Contributions to the Testing of Benford's Law

CONTRIBUTIONS TO THE TESTING OF BENFORD'S LAW

By Amanda BOWMAN, B.Sc.

A Thesis Submitted to the School of Graduate Studies in the Partial Fulfillment of the Requirements for the Degree of Master of Science

McMaster University © Copyright by Amanda BOWMAN 2016

McMaster University

Master of Science (2016)

Hamilton, Ontario (Department of Mathematics and Statistics)

TITLE: Contributions to the Testing of Benford's Law

AUTHOR: Amanda BOWMAN B.Sc. (University of Guelph)

SUPERVISOR: Dr. Fred M. HOPPE

NUMBER OF PAGES: x, 68

Abstract

Benford's Law is a statistical phenomenon stating that the distribution of leading digits in a set of naturally occurring numbers follows a logarithmic trend, where the distribution of the first digit is $P(D_1 = d_1) = \log_{10}(1 + 1/d_1)$, $d_1 \in \{1, 2, ..., 9\}$. While most commonly used for fraud detection in a variety of areas, including accounting, taxation, and elections, recent work has examined applications within multiple choice testing. Building upon this, we look at test bank data from mathematics and statistics textbooks, and apply three commonly used conformity tests: Pearson's χ^2, MAD, and SSD, and two simultaneous confidence intervals. From there, we run simulation studies to determine the coverage of each, and propose a new conformity test using linear regression with the inverse of the Benford probability function. Our analysis reveals that the inverse regression model is an improvement upon the χ^2 goodness of fit test and the regression model that was previously proposed in 2006 by A.D. Saville; however, still presents some asymptotic issues at large sample sizes. The proposed method is compared to the previously utilized tests through numerical examples.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Fred Hoppe, for his guidance throughout my thesis work. His continual support, encouragement, and expertise, in addition to the hours he spent working with me and the many stimulating discussions, made this thesis possible. It was a wonderful experience to work under his supervision.

I would also like to sincerely thank Dr. Alex Rosa and Dr. Franya Franek for being on my thesis committee.

Finally, I would like to thank my family and friends for all of their support and encouragement throughout my degree.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 A Brief History of Benford's Law

An intriguing statistical phenomenon, Benford's Law, is contrary to one's initial assumption that the leading significant digits of numbers in real-life datasets should be uniformly distributed, and instead states that they follow a logarithmic distribution. Although named after physicist Frank Benford, the first-digit law, referring to the leftmost digit in a number, was originally observed by the mathematician and astronomer Simon Newcomb in 1881 [11]. Newcomb realized that the pages at the start of his logarithmic book wore out much quicker than those later in the book, and therefore, numbers with a smaller leading digit appear more often [11]. He also noted the distribution of the second leading digits.

This effect was rediscovered in 1938 by Benford, and applied to 20 real datasets to develop an empirical result in an effort to prove the validity of the law without using a theoretical approach [2]. The 20 datasets were chosen from a variety of sources, from the surface areas of 335 rivers to American League baseball statistics to numbers appearing in Reader's Digest articles, with an effort to obtain a diverse collection. Without setting strict limits or criteria, the data collected ranged from 91 to 5000 observations, with a combined total of 20,229 values [2]. While some of the datasets examined did not conform to the first-digit law, the combined average was very close to the expected proportions and Benford showed that large datasets approximately conform to the logarithmic probabilities. The distribution for the first, second, and first two leading digits can be expressed in the following forms:

$$P(D_1 = d_1) = \log_{10}\left(1 + \frac{1}{d_1}\right) \qquad d_1 \in \{1, 2, ..., 9\} \tag{1.1}$$

$$P(D_2 = d_2) = \sum_{d_1=1}^{9} \log_{10}\left(1 + \frac{1}{d_1 d_2}\right) \qquad d_2 \in \{0, 1, ..., 9\} \tag{1.2}$$

$$P(D_1 D_2 = d_1 d_2) = \log_{10}\left(1 + \frac{1}{d_1 d_2}\right) \qquad d_1 d_2 \in \{10, 11, ..., 99\} \tag{1.3}$$

showing that the likelihood of, for example, a first digit being 1 is approximately 30.1% but only about 4.6% for it being a 9. The proportions for the first, second, and third leading digit are provided in Table 1.1. It can also be noted that the distribution of the digits becomes more uniform for the later digit positions: for example, by the third leading digit, the proportions only range from approximately 10.178% to 9.827%.

Neither Newcomb nor Benford provided any theoretical basis to explain or support

TABLE 1.1: Benford's Law proportions for the first, second, and third leading digits

Digit	First	Second	Third
0	-	0.11968	0.10178
1	0.30103	0.11389	0.10138
2	0.17609	0.10882	0.10097
3	0.12494	0.10433	0.10057
4	0.09691	0.10031	0.10018
5	0.07918	0.09668	0.09979
6	0.06695	0.09337	0.09940
7	0.05799	0.09035	0.09902
8	0.05115	0.08757	0.09864
9	0.04576	0.08500	0.09827

Benford's Law, and while Benford suggested that {1,2,3,..} is not the natural number scale, rather that nature counts $e^0, e^x, e^{2x}, ...$ since it appears that many natural functions are of the logarithmic form in base e [2], an analytical approach was not developed until Theodore Hill in 1995 [7]. In addition, Hill provided a generalization of (1.1) and (1.3), so that they could be extended to find the expected frequency of any combination of leading digits. The expression for this is as follows:

$$P(D_1 = d_1, D_2 = d_2, ..., D_k = d_k) = \log_{10}\left(1 + \frac{1}{\sum_{i=1}^{k} d_i \times 10^{k-i}}\right) \tag{1.4}$$

where $d_1 \in \{1, 2, ..., 9\}$ and $d_j \in \{0, 1, 2, ..., 9\}$ for $j \in \{2, ..., k\}$ for any positive integer k [7]. A 1976 article by Ralph Raimi [16] gave a thorough review of the proposed explanations of Benford's Law at that time, explaining hypotheses and results while omitting most proofs. While some believed that the phenomenon was the natural result of the number system we use [6, 20], the basis of this came from the idea that there is a natural way to calculate the "density" for the set of values beginning with a integer on the positive portion of the real number line that yields $\log(d_1 + 1)$. Although this result can be found through certain summability methods, it was stated without observed facts or supportive justification. Raimi supported the idea that mathematics alone cannot account for Benford's Law [16]. Additionally, he took issue with some of the proposed properties of the law, such as scale invariance as proposed by Pinkham [14], and the need for widely spread data [16].

Since Raimi's article, interest in Benford's Law greatly increased, though Hill's explanation, in which he used the assumptions of both scale and base invariance, is still seen as one of the most convincing arguments. This created a theoretical basis for the law using probability theory. In addition, Hill was able to show that, while not every dataset conforms to the law, as seen in Benford's research, a combination of random samples from a random selection of distributions do [7]. As Hill developed theoretical support for the base and scale invariance assumptions, other properties of datasets that follow, or would be expected to follow, Benford's Law were found, as will be discussed in the subsequent sections.

1.2 Properties

The assumption of base invariance described in Hill's work states that datasets that conform to Benford's Law will continue to do so if the base used is changed from base 10 to, say, base 8 or base 20. Hill defined base invariance as a probability measure P on $(\mathbb{R}^+, \mathscr{M})$ where $P(S) = P(S^{1/n})$ for all positive integers n and all $S \in \mathscr{M}$ [7] where

$\mathscr{M}$ is the decimal mantissa σ-algebra, which is a subfield of the Borels, so that:

$$S \in \mathscr{M} \Leftrightarrow S = \bigcup_{n=-\infty}^{\infty} B^n$$

for some Borel $B \subseteq [1, 10)$

The mantissa σ-algebra $\mathscr{M}$ has the following properties: every non-empty set in $\mathscr{M}$ is infinite with 0 and $+\infty$ having accumulations of points; $\mathscr{M}$ is closed under scalar multiplication ($s > 0$, $S \in \mathscr{M} \rightarrow sS \in \mathscr{M}$) and under integral roots ($n \in \mathbb{N}$, $S \in \mathscr{M} \rightarrow S^{1/n} \in \mathscr{M}$) but not under powers; $\mathscr{M}$ is self-similar, so if $S \in \mathscr{M}$ then $10^n S = S$ [7]. Hence, the probability measure for any set of real numbers in $(\mathbb{R}^+, \mathscr{M})$ should be the same for any base. Therefore, in base 10, every set of real numbers $S \in \mathscr{M}$ is identical to the set of real numbers $S^{1/2}$ in base 100 in $\mathscr{M}$.
In a similar manner, a probability measure P on $(\mathbb{R}^+, \mathscr{M})$ is said to be scale invariant if $P(S) = P(sS)$ for all $s > 0$ and all $S \in \mathscr{M}$ [7]. Therefore, multiplying a Benford set by a positive value will still produce a Benford set. For instance, converting company profits from Canadian dollars to Euros will not impact conformity.
In addition, the underlying logarithmic basis of the Law indicates that conformity requires the mantissas of the log of the dataset to be uniformly distributed, where the mantissa is the decimal portion of the log [13].
There are criteria for datasets that are expected to follow Benford's Law. First, one should not test the first two digits on sample sizes less than 300, and good conformity should not be expected for datasets smaller than 1000 observations, due to the commonly used χ^2 goodness of fit test which requires an expected cell count of 5 [13]. For the first two digits, 99 has an expected count of 4.36, which is generally considered close enough in practical settings. Moreover, there should not be a strict

minimum or maximum, other than if numbers are constrained to be positive; observations should not be values assigned as labels or for identification; there should be more small records than large, meaning the median should be greater than the mean and values should not be clustered tightly around an average [13]. With these criteria in mind, the next section will examine a sample of applications of Benford's Law.

1.3 Current Work

To date, Benford's Law has found applications in a diverse range of research areas, from forensic accounting to election data to fraud detection in scientific research. In addition, there are many mathematical series and sequences that have been found to follow the Law, including the Fibonacci sequence and most geometric series. In this section, we look at several detailed examples to illustrate some of the widespread applications.

1.3.1 Fraud Testing in Accounting

It has long been known that humans are not able to create sets of random numbers manually; analogously, it is also difficult to produce a set of numbers that follows Benford's Law. This allows for conformity tests to be used as a method of fraud detection, or at least to signify financial accounts that need to be examined to a more in-depth level. Mark Nigrini, a leading expert in the field, was one of the first to propose the use of Benford's Law as a testing tool for fraud in accounting data, and it has now become commonplace in digit analysis. In cases where there are significant deviations from the expected proportions, the likelihood of fraud having occurred is much greater.

The first study that utilized Benford's Law for such an application was by Charles Carslaw in 1988, when he conducted a second digit analysis on the profits of a sample of New Zealand firms. His results showed an excess of the second digit 0 and a lack of 9's, suggesting that managers round up the profit values to make them appear more impressive, showing goal oriented behaviour [4]. This is similar to psychological methods used when pricing goods, where a value of $1.99 appears significantly lower or more appealing to consumers than $2.00. This is thought to be due to the fact that humans place more emphasis on earlier leading digits [3].

Nigrini (2012) provides numerous examples of the use of Benford's Law in forensic accounting. In the case of *State of Arizona vs Wayne James Nelson*, Nelson, who was a manager in the office of the Arizona State Treasurer, was found guilty in a $2 million defraud case. The 23 fraudulent checks were all in amounts under $100,000, where values over $100,000 would likely have received more review or have required someone else's signature, and there were no round numbers or duplicates [13]. The amounts started small and increased until over 90% of the values began with a 7, 8, or 9, and did not conform to Benford's Law in the first or first-two digits [13]. In addition, 87, 88, 93, and 96 were all used twice as the first two digits, and 16, 67, and 83 reappeared as the final two digits, all of which would prove to be suspicious to an auditor.

1.3.2 Test Bank Questions

A novel application for Benford's Law was investigated in a 2015 paper by Slepkov et al., who tested if knowledge of the law could give an advantage in physics numerical multiple choice tests. They hypothesized that the correct answers should follow Benford's Law, while the distractors, if chosen at random, should not [19]. Three commonly used undergraduate physics textbooks were chosen, and end of chapter problems were recorded by hand, excluding unphysical numbers, numbers too narrowly confined in domain, all unit-less values, percentages, degrees, answers of exactly 0, and any non-numerical answers [19]. Using three conformity tests (MAD, SSD, and Pearson's χ^2 goodness of fit test), all three textbooks showed compliance with the Benford proportions. They then simulated 5,000 mock multiple choice questions, where the correct answers followed Benford's Law and the distractors were uniformly distributed. For 3-, 4-, and 5-option tests, the Benford approach of selecting answers, where one selects the answer with the lowest leading digit, proved to have an advantage with scores of 51%, 41%, and 33% respectively, compared to 33%, 25%, and 20% for random guessing [19]. Slepkov et al. then applied this to an actual physic test bank, and for the four option questions, a score of 24.6% was achieved using a "Benford attack", which is no better than randomly guessing [19]. This should not come as a surprise as distractors are not determined by random selection, and also followed Benford's Law, meaning that they are secure against a Benford approach.

Following the above, Hoppe developed a closed form solution for the probability of a correct answer when using a Benford approach, for test banks where the correct answers follow Benford's Law while the distractors follow a uniform distribution [8]. Recently, Nigrini examined test banks in accounting textbooks and the effect of the excessive use of large, rounded numbers, which in real data should be a sign for concern [12]. Results showed that the first digits of the textbook data follow Benford's Law but the second digits do not. In addition, there was an excessive amount of the second digit 0, where 80% of the numbers were multiples of 100 and 70% of 1000 [12]. While Nigrini's article does not look at Benford's Law in conjunction with test bank data as a method for improving test scores, it is posed as a future topic for research. Rather, it looks at the impact of the data on the views of accounting students and whether they will view the numbers commonly seen in class as unrealistic in a real forensic accounting setting. Nigrini states that while the first digits may conform, the subsequent digits can show significant deviations from Benford's Law and should be examined [12]. In addition, it should be emphasized to students that the examples seen in class and within textbooks are used for simplicity and should be considered suspicious in an analysis of real world data.

1.4 Motivation

Benford's Law is a complex problem, and while there are many explanations and hypotheses, none satisfactorily explain why such a wide variety of real life datasets have this distribution. Also unexplained is why a combination of data from multiple contexts, such as those seen in a test bank, would also conform to this law.

While the present work does not attempt to provide an explanation for the above questions, we will carry out an analysis of a collection of mathematics and statistics multiple choice test bank questions. Using this data and through simulations, we will examine some of the currently used tests for conformity and propose a new method utilizing linear regression.

This thesis is organized as follows. In Chapter 2, we describe our method of data collection and provide the methodology for the commonly used statistical techniques for testing for conformity with Benford's Law. We then apply these tests to our multiple choice dataset and provide the results in Chapter 3. This is followed by our proposed method of using linear regression as a test for conformity in Chapter 4, then present our conclusions in Chapter 5.

Chapter 2

Methodology

2.1 Data Collection

In order to collect a large sample of multiple choice questions, textbook banks were chosen based on both their availability within the McMaster Mathematics and Statistics department and publicly online. Nine textbooks were used in addition to a collection of midterm exams from Dr. George Wesolowsky, a professor emeritus at the DeGroote School of Business. Table 2.1 provides an index of the utilized sources and the number of rejected and accepted questions from each. Data was manually recorded after going through the entirety of each test bank, while adhering to a set of rejection criteria. Questions were rejected for having non-numerical answers, and for having options that each contained multiple numbers. Questions that reappeared in the test bank were only recorded once, and answers of exactly zero were excluded since there is no leading digit. Answers without units or context were also rejected;

however, differing from the method used by Slepkov et al. [19], percentages and proportions were not. It has been shown that numbers bounded by 0 and 1 satisfy Benford's Law and therefore they were not rejected here [1]. Overall, 13.6% were omitted due to a lack of units, 0.8% were duplicate questions, and approximately 68% of all questions were excluded due to being non-numerical, having multiple numbers per option, or having a value of exactly zero. This left the remaining 17.6%, which were accepted, and which were composed of 3-, 4-, 5-, and 6-option multiple choice questions, giving an overall sample size of 3683 observations.

2.2 Statistical Tests for Conformity

Testing a dataset's goodness of fit to Benford's Law can be accomplished in numerous ways, and a variety of tests are available for this purpose. In this section, we examine three test statistics that are commonly applied to assess conformity with Benford's Law.

2.2.1 Pearson's χ^2 Goodness of Fit Test

The most frequently used statistic to determine compliance with Benford's Law is the χ^2 goodness of fit test, which is calculated as follows:

$$\chi^2 = \sum_{i=1}^{K} \frac{(AC - EC)^2}{EC}$$

TABLE 2.1: Summary of accepted and rejected test bank question

Test Bank	Rejected due to units	Rejected due to non-numerical/zero/ multiple answers	Rejected due to repeat	Accepted	Total
Stewart Calculus: Early Transcendentals, 8th edition by Stewart (Cengage Learning, 2015)	357	696	24	90	1,167
Statistical Reasoning for Everyday Life, 1st edition by Bennett, Briggs, and Triola (Addison Wesley, 2000)	25	347	6	94	472
Elementary Statistics, 10th edition by Triola (Pearson, 2005)	60	181	0	154	395
The Basic Practic of Statistics, 7th edition by Moore, Notz, and Fligner (MacMillian Learning, 2015)	17	445	4	81	547
Probability and Statistics for Engineering and the Sciences, 8th edition by Devore (Duxbury Press, 2011)	17	256	0	13	286
Introduction to the Practice of Statistics, 2nd edition by G. McCabe and L. McCabe (W.H. Freeman, 1993)	1	42	0	9	52
The Basic Practice of Statistics, 3rd edition by Moore, Notz, and Fligner (W.H. Freeman, 2004)	41	383	2	162	588
Finite Mathematics, 3rd edition by Warner and Costenoble (Thomson Learning, 2004)	29	172	0	36	237
Introduction to Probability and Statistics, 14th edition by Mendenhall, R. Beaver, and B. Beaver (Cengage Learning, 2012)	74	499	3	139	715
Dr. Wesolowsky's Midterm Test Bank (McMaster, 1998-2006)	54	353	1	97	505
TOTAL	675	3,374	40	875	4,964

where AC and EC are the actual and expected counts of each leading digit respectively, and K is the number of possible leading digits, meaning if we are testing the first leading digits K=9 and if testing the first two then $K = 90$. The calculated statistic is then compared to a critical χ^2 value with $K-1$ degrees of freedom to test the null hypothesis that the data conforms to Benford's Law.

However, issues with the χ^2 statistic present themselves with large sample sizes (those approximately greater than 5000) [13]. The test statistic has an excess of power at close alternatives, and therefore small deviations from the expected values will cause a result of nonconformity that would not be an issue at a smaller sample size. This means a large dataset can be rejected, while a smaller dataset with larger deviations from the Benford proportions will be accepted as following the law.

2.2.2 Mean Absolute Deviation

An alternative test for conformity was proposed by Nigrini to negate the issues seen with the χ^2 goodness of fit test. The mean absolute deviation (MAD) test does not include the number of observations in its calculation and therefore, he states that it is not affected by sample size [13]. The formula for the test is as follows:

$$MAD = \frac{\sum_{i=1}^{K} |AP - EP|}{K}$$

where AP and EP are the actual and expected proportions of each leading digit, and K is the number of bins, again being 9 for the first digit and 90 for the first two.

To determine the ranges of MAD for conformity with Benford's Law, Nigrini empirically derived critical values based on personal experience and testing done on numerous datasets [13]. The ranges proposed for the first leading digits are 0.000-0.006 for close conformity, 0.006-0.012 for acceptable conformity, 0.012-0.015 for marginally acceptable conformity, and values greater than 0.015 show non-conformity. For the first two leading digits, these ranges become 0.0000-0.0012, 0.0012-0.0018, 0.0018-0.0022, and greater than 0.0022 respectively.

2.2.3 Sum of Squares Difference

While not as commonly utilized as Pearson's χ^2 goodness of fit test or MAD, sum of squares deviation (SSD) is used as a comparison measure when examining Benford's Law. Proposed by Kossovsky, SSD is a measure of the distance from the logarithm and not a test for conformity [9]. The formula is given by:

$$SSD = \sum_{i=1}^{K} (AP - EP)^2 \times 10^4$$

where again AP and EP are the actual and expected proportions of each leading digit respectively, and K is the number of possible leading digits.

As sample size is not included in the calculation, statistical theory cannot be used to identify critical values, and therefore, as with MAD, ranges for compliance were empirically derived. Kossovsky states that, for first digits, SSD values that are less than 2 are perfect Benford, those falling within [2, 25) are acceptably close, values between [25, 100] are marginally Benford, and values greater than 100 are non-Benford. For the first two leading digits these ranges become less than 2, [2, 10), [10, 50], and greater than 50 respectively. However, he also states that an SSD value should be subjectively judged to determine the distance from the logarithmic expectation [9].

2.3 Simultaneous Confidence Intervals

Since confidence intervals can provide more information about deviations from the Benford proportions than conformity tests, due to their ability to determine the values that are outside the confidence interval, we examined two simultaneous confidence intervals in order to take the multinomial proportions into account. The two simultaneous confidence intervals chosen were Goodman and Sison & Glaz, based on the examinations by Lesperance and her student Wong, for testing the first and first two digits respectively [10, 21]. After testing multiple simultaneous confidence intervals for multinomial proportions, the following two were recommended for assessing Benford's Law.

2.3.1 Goodman

The Goodman simultaneous confidence intervals modify the Quesenberry and Hurst calculations to create less conservative, and therefore shorter, intervals [5, 15]. Letting $n_1, n_2, ..., n_k$ be the observed cell frequencies from a multinomial distribution of size N, and $p_1, p_2, ..., p_k$ be the corresponding probabilities that an observation will fall into the i^{th} cell, the formula is as follows:

$$p_i = \frac{B + 2n_i \pm \sqrt{B[B + 4n_i(N - n_i)/N]}}{2(N + B)} \quad i = 1, 2, ..., k$$

where $B = \chi^2_{\alpha/k,1}$, the upper α/K quantile of the chi-square distribution with 1 degree of freedom, and k must be greater than 2. It should be noted that $p_i \geq 0$, $\sum_{i=1}^{k} p_i = 1$, and $\sum_{i=1}^{k} n_i = N$.

2.3.2 Sison & Glaz

The method of Sison and Glaz was the preferred choice by Lesperance and Wong; however, it has no closed form and therefore must be calculated computationally, so it should only be utilized if the computational power is available [10, 18, 21]. Let V_i be independent Poisson random variables with mean n_i, and let Y_i be its truncated form to $[n_i - \tau, n_i + \tau]$ for some constant τ. For a sample of N observations from a multinomial distribution, let $n_1^*, n_2^*, ..., n_k^*$ be the observed cell frequencies with probabilities $\hat{p}_1, \hat{p}_2, ..., \hat{p}_k$. The central and factorial moments of Y_i are denoted as:

$$\mu_i = E[Y_i]$$

$$\sigma_i^2 = Var[Y_i]$$

$$\mu_{(r)} = E[Y_i(Y_i - 1)...(Y_i - r + 1)]$$

$$\mu_{r,i} = E[Y_i - \mu_i]^r$$

In addition, we define the following:

$$\gamma_1 = \frac{\frac{1}{k}\sum_{i=1}^{k} \mu_{3,i}}{\sqrt{k}(\sum_{i=1}^{k} \sigma_i^2)^{3/2}}$$

$$\gamma_2 = \frac{\frac{1}{k}\sum_{i=1}^{k}\mu_{4,i} - 3\sigma_i^4}{\sqrt{k}(\sum_{i=1}^{k}\sigma_i^2)^2}$$

$$f_e(x) = \left(\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\right)\left\{1 + \frac{\gamma_1}{6}(x^3 - 3x) + \frac{\gamma_2}{24}(x^4 - 6x^2 + 3) + \frac{\gamma_1^2}{72}(x^6 - 15x^4 + 45x^2 - 15)\right\}$$

$$v(\tau) = \frac{n!}{n^n e^{-n}}\left\{\prod_{i=1}^{k} P(n_i - \tau \le V_i \le n_i + \tau)\right\} f_e\left(\frac{N - \sum_{i=1}^{k}\mu_i}{\sqrt{\sum_{i=1}^{k}\sigma^2}}\right)\frac{1}{\sqrt{\sum_{i=1}^{k}\sigma^2}}$$

The Sison and Glaz interval then takes the subsequent form:

$$\left(\hat{p}_i - \frac{\tau}{N} \le p_i \le \hat{p}_i + \frac{\tau + 2\gamma}{N}; i = 1, 2, \ldots k\right)$$

where $\gamma = \frac{(1-\alpha) - v(\tau)}{v(\tau+1) - v(\tau)}$ and τ satisfies $v(\tau) < 1 - \alpha < v(\tau + 1)$.

Chapter 3

Analysis

3.1 Histograms of Data

The collected test bank data were analyzed looking at first digits, first two digits, and second digits. Histograms were used to visualize the data, as seen in Figure 3.1, where the bars show the observed digit proportions for the subsets of the data and the continuous curve passes through the Benford's Law proportions. In all three cases, for the correct answers, distractors, and combined dataset, the observed first digit proportions are lower than the expected Benford values for the digit 1 and slightly higher than expected for the digits 7 through 9. In addition, the three plots of the first two digits show peaks on the intervals of 10, while the plots of the distractors and full data also show notable peaks at 25 and 75.

Due to the peaks observed at the multiples of 10, the makeup of the dataset was examined and it was noted that a large number of the collected questions had single digit answers, which would lead to values where the second digit is 0. A subset of the data was taken, where questions with two or more single digit answers were removed.

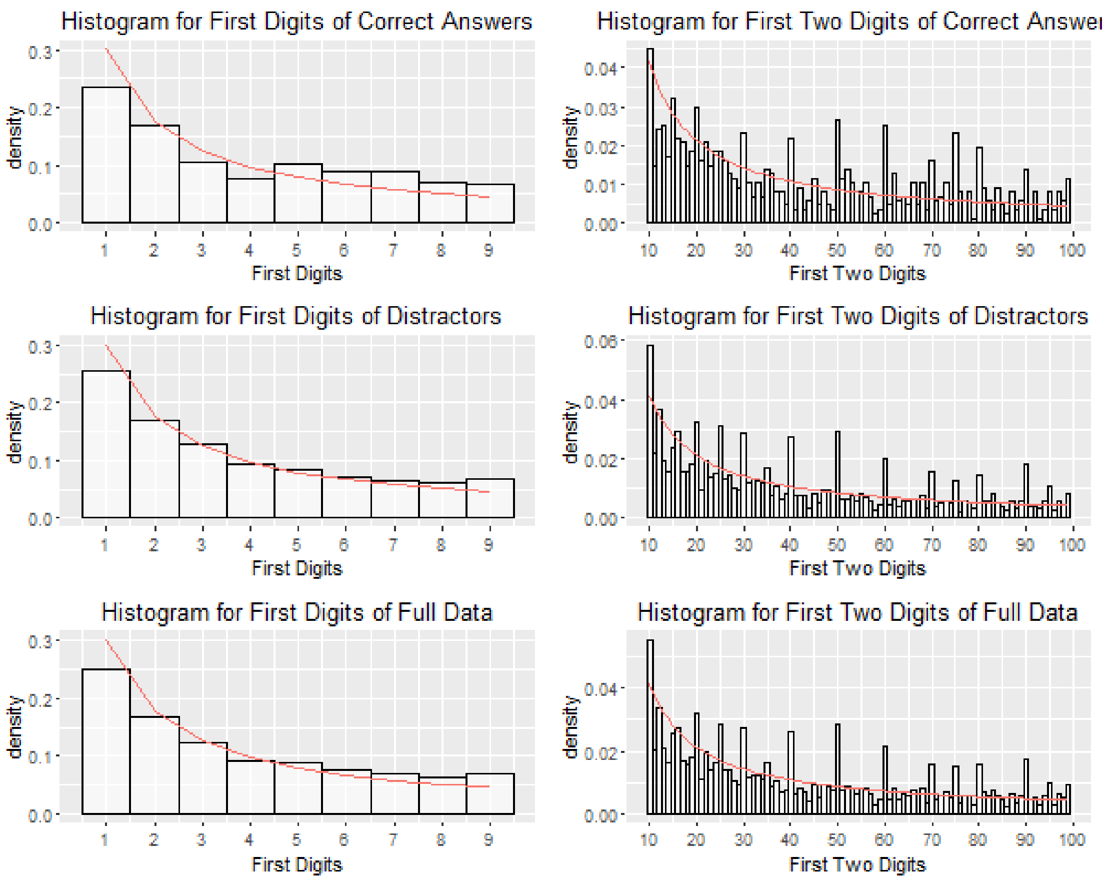

FIGURE 3.1: The first and first two digits of collected test bank data, with the true Benford proportions indicated with a red line

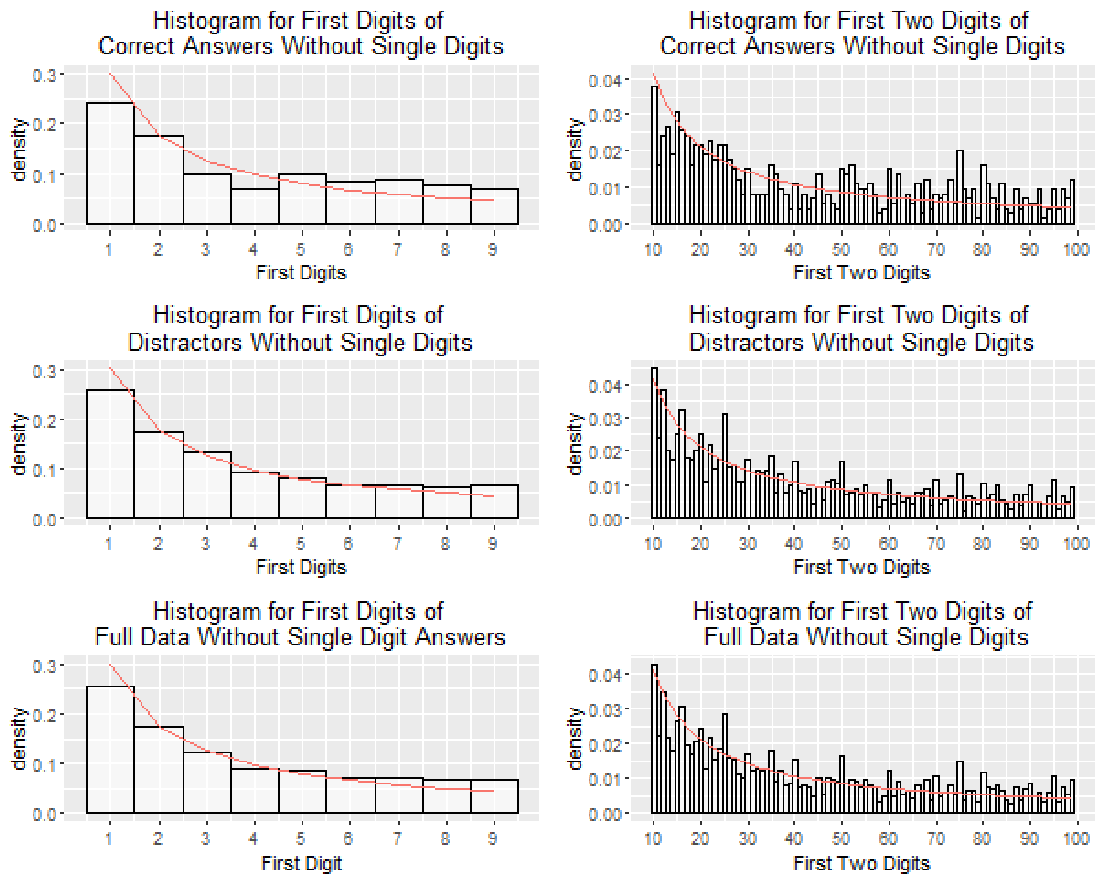

FIGURE 3.2: The first and first two digits of collected test bank data without single digit questions, with the true Benford proportions indicated with a red line

This data was plotted in Figure 3.2.

While the first digit distributions did not appear to change significantly with the removal of the single digit answers, the histograms of the first-two digits appear much closer to the true Benford proportions. It can still be noted that there are peaks at 75 and 50 for all three graphs, and at 25 for both the distractors and the full dataset.

The second digits were plotted for the correct answers, distractors, and the full data all with the single digit answers removed, as shown in Figure 3.3. The graphs show a large observed proportion of 0's and 5's, even when the single digit answers are

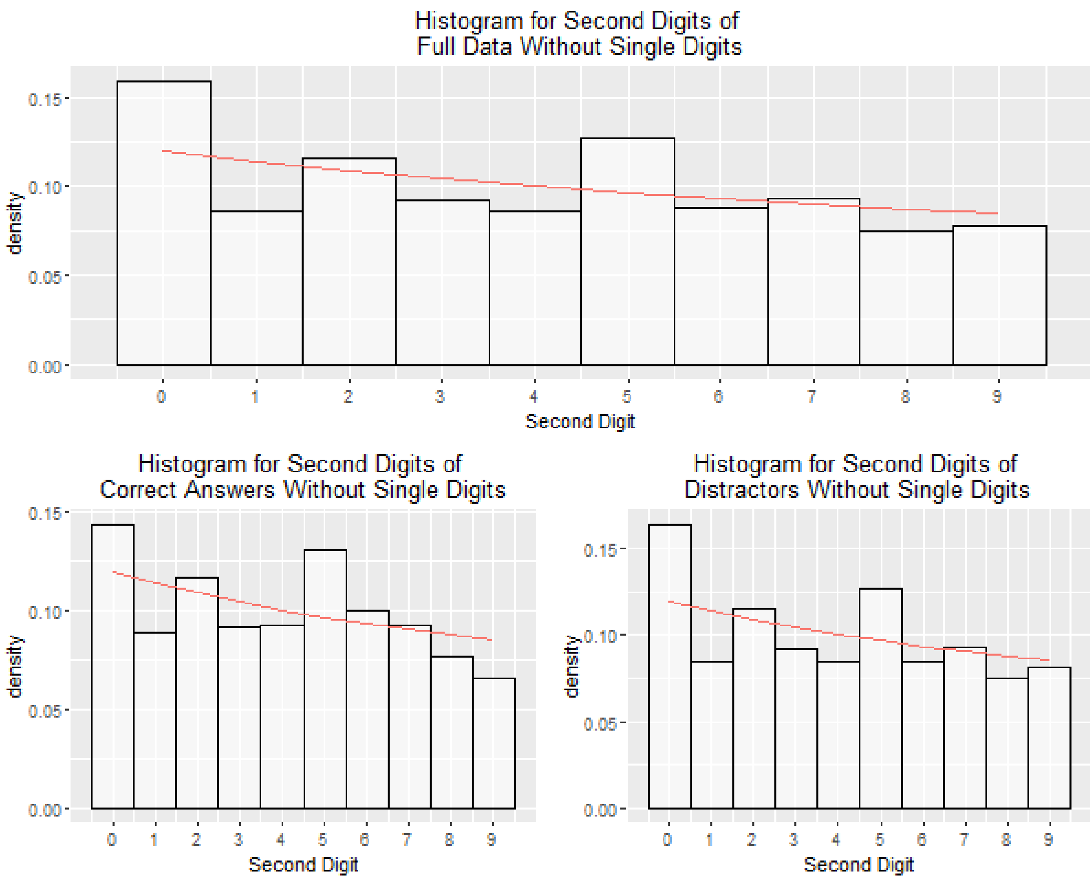

FIGURE 3.3: The second digits of collected test bank data without the single digit answers, with the true Benford proportions indicated with a red line

removed, which could be evidence of both rounding error and of the psychological preference for numbers ending in 0 and 5. In addition, the correct answers had significantly larger deviations in the proportions of the second digits, with 0 and 9 having smaller frequencies and 6 having a higher proportion than in the distractors or full dataset.

3.2 Statistical Tests for Conformity

Testing for conformity with Benford's Law for the first and first two digits of subsets of the full collected test bank data was completed using three tests: MAD (mean absolute deviation), Pearson's χ^2 goodness-of-fit, and SSD (sum of squares deviation). Tables 3.1 and 3.2 show the results, with none of the datasets conforming to Benford's Law according to the χ^2 goodness-of-fit test, although as previously stated, this test statistic is known to be overly sensitive to larger datasets.

TABLE 3.1: First digit tests for conformity with Benford's Law, applied to multiple choice test bank datasets

Dataset	MAD	Chi-square p-value	SSD
Correct Answers- Full	0.0260	$1.5470 \cdot 10^{-10}$	82.1983
Distractors- Full	0.0123	$8.6871 \cdot 10^{-11}$	28.8247
Combined- Full	0.0147	$2.5570 \cdot 10^{-19}$	36.5517
Correct Answers- Without Single Digit	0.0250	$1.3815 \cdot 10^{-08}$	75.0003
Distractors- Without Single Digits	0.0109	$1.7719 \cdot 10^{-07}$	25.0996
Combined- Without Single Digits	0.0130	$6.4233 \cdot 10^{-14}$	31.2360

In all cases, the datasets where single digit answers were removed had smaller test statistics than the corresponding full data. Using MAD for the first digits, the distractors when the single digit answers were removed showed acceptable conformity with Benford's Law, and the distractors for the full dataset and the full set, both with and without the single digit answers removed, all showed marginally acceptable conformity. Both sets of correct answers gave MAD values greater than 0.015, which

TABLE 3.2: First two digit tests for conformity with Benford's Law, applied to multiple choice test bank datasets

Dataset	MAD	Chi-square p-value	SSD
Correct Answers- Full	0.0044	$1.0494 \cdot 10^{-26}$	34.1571
Distractors- Full	0.0040	$9.6947 \cdot 10^{-122}$	34.3143
Combined- Full	0.0038	$5.9955 \cdot 10^{-164}$	31.2472
Correct Answers- Without Single Digit	0.0038	$1.4433 \cdot 10^{-07}$	22.7837
Distractors- Without Single Digits	0.0030	$1.6051 \cdot 10^{-26}$	16.0344
Combined- Without Single Digits	0.0028	$7.7156 \cdot 10^{-40}$	14.5272

shows nonconformity. The SSD statistics all gave values within the marginally Benford range, although for the distractors with the single digit answers removed, the SSD value of 25.0996 was only slightly greater than the cut off value of 25 for acceptable conformity.

Table 3.2 looks at the calculated conformity values for the first two digits, and as previously stated, the χ^2 test shows that none of the test bank subsets conform to Benford's Law. The MAD values conclude the same results, as the calculated statistics are all greater than the cut off value of 0.0022 for any level of conformity in the first-two digits. The SSD, on the other hand, produced all values between the range of 10 to 50, and therefore states marginal Benford. However, as noted in Section 2.2.3, SSD is a measure of the distance from the logarithm and not a test of conformity, and therefore the cut off values are considered to be rough guidelines [9].

3.2.1 Simultaneous Confidence Intervals

The results from running both the Goodman and Sison & Glaz simultaneous confidence intervals for multinomial proportions on the test bank datasets are provided in Tables 3.3 and 3.4. The tables show the digit proportions that fall outside the lower and upper limits of the calculated simultaneous confidence intervals.

TABLE 3.3: Observed digit proportions outside the simultaneous confidence intervals for testing first digit conformity with Benford's Law

Dataset	Goodman	Sison & Glaz
Correct Answers- Full	1 7 9	1 7
Distractors- Full	1 9	1 9
Combined- Full	1 7 8 9	1 9
Correct Answers- Without Single Digit	1 7 8 9	1
Distractors- Without Single Digits	1 8 9	1 9
Combined- Without Single Digits	1 7 8 9	1 9

The results show more values falling outside of the Goodman confidence intervals than the Sison & Glaz. Moreover, for the first digit analysis, the digit 1 consistently deviates from the expected Benford proportion using both methods. For the first two digits, the correct answers without the single digit options had the fewest deviations; however, it also has the smallest number of observations, and as the sample size increases the confidence intervals narrow. The leading digit 11 is identified to deviate in all cases except for the Goodman intervals for the correct answers without the single digit questions. This can be visually seen in Figures 3.1 and 3.2, where the observed proportion is much lower than the expected Benford line.

TABLE 3.4: Observed digit proportions outside the simultaneous confidence intervals for testing first two digit conformity with Benford's Law

Dataset	Goodman	Sison & Glaz
Correct Answers- Full	11 50 60 70 75 80 90	11 50 60 75
Distractors- Full	10 11 13 14 20 21 25 30 40 50 60 70 75 80 90 95	10 11 13 14 20 21 25 30 40 50 60 70 90
Combined- Full	10 11 13 14 20 21 25 30 40 44 50 60 70 75 80 90 95 99	10 11 13 14 20 21 25 30 40 50 60 70 75 80 90
Correct Answers- Without Single Digit	75 80	11
Distractors- Without Single Digits	11 14 25 50 75 90 95 99	11 13 14 25
Combined- Without Single Digits	11 14 25 50 75 80 95 99	11 13 14 25 75

3.3 Simulations

3.3.1 Simultaneous Confidence Intervals

Simultaneous confidence intervals are utilized when the goal is to obtain a set of k intervals with an overall coverage of $(1-\alpha)\times 100\%$. Often, k single $(1-\alpha)\times 100\%$ binomial confidence intervals are used with multinomial proportions, however the probability that all k intervals simultaneously contain the Benford proportions is not $(1-\alpha)\times 100\%$, rather often closer to $(1-k\alpha)\times 100\%$ [10]. Simultaneous $(1-\alpha)\times 100\%$ confidence intervals are utilized instead to create a set where the probability of the corresponding Benford proportion being contained in each interval is approximately $(1-\alpha)$. Simulations were run in R to identify the exact coverage for a sample size that matched that of our test bank dataset. Using a sample of size 3800 and sampling from a multinomial distribution with the Benford proportions, 10,000 simulations were run, with the coverage of the two simultaneous confidence intervals for the first digits being as follows:

- For Sison and Glaz, at the 95% level, coverages were 94.48% and 94.75% as the simulation was ran twice.
- At the 99% level for Sison and Glaz the coverages were 98.85% and 99.01%.
- For Goodman, the coverage at the 95% level was 95.22%, and was 99.11% at the 99% level.

The coverage for the first two digits was also simulated for Sison and Glaz, producing coverages of 94.62% and 99.01% for the 95% and 99% confidence levels, respectively. The first two digit intervals for Goodman produced coverages of 93.61% and 98.35%. Showing that at a sample size comparable to our dataset, the overall coverage of the intervals is close to the desired $(1-\alpha)\times 100\%$ level; however, the coverage of Sison & Glaz is slightly more accurate with a larger number of bins.

3.3.2 Pearson's χ^2 Goodness of Fit Test Statistic

Simulations were also run to examine the coverages for Pearson's goodness of fit test, again sampling from a multinomial distribution following Benford's Law and using sample sizes equal to that of our own data. The results showed that for samples of size 3800, the coverage of Pearson's χ^2 at the 95% level were 95.06% and 94.99% for the two simulations run, and at the 99% level, the coverages produced were 98.94% and 99.16%.

3.3.3 MAD

Due to the lack of statistical theory for the MAD test statistic and its critical values, the simulations run were more in depth than those in the previous two subsections. Using the same method as for the χ^2 test statistic, where we were sampling from a multinomial distribution with the Benford proportions and using a sample of size 3800 to be comparable to the test bank dataset, results showed that 96.94% of the simulations fell within the close conformity range and 3.06% fell within the acceptable conformity range, whereas none of the samples were considered to be marginally acceptable or to have nonconformity. Although Nigrini states that the MAD statistic ignores sample size since n is not included in its calculation [13], we wished to examine the distribution of the MAD values at various sample sizes when simulating samples from the Benford proportions, seen in Table 3.5. Since only N=10,000 simulations were run due to time constraints, values are rounded to three decimals places, since the accuracy of the fourth decimal value is not known.

TABLE 3.5: Acceptance probabilities for MAD conformity levels simulated from a Benford distribution; N=10,000

	Sample Size				
Conformity Ranges	100	500	1000	5000	10,000
Close Conformity (0.000 to 0.006)	0.000	0.046	0.252	0.995	1.000
Acceptable Conformity (0.006-0.012)	0.024	0.662	0.729	0.005	0.000
Marginally Acceptable Conformity (0.012-0.015)	0.062	0.219	0.018	0.000	0.000
Nonconformity (greater than 0.015)	0.915	0.072	0.001	0.000	0.000

While samples are expected to asymptotically approach the true distribution as sample sizes increase, by 10,000 observations 100% of the samples are within the close conformity range. If we treat MAD as a two-sided hypothesis test, where H_0 is that the sample conforms to Benford's Law and H_1 is that it does not, then the proportion of samples that fall within the nonconformity range is equivalent to α, or the Type I error. Since by samples of size 10,000 the rejection rate is 0%, and since MAD is often used to test samples much larger than this, one might expect an increase in the number of false negatives, or the Type II error, as the two error types are inversely related. In addition, Nigrini states that good conformity should not be expected for samples smaller than 1,000 [13], however for simulations of size 1000, only 25.2% fall within the close conformity range when sampling from Benford. It is worth noting that only 1% are rejected for nonconformity.

To take this further, MAD can be treated as three separate hypothesis tests, where one can test a null hypothesis that the sample has close conformity, has acceptable or better conformity, or conforms within any of the three ranges. This can be written as:

$$P[MAD \leq 0.006]$$

$$P[MAD \leq 0.012]$$

$$P[MAD \leq 0.015]$$

$$P[MAD \geq 0.015]$$

where the $P[MAD \geq 0.015]$ is equal to our α or P[Reject $H_0|H_0$ is true] for testing for any level of conformity. However, when testing if the sample has close conformity, our α level becomes the sum of the other three probabilities. As previously mentioned, as the sample sizes increases, α approaches 0 for all three possible tests, allowing for an increase in the P[Accept $H_0|H_0$ is false]. This may not pose an issue if one is interested in datasets that are approximately but not exactly Benford. However, one thing to note is that, unlike in the framework of statistical hypothesis testing, as the sample size changes, the α value changes rather than the critical values.

To examine this in more depth, simulations were run on samples from a multinomial distribution with proportions that were relatively close, but not exactly equal, to those expected under Benford's Law. The probability set chosen was $\{31.755, 16.11, 11.015, 8.287, 10.163, 6.028, 4.982, 5.037, 6.624\}$, which uses the proportions from a dataset of corporate payments used in Nigrini's 2012 book that contained over 185,000 observations [13]. The MAD of the dataset was 0.0132, which falls into the marginally acceptable range. Results from the simulations are seen in Table 3.6.

TABLE 3.6: Acceptance probabilities for MAD conformity levels simulated from a distribution with proportions {31.755, 16.11, 11.015, 8.287, 10.163, 6.028, 4.982, 5.037, 6.624}; N=10,000

	Sample Size				
Conformity Ranges	100	500	1000	5000	10,000
Close Conformity (0.000 to 0.006)	0.000	0.002	0.001	0.000	0.000
Acceptable Conformity (0.006-0.012)	0.010	0.138	0.218	0.178	0.110
Marginally Acceptable Conformity (0.012-0.015)	0.030	0.400	0.441	0.671	0.827
Nonconformity (greater than 0.015)	0.958	0.588	0.420	0.151	0.063

As before, as sample size increased the majority of the samples fell within the marginally acceptable conformity range since they asymptotically approach the true distribution. For large samples, none of the simulations fell within the close conformity range.

Chapter 4

Linear Regression as a Test of Conformity with Benford's Law

4.1 Linear Regression Using the Inverse of the Benford Probability Function

Given that the Benford probabilities are specified by:

$$p_i = \log_{10}\left(1 + \frac{1}{i}\right) \qquad i = 1, 2, \ldots, 9 \tag{4.1}$$

let X_1, X_2,..., X_9 be the number of observations with each leading digit. Therefore, $X_i \sim$ Binomial(n, p_i), where n is the sample size. Since the X_i's are Binomial(n, p_i), the estimates of the probabilities are $\hat{p}_i \sim \frac{1}{n}$Binomial(n, p_i). We now want to invert (4.1) and solve for i.

$$p_i = \log_{10}\left(1 + \frac{1}{i}\right)$$

$$\begin{aligned} 10^{p_i} &= 1 + \frac{1}{i} \\ 10^{p_i} - 1 &= \frac{1}{i} \\ i &= \frac{1}{10^{p_i} - 1} \end{aligned} \tag{4.2}$$

Here, i is the expected values of the leading digits (integer values from 1 to 9); however, we observe "$\hat{i}$", from now on referred to as U_i. Examining (4.2), we define:

$$\begin{aligned} U_i &= \frac{1}{10^{\hat{p}_i} - 1} \\ &= \frac{1}{10^{\frac{binomial(n,p_i)}{n}} - 1} \end{aligned} \tag{4.3}$$

Given that U_i is a random variable that should approximate i for large n, one would expect that the relationship between the observed and expected values could be utilized to determine whether the observed digits significantly deviate from Benford's Law. Linear regression can be applied to the inverse Benford model, comparing the slope and intercept parameters to the 1:1 line, as a sample with close conformity to the Benford proportions would yield almost perfect correlation. Therefore, the regression line takes the following form:

$$U_i = \beta_0 + \beta_1 i + \epsilon_i$$

where U_i is the observed leading digit value from the sample proportions; β_0 and β_1 are the intercept and slope parameters respectively; i is the expected leading digit value; ϵ_i is the random error term.

A similar model was proposed in a 2006 article by Saville, using the standard regression model to test for conformity with Benford's Law using the expected and observed proportions of the first leading digits [17]. His model is as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

where Y_i is the observed proportion of the i^{th} leading digit, X_i's are the known Benford probabilities, β_0 and β_1 are the intercept and slope parameters, and ϵ_i is the random error term, with an expected value of 0. He then proposed jointly testing if the intercept and slope differed from 0 and 1 respectively [17].

However, data following Benford's Law would not be expected to fit the statistical framework used in the ordinary least squares (OLS) regression model. The OLS model assumes linearity, errors that are normally distributed with a mean of 0 and constant variance, and observations, and therefore errors, that are independent of each other. Since the proportions must sum to 1, our observations cannot be independent as they are calculated from the observed proportions and as one increases another must decrease. Due to the aforementioned issues, simulations were run at various sample sizes to determine the true distribution of the β estimates for linear regression using the Inverse Benford model; issues with Saville's model are discussed in detail in Section 4.2. Ten thousand simulations were run for each sample size and the β estimates were plotted. The summary statistics are recorded in Table A.1. In addition, the values of the 2.5^{th} and 97.5^{th} percentiles are recorded to be used as critical values for two-sided hypothesis testing, along with the percentiles for the α= 0.01 and 0.10 levels of significance; these results are seen in Table A.5. This method was repeated using regression through the origin, and the results are seen in Tables A.2 and A.6.

The simulation size of 10,000 was chosen due to the number of sample sizes to be tested and, as a result, the time constraints. Therefore the critical values recorded are approximate.

Due to the formula for the inverse, each leading digit must appear at least once for this method to be used to test for conformity with Benford's Law, since an observed proportion of 0 for one digit will give a value of 0 in the denominator for the corresponding U_i. Therefore, this test only works for larger datasets, which through simulations, was determined to be samples of size at least 200. The simulated $\hat{\beta}$ values are plotted in Figure 4.1. Since the correlation between the β_0 and β_1 values is approximately -0.977 for all sample sizes, the overall shape of the β_0 and β_1 plots are almost reflections of each other. For small sample sizes, the β distributions are highly skewed and there appears to be a small second mode in the right tail.

While the aforementioned issue in (4.2) only appeared in the simulations for samples smaller than 200, the probability of U_i being undefined due to a denominator of 0 is greater than 0 at large sample sizes as well. In order to resolve this issue, we propose using the multivariate normal approximation of the multinomial distribution. The vector of the estimated probabilities, $\hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_2, ..., \hat{p}_9)$, are $\frac{1}{n}$Multinomial(n,$\mathbf{p}$), where $\mathbf{p}$ is the vector of Benford proportions. Using the multivariate normal approximation, $\hat{\mathbf{p}} \sim \frac{1}{n}$MVN(n$\mathbf{p}$, Σ) where Σ is the $k \times k$ symmetric covariance matrix with diagonal elements $np_i(1-p_i)$ and off-diagonal elements $-np_ip_j$ where $i \neq j$. Therefore, $\hat{\mathbf{p}}$ ~MVN($\mathbf{p}$, Σ^*), where $\Sigma^* = \frac{1}{n^2}\Sigma$, allowing us to rewrite equation 4.3 as:

$$\begin{aligned} \mathbf{U} &= \frac{1}{10^{\hat{\mathbf{p}}} - 1} \\ &= \frac{1}{10^{MVN(\mathbf{p},\Sigma^*)} - 1} \end{aligned} \tag{4.4}$$

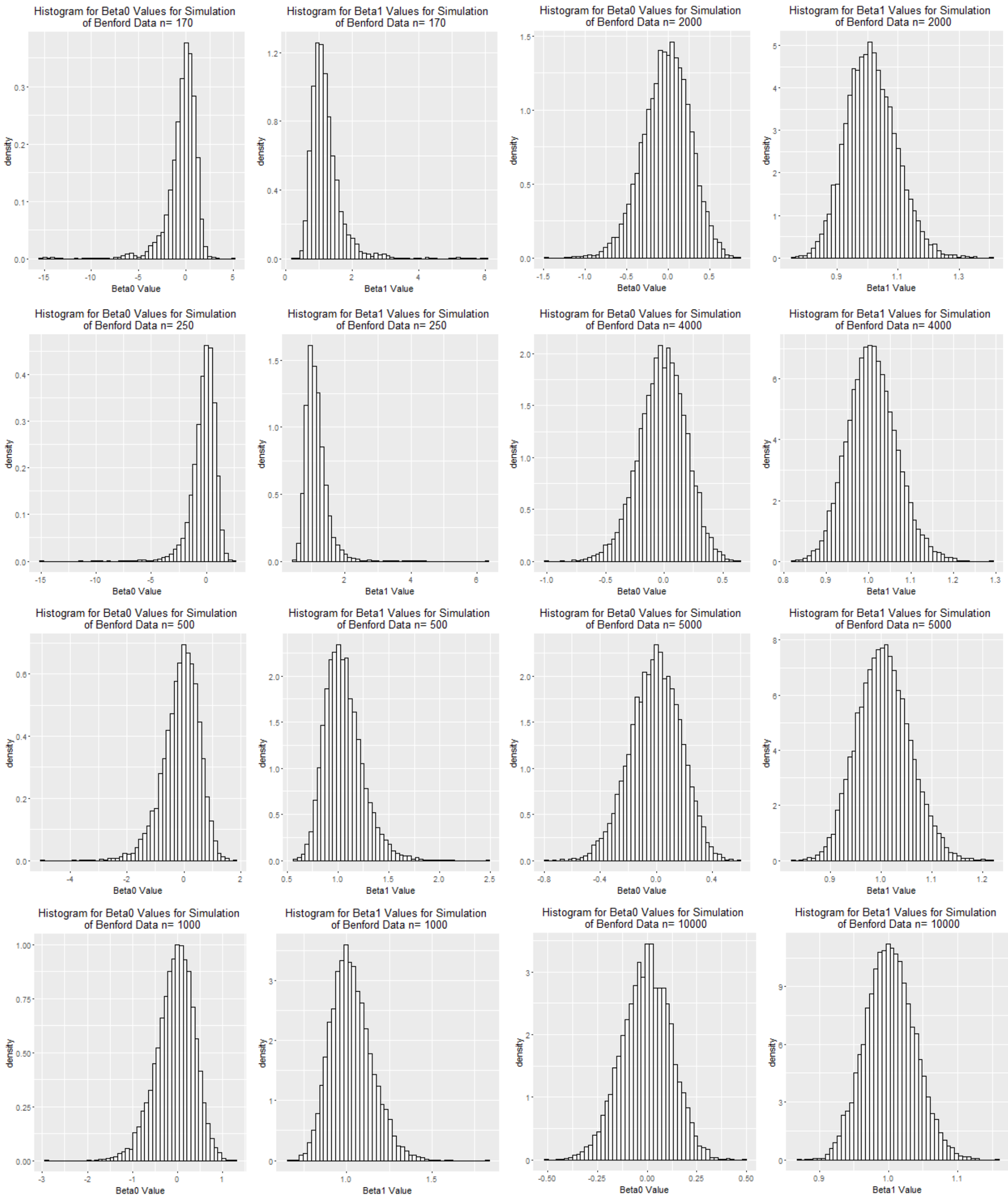

FIGURE 4.1: Simulated $\hat{\beta}$ distributions from the Inverse Benford Regression

This formulation removes the possibility of a denominator of 0, even if the proportion of one of the leading digits is 0, allowing it to be utilized for all sample sizes and for a wider variety of applications. Simulations were run to compare the critical values in Table A.5 to those identified through running the simulations using the multivariate normal approximation, and at the 5% level, the critical values were almost equivalent. The same appears to be true for the inverse regression through the origin using the multivariate normal approximation.

Simulations were run to determine the variability in the U_i values using the multinomial and multivariate normal formulations, plotted in Figures 4.2 and 4.3. The simulations both show heteroscedasticity, where, as the value of the leading digit increases, the variation in the estimated values becomes larger. The points are skewed to the right, and the plot using the multivariate normal approximation appears to have a slightly greater variation of estimated values for the higher leading digits. To compare summary statistics, Tables 4.1 and 4.2 contain the mean, median, and variance of the multinomial and multivariate normal forms respectively, at four sample sizes. The variance for samples of size 500 and 1000 is slightly greater at the higher leading digits using the multivariate normal approximation, as was seen when comparing Figures 4.2 and 4.3. However, excluding this, all three statistics from both tables are almost identical, showing that the multivariate normal approximation can be successfully utilized here.

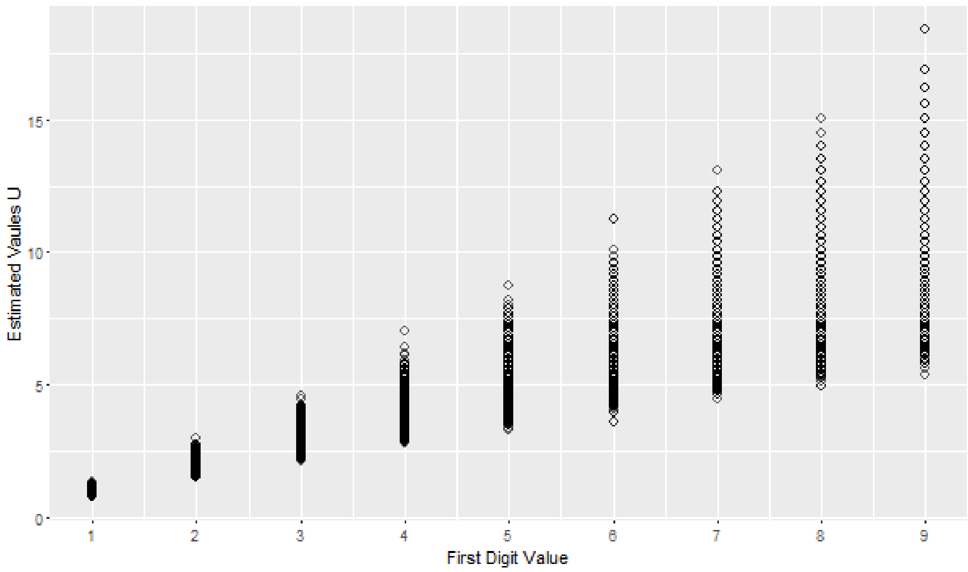

FIGURE 4.2: Simulated U_i values from multinomial distribution with Benford proportions for n=1000; N=10,000

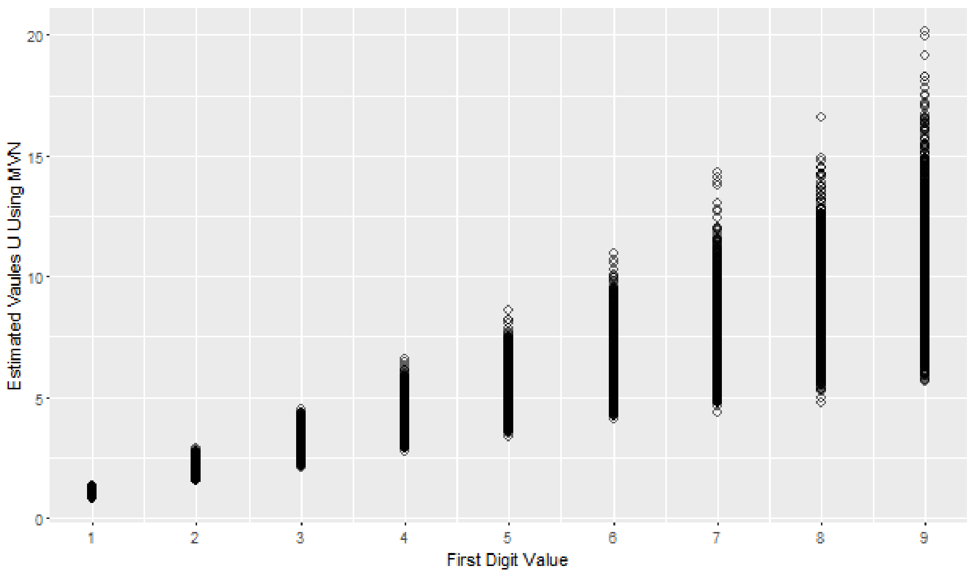

FIGURE 4.3: Simulated U_i values from multivariate normal approximation for n=1000; N=10,000

TABLE 4.1: Mean, median, and variance of simulated integer estimates from a multinomial distribution with Benford proportions; N=10,000

		Integer Values								
n	Summary Statistics	1	2	3	4	5	6	7	8	9
500	Mean	1.006	2.020	3.050	4.093	5.149	6.210	7.269	8.350	9.386
	Median	1.005	2.001	3.026	4.042	5.083	6.093	6.999	8.195	8.950
	Variance	0.009	0.059	0.182	0.414	0.840	1.443	2.338	3.372	4.867
1000	Mean	1.003	2.014	3.018	4.039	5.066	6.098	7.139	8.165	9.235
	Median	1.000	2.001	2.998	3.996	5.013	5.995	6.999	8.025	9.160
	Variance	0.005	0.028	0.085	0.202	0.372	0.623	1.023	1.536	2.175
5000	Mean	1.000	2.003	3.006	4.006	5.010	6.018	7.024	8.034	9.044
	Median	0.999	2.001	3.004	4.005	4.999	6.014	6.999	8.025	9.033
	Variance	0.001	0.005	0.017	0.037	0.071	0.117	0.184	0.276	0.397
10,000	Mean	1.000	2.001	3.002	4.005	5.009	6.006	7.009	8.021	9.025
	Median	1.000	1.998	3.001	4.000	5.006	5.995	6.999	8.009	9.012
	Variance	0.000	0.003	0.008	0.019	0.034	0.060	0.092	0.133	0.190

TABLE 4.2: Mean, median, and variance of simulated integer estimates from the multivariate normal approximation; N=10,000

		Integer Values								
n	Summary Statistics	1	2	3	4	5	6	7	8	9
500	Mean	1.007	2.024	3.053	4.084	5.135	6.216	7.297	8.346	9.453
	Median	1.002	1.999	2.999	3.991	4.995	6.024	7.007	7.983	8.994
	Variance	0.009	0.060	0.194	0.434	0.857	1.532	2.654	4.140	5.680
1000	Mean	1.004	2.012	3.026	4.041	5.063	6.104	7.137	8.156	9.183
	Median	1.000	2.001	2.998	3.994	4.992	6.006	7.012	7.981	8.971
	Variance	0.005	0.029	0.091	0.201	0.387	0.661	1.062	1.594	2.276
5000	Mean	1.001	2.002	3.005	4.008	5.016	6.018	7.018	8.029	9.034
	Median	1.000	2.000	2.998	3.998	5.002	6.000	6.996	7.996	8.992
	Variance	0.001	0.006	0.017	0.038	0.070	0.119	0.186	0.269	0.389
10,000	Mean	1.000	2.000	3.002	4.005	5.007	6.007	7.012	8.017	9.027
	Median	1.000	1.999	2.999	3.999	5.002	6.002	7.000	7.999	9.006
	Variance	0.000	0.003	0.008	0.019	0.036	0.059	0.092	0.137	0.189

In addition, for the smaller sample sizes, the median, being more robust than the mean, deviates only slightly from the expected integer values. While the probability distribution function of U_i is complicated analytically, and therefore finding the expected value posed problems, we were able to derive the formulation of the median. Using (4.3) and the definition that the median is any real number that satisfies both $P(U_i \leq median) \geq \frac{1}{2}$ and $P(U_i \geq median) \geq \frac{1}{2}$, by letting $1/\text{m}$ represent the median, we have:

$$
\begin{aligned}
P\left[U_i \geq \frac{1}{m}\right] &= \frac{1}{2} \\
P\left[\frac{1}{10^{X_i}-1} \geq \frac{1}{m}\right] &= \frac{1}{2} \\
P\left[10^{X_i}-1 \leq m\right] &= \frac{1}{2} \\
P\left[10^{X_i} \leq m+1\right] &= \frac{1}{2} \\
P\left[X_i \leq \frac{\log(m+1)}{\log 10}\right] &= \frac{1}{2}
\end{aligned}
$$

using

$$
Z = \frac{X_i - p_i}{\sqrt{\frac{p_i(1-p_i)}{n}}} \sim N(0,1)
$$

$$
P\left[Z \leq \frac{\frac{\log(m+1)}{log 10 - p_i}}{\sqrt{\frac{p_i(1-p_i)}{n}}}\right] = \frac{1}{2}
$$

However,

$$
\frac{\log(m+1)}{\log 10} = p_i
$$

and so,

$$
\begin{aligned}
\log(m+1) &= p_i \log 10 \\
m+1 &= 10^{p_i}
\end{aligned}
$$

But

$$p_i = log\left(1+\frac{1}{i}\right)$$

so,

$$\begin{aligned} 10^{p_i} &= 1+\frac{1}{i} \\ m &= \frac{1}{i} \\ \frac{1}{m} &= i \end{aligned}$$

Therefore, as the simulations showed, the median of U_i is i.

The heteroscedasticity seen in Figures 4.2 and 4.3, which can numerically be seen in Tables 4.1 and 4.2, suggests the need for weighted least squares regression to account for the increase in variation at larger values. The weights used are the inverse of the variance simulated using the multinomial approximation, and the critical values for this can be found in Table A.9.

4.2 Issues in Saville's Regression Analysis

As previously discussed, the model utilized by Saville to test for conformity with Benford's Law using the expected and observed proportions of leading digits is as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

where Y_i is the observed proportion of the i^{th} leading digit, X_i's are the known Benford probabilities, β_0 and β_1 are the intercept and slope parameters, and ϵ_i is the random error term, with an expected value of 0. The model was applied individually to datasets from 34 companies on the Johannesburg Stock Exchange by jointly testing if the intercept and slope differed from 0 and 1 respectively [17]. However, as mentioned in Section 4.1, the Benford model does not meet the assumptions of OLS regression and so his method was not statistically sound. One large violation of this is the lack of independence between observations. Saville used the observed proportions as his independent variable, however, since the proportions must sum to 1, they are directly related to each other since as one value decreases, another must increase. Furthermore, Saville jointly tests the hypotheses for the slope and intercept. In this case, however, these are directly related to each other, as testing that the intercept is at the origin also falls on the 1 to 1 line for which we are testing the slope for. Both Nigrini and Kossovsky critiqued his method and showed that it can not be utilized as a measure of conformity through examples [9, 13]. Kossovsky also derived the relationship between the intercept and slope, where intercept=(1-slope)/9 for the regression line [9].

In addition to the above, there appears to be some issues with the datasets that Saville used, as he mixed manipulatable and non-manipulatable numbers. In accounting, manipulatable numbers are defined as those easily manipulated, including a company's quarterly or annual profits and a taxpayer's taxable income, whereas non-manipulatable are values such as totals, subtotals, and numbers from tables or other pages [13]. He also initially classified each company as either errant or compliant, with the errant group being companies that had committed or were alleged to have committed accounting fraud and had their shares suspended or delisted. Nigrini states that the comparison is more between small companies that are doing poorly and large, successful companies, rather than errant and compliant companies[13].

Both Nigrini and Kossovsky note that the method used by Saville will reject conforming datasets and is believed to have a large error margin[9, 13].

In order to support our belief that the Benford model does not follow the OLS assumptions that Saville used, we simulated 10,000 samples from a multinomial distribution with Benford proportions to determine the rejection rate at three significance levels: α=0.01, 0.05, and 0.10. The results are shown in Table 4.3. From these results,

TABLE 4.3: Rejection rate of Saville's Benford Regression using OLS critical values at three α levels; N=10,000

n	α=0.01	α=0.05	α=0.1
50	0.0339	0.1226	0.2025
100	0.0320	0.1198	0.2064
500	0.0324	0.1224	0.2145
1000	0.0330	0.1232	0.2046
2000	0.0298	0.1154	0.2097
4000	0.0324	0.1247	0.2030
5000	0.0326	0.1280	0.2123
10,000	0.0332	0.1260	0.2055

we can see that following Saville's method, and therefore using the OLS critical values, the Type I error rate is, on average, at least twice the significance level for α=0.05 and 0.10, and three times as large for α=0.01. Accordingly, the β values are not normally distributed and the OLS assumptions do not hold. Using this method and testing for a regression line with a slope of 1 and an intercept of 0, the β values for the intercept and slope become highly correlated since they fall on the same 1 to 1 line. Due to this, we decided to set the intercept to 0 and examine Saville's model using regression through the origin (RTO). While RTO is seen as a controversial method and can introduce bias into the model, we felt it was worth examining since 1 is the smallest possible first leading digit and so the line should not be able to vary at any lower values. Table 4.4 shows the results of the Type I error simulation for this model. It

TABLE 4.4: Rejection rate of Saville's Benford Regression through the origin using OLS critical values at three α levels; N=10,000

n	α=0.01	α=0.05	α=0.1
50	0.00	0.00	0.0003
100	0.00	0.00	0.0005
500	0.00	0.00	0.0003
1000	0.00	0.00	0.0001
2000	0.00	0.00	0.0001
4000	0.00	0.00	0.0004
5000	0.00	0.00	0.0004
10,000	0.00	0.00	0.0001

can be seen that the rejection rate is essentially 0 for all three α levels, showing the regression assumptions do not hold here either.

Following this, we ran simulations using Saville's model, both in it's original form and regressing through the origin, to determine the true distribution of the β values, as was done with the inverse Benford regression. The results of the distribution of the β values are presented in Tables A.3 and A.4. Table A.3 shows that the correlation between the $\hat{\beta}_0$ and $\hat{\beta}_1$ values is -1 for all sample sizes, which gives further evidence to the fact that one does not have to simultaneously test both hypotheses as Saville was doing. Due to the nature of the data, where the proportions must sum to 1, it would not be possible to get a slope of 1 and an intercept that differs from 0. Table A.1 provided the correlation between the β values for our proposed inverse regression model, and for all sample sizes the correlation was approximately -0.977, which, while not quite -1, also suggests we do not need to test simultaneous hypotheses but rather testing the slope alone should be sufficient. Appendix A contains the approximate critical values for all of the regression models at α=0.01, 0.05, and 0.10 in Tables A.5 to A.9.

4.2.1 Power

Using the same proportions as in Table 3.6, the power of each of the five regression methods was tested at various sample sizes. This probability set was chosen since it provides similar but not exactly Benford values. As previously mentioned, it falls within the marginally acceptable conformity range when using the MAD test statistic. The rejection rates for the weighted inverse Benford regression are included on Table 4.5, and the results for the other four regression methods are included in Appendix A in Tables A.10 to A.13 due to the similarity of all 5 tables. Since we want a test statistic that is not overly sensitive to small deviations, the regression techniques provide a lower rejection rate than seen using MAD in 3.6, at least for these proportions.

TABLE 4.5: Rejection rate of Weighted Inverse Benford Regression simulated from a distribution with proportions {31.755, 16.11, 11.015, 8.287, 10.163, 6.028, 4.982, 5.037, 6.624}; N=10,000

n	α=0.01	α=0.05	α=0.1
200	0	0	0
500	0	0	0
1000	0	0	0
2000	0	0	0
4000	0	0.0001	0.0001
5000	0.0001	0.0001	0.0001
10,000	0.0001	0.0001	0.0001

4.2.2 Applied Examples

In this section, we look at eight datasets and test for conformity with Benford's Law using the five regression models, Pearson's χ^2, MAD, and SSD. For the regression models, the results using OLS assumptions were also provided for the comparison of the model Saville suggested, to the same model using the critical values that fit the true distribution, at the 5% level. The datasets used were: the first 1002 terms of the Fibonacci sequence, the first 1002 terms of the powers of two and the powers of $10^{0.05}$, the Sino Forest dataset which contained the numbers from Sino Forest Corporation's 2010 financial report, two datasets made from the sets of proportions used by Kossovsky to critique Saville's regression model [9], and two datasets used in Nigrini's 2012 book, which were the 2005 journal entries in a company's accounting system and a collection of the daily returns for Apple [13]. Results are presented in Tables 4.6 through 4.13. All five regression models matched the test statistic results for the Fibonacci sequence, powers of 2, the Sino Forest data, and the second set of Kossovsky's proportions. It must be noted that while Saville's regression accepts conformity to Benford's Law here, this is only due to this identification of the new critical values through the previous run simulations. Saville used the OLS model in his paper, and therefore adopts the assumption of normally distributed β values, and that method rejects conformity here [17].

Disagreements between the different regression models and the test statistics are seen for both the powers of $10^{0.05}$ and the first set of Kossovsky's proportions. For the powers of $10^{0.05}$, both Pearson's Goodness-of-Fit test and MAD show nonconformity, and while the SSD statistic has a relatively high value, it shows marginal conformity. Using the regression methods, only the inverse Benford regression model and the weighted inverse regression model identified nonconformity with Benford's Law. In the first set of Kossovsky's proportion, which he used in his work to show that Saville's method would accept data that strongly deviated from Benford's Law, the same result was seen. All three test statistics show nonconformity; however, both Saville's regression and his model regressing through the origin show conformity. All three inverse regression models identified that the data did not follow Benford's Law, which shows improvements upon Saville's model even when using critical values from the true distribution.

The journal entry and Apple returns datasets apply the conformity tests to real data of larger sample sizes. The journal entry data has 154,935 observations and shows where the χ^2 goodness of fit test fails. MAD and SSD show close and acceptably close levels of conformity respectively, while the χ^2 test shows nonconformity. This is expected for such a large sample size, since it is known that the χ^2 becomes overly sensitive as the sample size increases. The results of the regression tests are interesting, as both tests following Saville's model reject conformity, as does the weighted inverse Benford regression and inverse Benford through the origin. Only the inverse Benford provides conformity results. The Apple returns data, with sample size 6799, also showed the issues with the χ^2 test statistic which rejected the null hypothesis, while both MAD and SSD showed acceptably close conformity. All 5 regression tests gave results of nonconformity as well. The proportions of leading digits for this data are {0.30099, 0.20186, 0.13758, 0.10063, 0.074196, 0.06819, 0.04566, 0.036648, 0.03424}, which are very close to the actual Benford probabilities, showing some issues that exist with the regression methods. It should be noted that simulations were run to calculate the approximate critical values for both sample sizes of 154,935 and 6799 to have accurate results. While Tables 4.6 through 4.13 look at a 5% significance level for the regression methods, as mentioned in Section 3.3.3, the MAD test statistic is using an α of approximately 0 for these larger sample sizes; therefore, one can not make a fair comparison between these methods. The critical values for these sample sizes were calculated for smaller significance levels, such as α=0.005 and 0.0001, using the weighted inverse Benford regression. While the method still rejected conformity in both cases, the critical values became closer to the estimated slope and intercept, and so an increase in the number of simulations run would provide more accurate critical values and could change the result seen.

TABLE 4.6: Conformity tests for the Fibonacci Sequence; n=1002

Test Statitics			Conform to Benford's Law	
Chi-square p-value	0.9999995		close conformity	
MAD	0.0007385		close conformity	
SSD	0.0710728		close conformity	
Regression Type	$\hat{\beta}_0$ Estimate	$\hat{\beta}_1$ Estimate	Conform to Benford's Law	
			Using OLS	Using New Criteria
Inverse Benford Regression	-0.005938	1.002407	yes	yes
Inverse Benford Regression Through the Origin		1.001469	yes	yes
Weighted Inverse Benford Regression	-0.004310	1.002113	yes	yes
Saville's Benford Regression	-0.0002271	1.0020438	yes	yes
Saville's Regression Through the Origin		1.0006713	yes	yes

TABLE 4.7: Conformity tests for the Powers of 2; n=1002

Test Statitics			Conform to Benford's Law	
Chi-square p-value	0.9999983		close conformity	
MAD	0.00079596		close conformity	
SSD	0.09970152		close conformity	
Regression Type	$\hat{\beta}_0$ Estimate	$\hat{\beta}_1$ Estimate	Conform to Benford's Law	
			Using OLS	Using New Criteria
Inverse Benford Regression	-0.004204	1.01247	yes	yes
Inverse Benford Regression Through the Origin		1.005836	yes	yes
Weighted Inverse Benford Regression	-0.005695	1.002832	yes	yes
Saville's Benford Regression	-0.0003138	1.002824	yes	yes
Saville's Regression Through the Origin		1.0009275	yes	yes

TABLE 4.8: Conformity tests for the Sino Forest dataset; n=772

Test Statitics			Conform to Benford's Law	
Chi-square p-value	0.4682064		conformity	
MAD	0.00659813		acceptable conformity	
SSD	7.08275		acceptable conformity	
Regression Type	$\hat{\beta}_0$ Estimate	$\hat{\beta}_1$ Estimate	Conform to Benford's Law	
			Using OLS	Using New Criteria
Inverse Benford Regression	0.5229	0.85	yes	yes
Inverse Benford Regression Through the Origin		0.93251	yes	yes
Weighted Inverse Benford Regression	0.04945	0.98818	yes	yes
Saville's Benford Regression	0.005166	0.953505	yes	yes
Saville's Regression Through the Origin		0.98473	yes	yes

TABLE 4.9: Conformity tests for Powers of $10^{0.05}$; n=1002

Test Statitics			Conform to Benford's Law	
Chi-square p-value	1.902427e-12		nonconformity	
MAD	0.02037764		nonconformity	
SSD	60.02087		marginal conformity	
Regression Type	$\hat{\beta}_0$ Estimate	$\hat{\beta}_1$ Estimate	Conform to Benford's Law	
			Using OLS	Using New Criteria
Inverse Benford Regression	1.0704	0.8182	yes	no
Inverse Benford Regression Through the Origin		0.98717	yes	yes
Weighted Inverse Benford Regression	-0.1146	1.1535	yes	no
Saville's Benford Regression	0.006046	0.945583	yes	yes
Saville's Regression Through the Origin		0.98213	yes	yes

TABLE 4.10: Conformity tests for {21.7%,36.8%,9.6%,14.5%,1.0%,1.0%,3.4%,6.5%,5.5%}; n=1000

Test Statitics			Conform to Benford's Law	
Chi-square p-value	8.418438e-79		nonconformity	
MAD	0.05846416		nonconformity	
SSD	559.2187		nonconformity	
Regression Type	$\hat{\beta}_0$ Estimate	$\hat{\beta}_1$ Estimate	Conform to Benford's Law	
			Using OLS	Using New Criteria
Inverse Benford Regression	5.3442	1.6112	yes	no
Inverse Benford Regression Through the Origin		1.9908	yes	no
Weighted Inverse Benford Regression	-2.085	3.134	yes	no
Saville's Benford Regression	0.0000156	0.999986	yes	yes
Saville's Regression Through the Origin		0.99996	yes	yes

TABLE 4.11: Conformity tests for {30.4%,17.8%,12.6%,9.7%,7.9%,6.6%,5.6%,5.0%,4.4%}; n=1000

Test Statitics			Conform to Benford's Law	
Chi-square p-value	0.9999928		close conformity	
MAD	0.00134		close conformity	
SSD	0.2291279		close conformity	
Regression Type	$\hat{\beta}_0$ Estimate	$\hat{\beta}_1$ Estimate	Conform to Benford's Law	
			Using OLS	Using New Criteria
Inverse Benford Regression	-0.146443	1.048685	no	yes
Inverse Benford Regression Through the Origin		1.025563	no	yes
Weighted Inverse Benford Regression	-0.035870	1.016932	no	yes
Saville's Benford Regression	-0.0021241	1.019117	no	yes
Saville's Regression Through the Origin		1.006279	yes	yes

TABLE 4.12: Conformity tests for the Journal Entry data (Nigrini 5.16 [13]); n=154,935

Test Statitics			Conform to Benford's Law	
Chi-square p-value	3.200135e-54		nonconformity	
MAD	0.00464314		close conformity	
SSD	4.826008		acceptably close conformity	
Regression Type	$\hat{\beta}_0$ Estimate	$\hat{\beta}_1$ Estimate	Conform to Benford's Law	
			Using OLS	Using New Criteria
Inverse Benford Regression	0.04682	1.02009	yes	yes
Inverse Benford Regression Through the Origin		1.02748	yes	no
Weighted Inverse Benford Regression	-0.15085	1.07305	no	no
Saville's Benford Regression	-0.008274	1.074470	yes	no
Saville's Regression Through the Origin		1.02446	yes	no

TABLE 4.13: Conformity tests for Apple Returns data (Nigrini 11.16 [13]); n=6799

Test Statitics			Conform to Benford's Law	
Chi-square p-value	3.309764e-08		non-conformity	
MAD	0.009638653		acceptably close conformity	
SSD	13.59194		acceptably close conformity	
Regression Type	$\hat{\beta}_0$ Estimate	$\hat{\beta}_1$ Estimate	Conform to Benford's Law	
			Using OLS	Using New Criteria
Inverse Benford Regression	-1.4816	1.4746	no	no
Inverse Benford Regression Through the Origin		1.24071	no	no
Weighted Inverse Benford Regression	-0.1702	1.1014	yes	no
Saville's Benford Regression	-0.008528	1.076749	yes	no
Saville's Regression Through the Origin		1.02521	yes	no

Chapter 5

Conclusions

A remarkable distribution in the statistical literature is Benford's law. This law describes a phenomenon wherein the leading digits of a set of naturally occurring numbers follows a decreasing logarithmic trend. In the present thesis, we examined three conformity tests by applying them to mathematics and statistics multiple choice test banks, and by using more in depth simulation studies. While the full test bank dataset and its subsets did not conform to Benford's Law using Pearson's χ^2 goodness of fit test, the full dataset, both with and without the single digit answers, and the full distractors had marginally acceptable conformity in the first digit when using MAD and SSD. The distractors without the single digit answers had acceptable conformity, again when using MAD and SSD. When testing the first two digits, both the χ^2 and MAD resulted in non-conformity, while SSD produced all values within the marginally acceptable conformity range. We examined the power of the χ^2 test statistic and the two simultaneous confidence intervals, Goodman and Sison & Glaz, at the sample size of our dataset. Since the confidence intervals and χ^2 statistic have significance levels, we were able to compare the accuracy, which was as expected.

Although the test bank questions did not conform to Benford's Law for both the correct answers and the distractors, as was seen in the study done by Slepkov et al. on physics test banks, it is worth noting that Selpkov only examined the first leading digits. Nigrini states that the first digit test is at too high of a level to be utilized in a thorough data analysis, and significant deviations can be present in the later digits [13]. So while the physics multiple choice test bank questions conformed in the first digit, an investigation into the first two or second digits may result in non-conformity, as was seen in both the present study and Nigrini's examination of accounting test banks [12].

The MAD statistic was examined at various sample sizes to develop a statistical framework in the context of hypothesis testing. Through simulations to determine the power of the test, it was observed that the α level adjusts as the number of observations changes since the critical values are fixed. By large sample sizes, where this test is often applied, the α value is essentially 0. In addition, the test statistic can be viewed at three separate hypothesis tests as one can test if the observed MAD value falls below the upper limit of any of the three conformity ranges.

An inverse Benford regression procedure was proposed as an alternative conformity test in an effort to form a test with statistical framework that was not as sensitive to slight deviations as the χ^2 goodness of fit test at large sample sizes. The approximate critical values were simulated, both when including and excluding an intercept term. Additionally, the regression method suggested by Saville [17] was examined, showing that the model does not follow the OLS assumption that he utilized, and critical values reflecting the true distribution of the β estimates were found. As the residuals of the inverse regression method were heteroscedastic, we proposed using weighted regression as the model. Applied examples were utilized to compare the five regression methods with the three test statistics. While the inverse regression models slightly outperformed the χ^2 test statistic, there appear to some similar asymptotic issues present at larger sample sizes.

Overall, our contributions included the proposal of the aforementioned inverse Benford regression model, the thorough examination of the MAD test statistic and the α levels associated with treating it as a hypothesis test, and further justification that Saville's regression model does not meet the OLS assumptions. We also showed approximate critical values for all 5 regression models.

Due to the computational time required to run the critical value simulations, the tables provided are only approximate. Future work could explore running 100,000 or more simulations to get more accurate critical values. In addition, one could extend the weighted inverse Benford regression model to the first two digits, which may have improved performance at large sample sizes.

Appendix A

Chapter 4 Tables

TABLE A.1: Summary statistics for Inverse Benford Regression; N=10,000

n	Estimate	Mean	Standard Deviation	Correlation
200	$\hat{\beta}_0$	-0.299	1.357	-0.978
	$\hat{\beta}_1$	1.157	0.405	
500	$\hat{\beta}_0$	-0.103	0.636	-0.977
	$\hat{\beta}_1$	1.055	0.186	
1000	$\hat{\beta}_0$	-0.049	0.417	-0.977
	$\hat{\beta}_1$	1.027	0.121	
2000	$\hat{\beta}_0$	-0.022	0.282	-0.976
	$\hat{\beta}_1$	1.013	0.082	
4000	$\hat{\beta}_0$	-0.016	0.199	-0.977
	$\hat{\beta}_1$	1.008	0.058	
5000	$\hat{\beta}_0$	-0.008	0.178	-0.977
	$\hat{\beta}_1$	1.005	0.051	
10,000	$\hat{\beta}_0$	-0.004	0.123	-0.976
	$\hat{\beta}_1$	1.002	0.036	

TABLE A.2: Summary statistics for Inverse Benford Regression through the Origin; N=10,000

n	Estimate	Mean	Standard Deviation
200	$\hat{\beta}_1$	1.076	0.196
500	$\hat{\beta}_1$	1.037	0.088
1000	$\hat{\beta}_1$	1.018	0.058
2000	$\hat{\beta}_1$	1.009	0.040
4000	$\hat{\beta}_1$	1.005	0.028
5000	$\hat{\beta}_1$	1.003	0.025
10,000	$\hat{\beta}_1$	1.002	0.017

TABLE A.3: Summary statistics for Saville's Benford Regression; N=10,000

n	Estimate	Mean	Standard Deviation	Correlation
50	$\hat{\beta}_0$	0.0004669	0.028	-1
	$\hat{\beta}_1$	0.9958	0.254	
100	$\hat{\beta}_0$	0.00004	0.020	-1
	$\hat{\beta}_1$	0.9996	0.180	
500	$\hat{\beta}_0$	-0.00002412	0.009	-1
	$\hat{\beta}_1$	1.000	0.079	
1000	$\hat{\beta}_0$	-0.00001818	0.006	-1
	$\hat{\beta}_1$	1.000	0.056	
2000	$\hat{\beta}_0$	-0.00001587	0.005	-1
	$\hat{\beta}_1$	0.9999	0.041	
4000	$\hat{\beta}_0$	-0.00001334	0.003	-1
	$\hat{\beta}_1$	1.000	0.028	
5000	$\hat{\beta}_0$	-0.0000574	0.003	-1
	$\hat{\beta}_1$	1.001	0.025	
10,000	$\hat{\beta}_0$	-0.000006261	0.002	-1
	$\hat{\beta}_1$	1.000	0.018	

TABLE A.4: Summary statistics for Saville's Benford Regression through the Origin; N=10,000

n	Estimate	Mean	Standard Deviation
50	$\hat{\beta}_1$	1.000	0.083
100	$\hat{\beta}_1$	1.000	0.059
500	$\hat{\beta}_1$	0.999	0.026
1000	$\hat{\beta}_1$	1.000	0.019
2000	$\hat{\beta}_1$	1.000	0.013
4000	$\hat{\beta}_1$	1.000	0.009
5000	$\hat{\beta}_1$	1.000	0.008
10,000	$\hat{\beta}_1$	1.000	0.006

TABLE A.5: Critical values for Inverse Benford Regression; N=10,000

		α=0.01		α=0.05		α=0.1	
n	Estimate	0.5%	99.5%	2.5%	97.5%	5%	95%
200	$\hat{\beta}_0$	−5.355	1.937	−3.520	1.460	−2.591	1.239
	$\hat{\beta}_1$	0.543	2.765	0.652	2.124	0.713	1.832
500	$\hat{\beta}_0$	−2.327	1.209	−1.516	0.942	−1.218	0.792
	$\hat{\beta}_1$	0.674	1.712	0.759	1.475	0.796	1.383
1000	$\hat{\beta}_0$	−1.316	0.880	−0.927	0.691	−0.764	0.587
	$\hat{\beta}_1$	0.763	1.397	0.816	1.285	0.845	1.236
2000	$\hat{\beta}_0$	−0.822	0.634	−0.608	0.487	−0.498	0.412
	$\hat{\beta}_1$	0.827	1.248	0.865	1.183	0.885	1.153
4000	$\hat{\beta}_0$	−0.565	0.451	−0.426	0.359	−0.352	0.306
	$\hat{\beta}_1$	0.872	1.167	0.900	1.125	0.914	1.103
5000	$\hat{\beta}_0$	−0.492	0.417	−0.375	0.322	−0.306	0.276
	$\hat{\beta}_1$	0.884	1.146	0.909	1.109	0.923	1.091
10,000	$\hat{\beta}_0$	−0.349	0.307	−0.253	0.233	−0.212	0.194
	$\hat{\beta}_1$	0.914	1.103	0.934	1.074	0.944	1.063

Table A.6: Critical values for Inverse Benford Regression through the Origin; N=10,000

		α=0.01		α=0.05		α=0.1	
n	Estimate	0.5%	99.5%	2.5%	97.5%	5%	95%
200	$\hat{\beta}_1$	0.810	2.025	0.855	1.579	0.883	1.451
500	$\hat{\beta}_1$	0.858	1.326	0.892	1.234	0.910	1.196
1000	$\hat{\beta}_1$	0.889	1.197	0.917	1.144	0.930	1.121
2000	$\hat{\beta}_1$	0.920	1.124	0.938	1.093	0.947	1.078
4000	$\hat{\beta}_1$	0.939	1.079	0.953	1.060	0.961	1.050
5000	$\hat{\beta}_1$	0.944	1.070	0.956	1.053	0.964	1.045
10,000	$\hat{\beta}_1$	0.959	1.049	0.969	1.037	0.974	1.031

TABLE A.7: Critical values for Saville's Benford Regression; N=10,000

		α=0.01		α=0.05		α=0.1	
n	Estimate	0.5%	99.5%	2.5%	97.5%	5%	95%
200	$\hat{\beta}_0$	−0.037	0.035	−0.027	0.027	−0.023	0.023
	$\hat{\beta}_1$	0.681	1.329	0.757	1.247	0.793	1.206
500	$\hat{\beta}_0$	−0.023	0.022	−0.017	0.017	−0.015	0.015
	$\hat{\beta}_1$	0.800	1.203	0.844	1.157	0.869	1.133
1000	$\hat{\beta}_0$	−0.016	0.016	−0.012	0.012	−0.010	0.010
	$\hat{\beta}_1$	0.856	1.143	0.893	1.110	0.908	1.092
2000	$\hat{\beta}_0$	−0.011	0.011	−0.009	0.009	−0.007	0.007
	$\hat{\beta}_1$	0.898	1.102	0.921	1.078	0.935	1.065
4000	$\hat{\beta}_0$	−0.008	0.008	−0.006	0.006	−0.005	0.005
	$\hat{\beta}_1$	0.928	1.073	0.946	1.054	0.954	1.045
5000	$\hat{\beta}_0$	−0.007	0.007	−0.005	0.006	−0.005	0.005
	$\hat{\beta}_1$	0.934	1.064	0.950	1.049	0.959	1.041
10000	$\hat{\beta}_0$	−0.005	0.005	−0.004	0.004	−0.003	0.003
	$\hat{\beta}_1$	0.955	1.046	0.965	1.035	0.971	1.030

TABLE A.8: Critical values for Saville's Benford Regression through the Origin; N=10,000

		α=0.01		α=0.05		α=0.1	
n	Estimate	0.5%	99.5%	2.5%	97.5%	5%	95%
200	$\hat{\beta}_1$	0.892	1.107	0.918	1.081	0.931	1.068
500	$\hat{\beta}_1$	0.936	1.069	0.949	1.052	0.957	1.044
1000	$\hat{\beta}_1$	0.952	1.048	0.964	1.037	0.969	1.030
2000	$\hat{\beta}_1$	0.965	1.034	0.974	1.026	0.978	1.022
4000	$\hat{\beta}_1$	0.976	1.023	0.982	1.018	0.985	1.015
5000	$\hat{\beta}_1$	0.978	1.021	0.983	1.016	0.986	1.014
10000	$\hat{\beta}_1$	0.985	1.015	0.988	1.011	0.990	1.010

TABLE A.9: Critical values for Weighted Inverse Benford Regression; N=10,000

		α=0.01		α=0.05		α=0.1	
n	Estimate	0.5%	99.5%	2.5%	97.5%	5%	95%
200	$\hat{\beta}_0$	−0.832	0.806	−0.639	0.557	−0.552	0.443
	$\hat{\beta}_1$	0.677	1.568	0.774	1.421	0.824	1.358
500	$\hat{\beta}_0$	−0.464	0.457	−0.355	0.314	−0.307	0.256
	$\hat{\beta}_1$	0.819	1.266	0.876	1.199	0.900	1.173
1000	$\hat{\beta}_0$	−0.309	0.303	−0.243	0.224	−0.206	0.185
	$\hat{\beta}_1$	0.873	1.166	0.906	1.129	0.924	1.111
2000	$\hat{\beta}_0$	−0.219	0.212	−0.166	0.157	−0.141	0.132
	$\hat{\beta}_1$	0.908	1.111	0.932	1.085	0.944	1.073
4000	$\hat{\beta}_0$	−0.150	0.144	−0.116	0.111	−0.099	0.091
	$\hat{\beta}_1$	0.937	1.074	0.951	1.058	0.960	1.049
5000	$\hat{\beta}_0$	−0.130	0.135	−0.100	0.099	−0.086	0.082
	$\hat{\beta}_1$	0.940	1.064	0.956	1.049	0.964	1.042
10,000	$\hat{\beta}_0$	−0.094	0.095	−0.073	0.071	−0.061	0.059
	$\hat{\beta}_1$	0.957	1.045	0.968	1.036	0.974	1.030

TABLE A.10: Rejection rate of Saville's Benford Regression simulated from a distribution with proportions {31.755, 16.11, 11.015, 8.287, 10.163, 6.028, 4.982, 5.037, 6.624}; N=10,000

n	α=0.01	α=0.05	α=0.1
200	0	0	0
500	0	0.0001	0.0001
1000	0	0	0
2000	0	0	0
4000	0	0	0.0001
5000	0	0	0
10,000	0	0	0.0001

TABLE A.11: Rejection rate of Saville's Benford Regression through the Origin simulated from a distribution with proportions {31.755, 16.11, 11.015, 8.287, 10.163, 6.028, 4.982, 5.037, 6.624}; N=10,000

n	α=0.01	α=0.05	α=0.1
200	0	0	0
500	0	0	0
1000	0	0	0
2000	0	0	0.0001
4000	0	0.0001	0.0001
5000	0	0	0
10,000	0	0	0.0001

TABLE A.12: Rejection rate of Inverse Benford Regression through the Origin simulated from a distribution with proportions {31.755, 16.11, 11.015, 8.287, 10.163, 6.028, 4.982, 5.037, 6.624}; N=10,000

n	α=0.01	α=0.05	α=0.1
200	0	0	0
500	0	0	0
1000	0	0	0.0001
2000	0	0	0
4000	0	0	0.0001
5000	0	0.0001	0.0001
10,000	0.0001	0.0001	0.0001

TABLE A.13: Rejection rate of Inverse Benford Regression simulated from a distribution with proportions {31.755, 16.11, 11.015, 8.287, 10.163, 6.028, 4.982, 5.037, 6.624}; N=10,000

n	α=0.01	α=0.05	α=0.1
200	0.0001	0.0001	0.0001
500	0	0.0001	0.0001
1000	0	0	0
2000	0.0001	0.0001	0.0001
4000	0.0001	0.0001	0.0001
5000	0.0001	0.0001	0.0001
10,000	0.0001	0.0001	0.0001

Bibliography

[1] V. Baláž, K. Nagasaka, and O. Strauch. "Benford's law and distribution functions of sequences in (0, 1)". *Mathematical Notes* 88.3 (2010), pp. 449–463. ISSN: 1573-8876. DOI: 10.1134/S0001434610090178.

[2] F. Benford. "The law of anomalous numbers". *Proceedings of the American Philosophical Society* 78.4 (1938), pp. 551–572.

[3] G.A. Brenner and R. Brenner. "Memory and Markets, or Why Are You Paying $2.99 for a Widget?" *The Journal of Business* 55.1 (1982), pp. 147–158. DOI: 10.2307/2352858.

[4] C. Carslaw. "Anomalies in Income Numbers: Evidence of Goal Oriented Behavior". *The Accounting Review* 63.2 (1988), pp. 321–327.

[5] L.A. Goodman. "On simultaneous confidence intervals for multinomial proportions". *Technometrics* 7.2 (1965), pp. 247–254. DOI: 10.1080/00401706.1965.10490252.

[6] S.A. Goudsmit and W.H. Furry. "Significant Figures of Numbers in Statistical Tables". *Nature* 154 (Dec. 1944), pp. 800–801. DOI: 10.1038/154800a0.

[7] T.P. Hill. "A Statistical Derivation of the Significant-Digit Law". *Statistical Sciences* 10.4 (1995), pp. 354–363.

[8] F.M. Hoppe. “Benford’s law and distractors in multiple choice exams”. *International Journal of Mathematical Education in Science and Technology* 47.4 (2016), pp. 606–612. DOI: 10.1080/0020739X.2015.1091515.

[9] A.E. Kossovsky. *Benford’s Law: Theory, the General Law of Relative Quantities, and Forensic Fraud Detection Applications*. Series in Computer Vision. World Scientific Publishing Company, 2014. ISBN: 9789814583701.

[10] M. Lesperance, W. J. Reed, M. A. Stephens, C. Tsao, and B. Wilton. “Assessing Conformance with Benford’s Law: Goodness-Of-Fit Tests and Simultaneous Confidence Intervals”. *PLoS ONE* 11.3 (Mar. 2016), pp. 1–20. DOI: 10.1371/journal.pone.0151235.

[11] S. Newcomb. “Note on the frequency of use of the different digits in natural numbers”. *American Journal of Mathematics* 4 (1881), pp. 39–40. DOI: 10.2307/2369148.

[12] M.J. Nigrini. “The Implications of the Similarity Between Fraud Numbers and the Numbers in Financial Accounting Textbooks and Test Banks”. *Journal of Forensic Accounting Research* 0.0 (2016), null. DOI: 10.2308/jfar-51465.

[13] M.J. Nigrini and J.T. Wells. *Benford’s Law: Applications for Forensic Accounting, Auditing, and Fraud Detection*. Wiley Corporate F&A. Wiley, 2012. ISBN: 9781118152850.

[14] R.S. Pinkham. “On the Distribution of First Significant Digits”. *The Annals of Mathematical Statistics* 32.4 (Dec. 1961), pp. 1223–1230. DOI: 10.1214/aoms/1177704862.

[15] C.P. Quesenberry and D.C. Hurst. “Large sample simultaneous confidence intervals for multinomial proportions”. *Technometrics* 6.2 (1964), pp. 191–195. DOI: 10.1080/00401706.1964.10490163.

[16] R.A. Raimi. “The First Digit Problem”. *American Mathematical Monthly* 83.7 (1976), pp. 521–538. DOI: 10.2307/2319349.

[17] A.D. Saville. “Using Benford’s law to detect data error and fraud: An examination of companies listed on the Johannesburg Stock Exchange”. *South African Journal of Economic and Management Sciences* 9.3 (2006), pp. 341–354.

[18] C.P. Sison and J. Glaz. “Simultaneous confidence intervals and sample size determination for multinomial proportions”. *Journal of the American Statistical Association* 90.429 (1995), pp. 366–369. DOI: 10.1080/01621459.1995.10476521.

[19] A.D. Slepkov, K.B. Ironside, and D. DiBattista. “Benford’s Law: Textbook Exercises and Multiple-Choice Testbanks”. *PLoS ONE* 10.2 (2015), pp. 1–13. DOI: 10.1371/journal.pone.0117972.

[20] W. Weaver. *Lady Luck: The Theory of Probability*. Dover Books on Mathematics. Dover, 1963. ISBN: 9780486243429.

[21] S.C.Y. Wong. “Testing Benford’s Law with the first two significant digits”. MA thesis. Victoria, B.C.: University of Victoria, 2010.