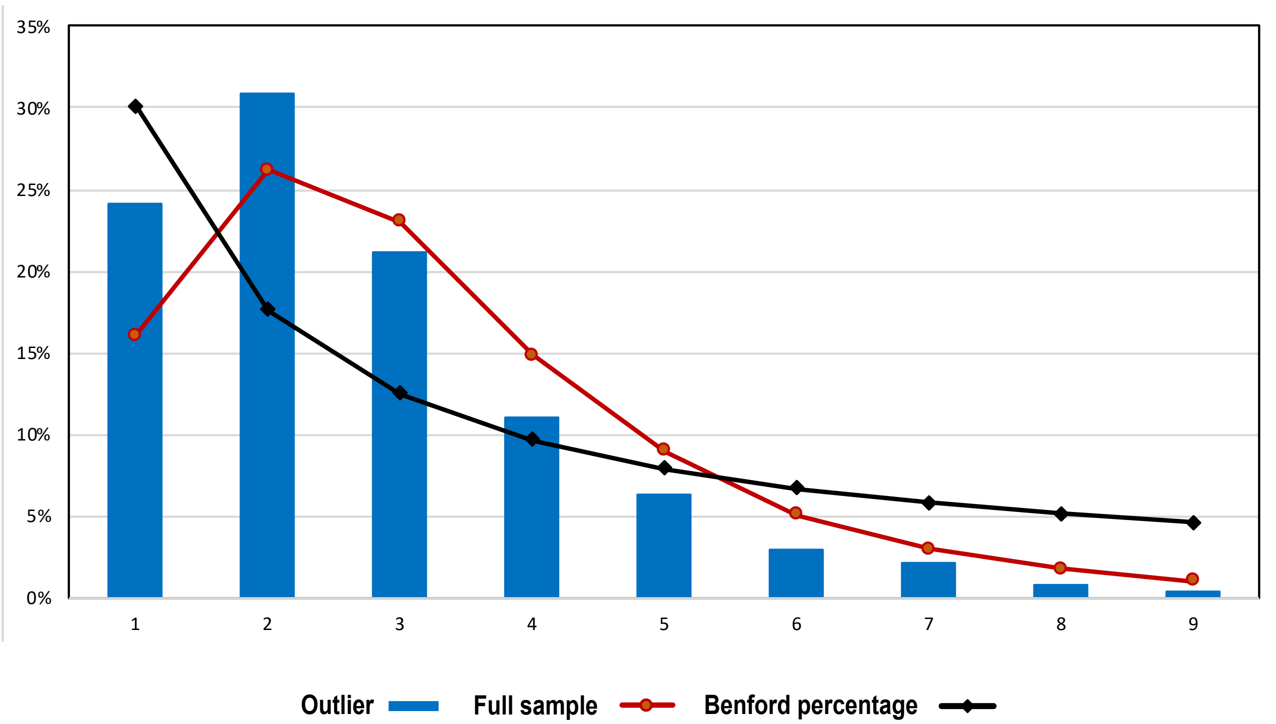
	Second Digit Summary									
	0	1	2	3	4	5	6	7	8	9
Digit counts	29,841	25,094	24,566	26,946	23,912	26,804	22,768	22,311	23,609	21,124
Sample percentage	12.08%	10.16%	9.95%	10.91%	9.68%	10.85%	9.22%	9.03%	9.56%	8.55%
Benford predicted percentage	11.97%	11.39%	10.88%	10.43%	10.03%	9.67%	9.34%	9.04%	8.76%	8.50%
Absolute deviation	0.11%	***1.23%***	***0.94%***	**0.48%**	***0.35%***	***1.19%***	*0.12%*	0.01%	***0.80%***	0.05%
Mean absolute deviation	***0.53%***									

$19.99 or $100 and $199.99) and that more begin with 2, 3, or 4 (mostly checks in the $20 to $49.99 range, with some between $200 and $499.99), as compared to the Benford prediction.[6] Thus, analysis of the second-digit distribution is again likely to be more appropriate here. Exhibit 8 shows that the second-digit distribution conforms to the Benford prediction reasonably closely, with a mean absolute deviation of only 0.53 percent, well below our practical threshold of 1 percent. Thus, despite the deviation of the first digits, this analysis suggests that these numbers are not suspicious overall.

[6] No checks reached or exceeded $1,000.

Exhibit 9

First-digit frequency distribution for chain restaurant checks: Largest outlier vs. full sample distribution and Benford's Law



In a large dataset like that of this franchisee, it would also be worthwhile to repeat the analysis at the individual restaurant level. Such an analysis would reveal whether there are any large outliers relative to the overall dataset, based on an examination of the mean absolute deviation relative to the full-sample distribution (and relative to Benford's Law). Exhibit 9 presents the first-digit analysis for the biggest outlier restaurant in the chain restaurant dataset relative to both Benford's Law and the sample average. This analysis suggests that something unusual is occurring at this particular restaurant, which may be worth further investigation.[7] A similar approach could be used to compare data from a single month to the historical average.

---

[7] The authors are working with anonymized historical data for this example and thus are not able to further investigate this outlier. It is presented here purely as an example of how larger datasets can be subdivided for more detailed analysis.

Exhibit 10

Potential uses of Benford digit analysis

When Benford analysis is likely to be useful	When Benford analysis is not likely to be useful
Numbers that result from the combination of other numbers. *Examples*: restaurant checks (quantity × price); hotel bills (sum of charges); accounts receivable (number sold × price); accounts payable (number bought × price)	Numbers that are assigned or chosen by humans. *Examples*: prices; invoice numbers; zip codes; ATM withdrawals
Large datasets: the more numbers the better, no need to choose a sample to analyze.	Small datasets: fewer than 100 observations will be very noisy.
Datasets in which ranges of values are possible.	Datasets in which there is a built-in maximum or minimum or many transactions of identical size by contract. *Example*: contractual payments that do not vary over time
	Where no transaction is recorded. *Examples*: thefts, kickbacks

Practical Applications

The Benford digit analysis tool provides an easy way for managers to quickly assess whether there are any suspicious patterns in their financial data. Exhibit 10 summarizes when Benford digit analysis might be useful. For example, a hotel or restaurant manager could use the tool to check for suspicious patterns in charges on a daily, weekly, or monthly basis. For financial data that do not fit the Benford first-digit distribution for well-known, logical reasons (such as the chain-restaurant checks found in dataset 3), second-digit analysis is a more useful technique. Managers can also develop their own Benford benchmark range over time. For example, if a financial series repeatedly has a first-digit mean absolute deviation between 5 and 6 percent, a spike to 8 percent would be a signal to investigate further.

We again emphasize an important caveat. When a series of financial data deviates strongly from the predicted Benford distribution, it does not necessarily mean financial fraud has occurred (as demonstrated in datasets 2 and 3). A high (or higher than usual) mean absolute deviation should be viewed as only a signal that this data series merits more checking—for example, drilling down into why there are so many more (or fewer) observations of particular first digits. In some cases (especially in the first-digit distribution), there may be clear business reasons why the distribution differs, such as the relatively few records starting with 1 in the chain restaurant or upscale hotel data we examined. Nevertheless, a seemingly logical explanation may be less than the whole story, as in the example of reimbursement fraud being signaled by an excess of first-digit 1s. ■

Benford Digit Analysis Tool

The Benford digit analysis tool provides a quick and easy way to analyze how well any set of data conforms to the Benford Law's first-digit and second-digit distributions. The tool is available as an Excel file that is downloadable for free from the Center for Hospitality Research Reports website. To use the tool:

Open the workbook and go to the first spreadsheet, labeled "Input." *

Hit "Clear" button to clear input, then copy in a column of data of any size (up to the limits of Excel).

Hit "Submit" button.

Open spreadsheets labeled "Tables" and "Graphs" to see results of Benford digit analysis.

* If "Protected View" warning appears at the top, hit "Enable Editing" box, and if "Security Warning" appears at the top, hit "Enable Content" box. These messages appear automatically because the spreadsheets contain programming to perform the calculations and create tables and graphs automatically.

CHR Advisory Board

Cornell Hospitality Report
Vol. 18, No. 10 (December 2018)

Cornell Hospitality Reports are produced for the benefit of the hospitality and service industries by The Center for Hospitality Research at Cornell University

Glenn Withiam, Editor
Christopher K. Anderson, Academic Director
Carol Leitch-Zhe, Program Manager
Kate Walsh, Dean, E.M. Statler Professor, School of Hotel Administration

The Center for Hospitality Research
School of Hotel Adminstration
Cornell SC Johnson College of Business
Cornell University
Statler Hall
Ithaca, NY 14853

607-254-4505
chr.cornell.edu