Università Ca'Foscari Venezia

Master's Degree programme
in
Data Analytics for Business and Society

Final Thesis

Benford Law and its application to electoral data

Supervisor
Prof.ssa Angela Andreella

Graduand
Francesco Toffoletto
Matriculation Number 873373

Academic Year
2023 / 2024

Contents

1 Introduction

1.1. Purpose of the thesis

Why are we interested in finding rules that govern our data? In today's interconnected world, vast amounts of data are generated and accessible every day, influencing decisions in diverse fields such as science, business, government, and technology. It is essential to identify patterns or rules that govern our data to better understand and predict them, as well as to detect anomalies. If some data deviate from the expected distribution or identified trends, such inconsistencies could indicate changes in behaviour, new trends or underlying problems such as errors, anomalies or even potential manipulations.

Benford's law is one such rule, proven to govern both some natural and non-natural data. Discovered in the late nineteenth century, this law reveals a predictable pattern in the frequency of leading digits within datasets. Recognising these patterns helps us to exploit our data in the best possible way and sometimes anticipate trends or detect errors.

Interest in this law has grown recently in terms of the number of publications and increasing attention. We can see that there have been more publications in recent years than in previous ones, according to the online database provided by Berger and Hill: `http://www.benfordonline.net`, 2009. (Last accessed: 23 November 2024).

The idea of this thesis is quite simple: If it has been proven or verified that certain categories of data belong or follow a Benford's distribution, we can check for any anomalies by comparing our datasets with the Benford's distribution. In the specific case of Benford's distribution, falsified datasets can be identified by comparing observed data to the logarithm formula of the distribution. In this document, we will focus on electoral data. Therefore, the aim of this document is to analyse electoral data to determine if there are any anomalies. We will examine official data provided by government agencies, acknowledging that, unfortunately, such data may already have been manipulated. Of course, the results will not constitute definitive proof, but can act as a warning sign, raising suspicions, and prompting further investigation.

By **electoral data**, we refer to data from a specific election in a particular country. To better analyse the data and assess their conformity with Benford's law, we will use the most granular level of detail available, aggregated at the smallest polling station level. Using the most detailed level possible, we will encounter some problems specific to statistical tests, such as the excess power problem. However, with the combined use of statistical tests and metrics, we want to determine whether Benford's law is not merely theoretical but can actually help identify fraud, anomalies, or reporting errors.

2 Benford's law

In nature, there are certain rules to which the numbers must comply. At first glance, these rules may seem illogical or, at the very least, quite strange. For example, when we think of a set of random numbers, we might assume that each number from 1 to 9 has the same probability of appearing as the first digit; for simplicity, we exclude zero and digits beyond the first. However, this assumption is incorrect. A similar error occurs when we try to create a set of natural numbers randomly. For example, if we consider a set of 100 numbers, we might expect each digit from 1 to 9 to appear as the first digit with an equal probability of approximately 11%.

In reality, the distribution of the first digits is more complex. The probabilities are not uniformly distributed and exhibit a pattern that may initially seem unusual. However, upon closer inspection, it becomes evident that many different sets of numbers adhere to the same type of non-uniform distribution. This phenomenon was first identified by Simon Newcomb (Newcomb, 1881) and later disclosed by Frank Benford (Benford, 1938). Lower numbers, such as 1, were shown to have a higher probability of appearing as the first significant digit compared to higher numbers.

Specifically, the number 1 appears more frequently as the leading digit than the number 2, the number 2 appears more frequently than the number 3, and so on, up to the digit 9. The strange fact is that the digit 1 occurs as the first digit a little more than six times more than 9. The two smallest digits occur as the first significant digit with a combined probability close to 50%, whereas the largest digits together have a probability of less than 10%.

Similarly to how the normal distribution can approximate various natural phenomena, Benford's law applies to many natural situations. In other words, for many types of data originating from nature, we are more likely to encounter numbers with smaller leading digits than larger ones.

2.1. Brief history

The first documented and verified evidence for this statistical law originates from Simon Newcomb's discovery in 1881 (Newcomb, 1881). Newcomb, a Canadian-American astronomer, noticed an intriguing detail while working with logarithmic tables. The first pages, particularly those with numbers beginning with the digit 1, appeared significantly more worn than the pages for the higher digits. Since the arguments of a logarithmic table are known to be arranged linearly and uniformly, this observation seemed peculiar.

From this simple observation, Newcomb hypothesised that the use of logarithm tables was skewed toward numbers starting with the digit 1 and less so towards numbers starting with the digit 9. His intuition was correct and he sought to express it mathematically. This is what Newcomb (Newcomb, 1881) wrote in his paper: "That the ten digits do not occur with equal frequency must be evident to anyone making use of logarithm tables, and noticing how much faster the first pages wear out than the last ones. The first significant figure is oftener 1 than any other digit, and the frequency diminishes up to 9."

Newcomb's work is recognised as the first recorded instance of this phenomenon. Following his initial observation, he published a paper proposing a theoretical distribution that applied not only to the first digit but also to subsequent digits. In 1881, in the American Journal of Mathematics (Newcomb, 1881), he proposed that the probability of a number N being the first digit of a naturally occurring set of numbers could be determined by the logarithmic formula:

$$\Pr(N = n) = \log_{10}\left(\frac{n+1}{n}\right) \quad n \in \{1, \dots, 9\}. \tag{1}$$

Why then do we refer to this rule as Benford's law throughout this document and in the related literature? The explanation lies with Frank Benford, an American electrical engineer and physicist, who independently reached the same insight as Newcomb through his own observations of logarithmic tables. In his 1938 paper, Benford noted: "It has been observed that the first pages of a table of common logarithms show more wear than do the last pages, indicating that more used numbers begin with the digit 1 than with the digit 9." (Benford, 1938, p. 551).

Benford analysed 20229 different data values and found results that mirrored those observed by Newcomb decades earlier. Why did Benford's independent findings align so closely with Newcomb's observations? There are several factors that may explain this. For example, in Benford's research, the data and observations spanned different orders of magnitude. Benford also took numbers from various fields and combined them (Miller, 2015, chapter 4). In addition to confirming Newcomb's observations, Benford aimed to provide explanations and justifications for this apparent anomaly.

After forming his hypothesis, Benford tested his law on 20 datasets (Benford, 1938, p. 553), including data on river lengths, mathematical relationships, and populations, all of which yielded results consistent with Newcomb's earlier work. He referred to this numerical behaviour as "The Law of Anomalous Numbers."

However, it is worth noting that Benford did not propose any practical applications or suggestions on how this law could be used. His work was limited to collecting data, analysing it, and ultimately identifying a rule that governs numerical behaviour in the natural world. In his paper, there are no claims regarding how this theory might be applied constructively. The intuition that this strange pattern of numbers could be used for detecting fraud came many years later, by other researchers. A key development came from Nigrini, who summarised the intuition behind his applications: "If there were indeed predictable patterns to the digits in tabulated data, then perhaps auditors could use these expected patterns to test whether the data were authentic or fraudulent." (Nigrini, 2012)

2.2. Mathematical formulation of the law

Before providing a mathematical formulation, we must define what the law states and what it means. We can use the following definition given by Miller (2015, p. 7) in his book: "We say a data satisfies Benford's Law for the Leading Digit if the probability of observing a first digit of d is approximately $\log_{10}\left(\frac{d+1}{d}\right)$" Miller (2015, p. 7).

There is also a heuristic explanation for Benford's law, or more precisely, of Benford's first digit law: "If a random variable $x > 10$ has a distribution ranging over several orders of magnitude, its first leading digit is likely to follow Benford's law" (Miller, 2015, p. 119). Thus, having adopted this definition, we must explain and define what is meant by the first digit and why, in this context, we have chosen to consider only this digit.

The significant digit can be defined as the first non-zero digit (a number from 1 to 9) to the left of a number. It is intuitive to understand that by observing only the first digit, we can compare many numbers without considering their different scales because every number has at least one digit. The sign of negative numbers is ignored, so, for example, the first two digits of the number -345 are considered to be 34.

We also know that any positive number n can be expressed in the form $S(n) \cdot 10^k$ where S is called significand and k is called the exponent. S is a number between 1 and

10, formally $S(x) \in [1, 10)$, and k is an integer. Formally, as Nigrini said in his book (Nigrini, 2012), the first significant digit is:

$$\text{FirstDigit}(x) = \lfloor \text{Significand}(a) \rfloor \text{ where } x = a \times 10^n \text{ (with } 1 \leq a < 10,\ n \in \mathbb{Z}).$$

This formula can also be extended to define digits different from the first digit.

In summary, we want to know the distribution of the leftmost nonzero digit (ranging from 1 to 9) as the first digit.

Another definition comes from Berger and Hill's book (Berger and Hill, 2015): "Informally, the first significant (decimal) digit of a positive real number x is the first non-zero digit appearing in the decimal expansion of x". We can infer that the first significant digits of two different numbers can be the same. For example, the numbers 3017 and 0.00387 have the same first significant digits.

It is also possible to generalise this reasoning for digits that are different from one. We can state that, although the first significant digit is never zero, the second, third, etc., may be any integer number in the set $I = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Since, in our formulation, we only consider real integer numbers, we sometimes need to transform non-integer real numbers into integers by removing the decimal part in order to identify the first significant digit.

The general formula without a numerical base is defined as:

$$\Pr(N = n) = \log_b \left(\frac{n+1}{n} \right), \quad n \in \{1, \dots, 9\}$$

where b is the base in which our numbers are written. For now, we will use only the decimal base, i.e., $b = 10$, and the first digit. So we refer to the formula defined in Equation (1) by Newcomb (Newcomb, 1881):

$$\log_{10} \left(\frac{n+1}{n} \right) \tag{2}$$

To better understand how the mechanism works, we can take, as an example, the number 314.

$$\log_{10} \left(\frac{315}{314} \right) = 0.00138.$$

Thus, the probability that, in a dataset that follows Benford's rule, a number begins with the exact three digits 314 is 0.138%.

We can extend the formula defined in Equation (2) to calculate the exact probability that the second, third, or subsequent digits are equal to a number from 0 to 9. We must introduce the summation sign because we want to know the cumulative probability of having a digit d at position p. The generalised formula for two or more digits is:

$$\Pr(N_p = d) = \sum_{k=10^{p-2}}^{10^{p-1}-1} \log_{10}\left(1 + \frac{1}{10k + d}\right), \quad p \in \mathbb{Z}^+,\ p \geq 2, \quad d \in \{0, 1, 2, \ldots, 9\} \tag{3}$$

Now that we have defined a formulation, we can compute other examples. For example, if we wanted to know the probability of a digit being equal to 1, we would use the formula defined in Equation (2) for the first digit and would easily reach the result of 0.30103 and so 30.1%. Extending this example to the second digit, we would use the formulation in Equation (3). To find the probability that the second digit is equal to 1, we would sum the probabilities of all numbers with the digit 1 as the second digit. The formula would then become: $\sum_{k=10^{2-2}}^{10^{2-1}-1} \log_{10}\left(1 + \frac{1}{10k+1}\right)$. In this formula logarithms from $\log_{10}\left(\frac{12}{11}\right)$ to $\log_{10}\left(\frac{92}{91}\right)$ are included. The result is 0.11389, which means that the probability that the second digit is equal to 1 is 11.39%.

We can use the same formula defined in Equation (3) for the third position, the fourth position, and so on. Applying the formula, we find that from the third position onwards the differences between the digits become smaller, approaching a uniform distribution where all probabilities are 0.1, or 10%.

Table 1 shows the theoretical probabilities for the first four positions, indicated as a column. The digits are all integers from 0 to 9, with the exception of 0 for the first digit (since, by definition, there is no number with a first significant digit equal to 0, except for the number 0 itself). Upon initial inspection, we can see that the theory is confirmed. The table shows that as we move from left to right, the digits tend toward being evenly distributed. For numbers with four or more digits, we consider the distribution to be uniform.

In Figure 1, three histograms are shown with the frequencies of the digits from 0 to 9 for the first, second, and third digit positions. We can observe that, for the first digit, the distribution is asymmetric. The bars corresponding to values from 1 to 4 have higher frequencies compared to those from 5 to 9. However, this asymmetry and difference become less important as the positional digit increases. As we can clearly see from the graph, by the second position, the distribution is less asymmetric, and by the third position, it is almost uniform. The black lines represent the frequency of probability for a uniform distribution, which is fixed at $1/9$ (approximately 0.11) for the first digit (since it cannot be 0) and at 0.10 for the second and third digits.

Digit	**First position**	**Second position**	**Third position**	**Fourth position**
0	0.000	0.120	0.102	0.100
1	0.301	0.114	0.101	0.100
2	0.176	0.109	0.101	0.100
3	0.125	0.104	0.101	0.100
4	0.097	0.100	0.100	0.100
5	0.079	0.097	0.100	0.100
6	0.067	0.093	0.099	0.100
7	0.058	0.090	0.099	0.100
8	0.051	0.088	0.099	0.100
9	0.046	0.085	0.098	0.100

Table 1: Theoretical probabilities of the Benford's law

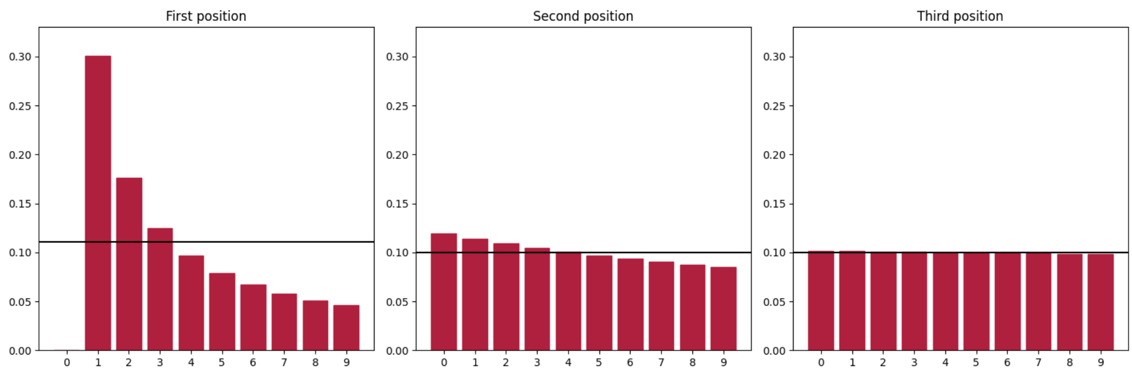

Figure 1: Probabilities for the first three digits.

The Benford's Law proportions are irrational numbers. For that reason, it is not possible to express them as a set of fractions that simply give us the exact numbers of the law. However, as the number of rows in a dataset increases, we can get closer to the exact Benford proportions. This is one of the reasons why we need a huge amount of data. The literature suggests using a large data table with numbers that have four or more digits for a good fit, even if we do not have a fixed minimum number of rows (Nigrini, 2012). A suggestion of a possible minimum requirement comes from Nigrini (2012), who writes in his book that the data set should have at least 1000 rows to expect good compliance.

2.3. Some known applications

The applications of Benford's rule are remarkably broad. Since in many real-life datasets the leading significant digits tend to be heavily skewed toward smaller digits, there are numerous examples of data that can be shown to conform to Benford's law. As Professor Raimi observed in 1976 (Raimi, 1976, p.521): "This particular logarithmic distribution of first digits, while not universal, is so common and yet so surprising at first glance that it has given rise to a varied literature, among the authors of which are mathematicians, statisticians, economists, engineers, physicists, and amateurs." This observation is well supported in the literature, and we know the applications of the First Digit Law are wide-ranging. The law can be used to analyse various numerical series, such as the Fibonacci sequence (Miller, 2015, chapter 1), some health-related data, such as cancer incidence rates (Crocetti and Randi, 2016), and many others. This wide applicability also extends to fraud detection in fields such as accounting (auditing), scientific research, and electoral data (Nigrini, 2012). We also know that Benford's law has been included in the ordinary audit check; for example, it is inside certain software such as IDEA and ACL, which are two audit software platforms (Nigrini, 2012). Many sequences were tested, but the conclusion is that there are more natural sequences in the world that do not respect Benford's law than conform to Benford (Nigrini, 2012). One example of a sequence that does not follow Benford's law is the sequence of prime numbers.

Empirical evidence supporting Benford's law continues to grow. For example, Becker (1982) identified conformity with Benford's law in the decimal part of the failure rates and the Mean Time To Failure (MTTF) values listed in the tables. Similarly, studies on stock exchange indices, such as the Dow Jones Industrial Average Index (DJIA) and the Standard & Poor's Index (S&P), have demonstrated their compliance with Benford's law (Ley, 1996). One particularly intriguing application comes from Mark Nigrini's work (Nigrini, 1996) on fraud detection: he examined more than 90000 entries for "Interest Received" in US tax returns from IRS Individual Tax Model Files and found that they adhered to Benford's law. Beyond these, more advanced fields such as time-variant processes, multi-dimensional processes, forecast evaluations, and anomaly detection in images also exhibit coherence with Benford's law (Berger and Hill, 2015). However, these applications are outside the scope of this document.

Nevertheless, there does not exist a general criterion that fully explains when and why

Benford's law holds for a generic set of data (Campanelli, 2022). For that reason, it is not possible to know in advance if a set of data is expected to follow Benford's law or not, but there are some conditions that make it easier.

2.3.1. Fraud Detection and the work of Mark Nigrini

One of the greatest contributors to this subject is Mark Nigrini, who has worked on numerous applications of Benford's law and in particular in fraud detection analysis of accounts payable data. His research is summarised in a section of Miller's book (Miller, 2015, Chapter 8.3.4). He analysed US taxation data and found certain anomalies. Specifically, he identified in a dataset of US declarations some spikes: instances where the actual proportion of first-two digits exceeded the expected proportion by a significant margin. These spikes prompted investigators to examine the specific numbers causing the anomalies, focussing on the frequently repeated values. The excessive occurrences of particular numbers in that type of data can be a sign of inefficiencies, errors, biases, or fraud. Nigrini identified, for example, spikes around $\$10$ and $\$100$ invoices. These anomalies, upon further analysis, were found to correspond to freight charges ($\$10$ invoices) and travel advances for sales staff ($\$100$ invoices). A second spike, observed at $\$15$ and $\$150$, was traced to $\$15$ freight invoices and $\$1500$ programmer bonuses for meeting deadlines. In the end, that analysis did not reveal fraud, but provided insight into underlying business practices. This is to say that Benford's law may raise some suspicions but these are not sufficient to identify and confirm fraud. They can only serve as indications. Sometimes, however, frauds detected through this rule were later confirmed by a more in-depth analysis. Nigrini reported three cases of fraud that were later appropriately verified and confirmed by investigations (Miller, 2015).

2.4. Conditions that make a dataset a good candidate to be Benford

As we have written in Section 2.3, the topic has broad applications. But there are also many data sets or numbers that do not follow Benford's law in any sense. Telephone numbers in a specific region, lottery numbers, or human weight tables are typically distributed in another way. This also occurs for "pure" mathematical data such as square-root tables of integers, prime numbers, and others. But why are there some datasets that follow the First-Digit Law and others that do not? Are there any types of circumstances that are helpful in predicting which empirical data follow Benford's law?

We know that this disparity arises, for example, when measurements are taken on a scale confined to a specific range rather than spanning multiple orders of magnitude (Miller, 2015, p. 10-12). Examples of this include temperature readings, heights of individuals, university exam scores, and weights. Such data sets typically do not follow Benford's law because their values do not span the wide range of scales necessary for the law to manifest. We also know from the literature that, as defined in Chapter 1.5 of Miller's book (Miller, 2015, p. 8-16) and in Chapter 5 of Hill and Berger's book (Berger and Hill, 2015, chapter 5), there are some conditions that help to make a data set more likely to follow Benford's law. These conditions can be summarised in three points:

1. **Spread hypothesis**: If a dataset is distributed over several orders of magnitude, it is more likely to be a Benford dataset. This is a useful hypothesis but not a certainty. There are datasets that follow this condition but that are not coherent with Benford's law. Consider, for example, a dataset containing only numbers with the first digit equal to 1 but which is distributed across different orders of magnitude. We can see that this dataset satisfies the above-mentioned condition but not Benford's rule. Explicitly, we can think of this dataset as

$$\text{dataset} = \left\{1, 10, 100, 1000, \ldots, 10^{500}\right\}.$$

   Instead of 500, we could have written whatever number we wanted, up to infinite. It is easy to notice that, even if this is a spread dataset, this does not follow Benford's law. All the numbers contained have in fact the first significant digit equal to 1. So, by necessity, we need other conditions or different explanations.

2. **Geometric explanation**: If we have a process with a constant growth rate, the speed of the increase (which in this case we take as constant) will cause more time spent on lower digits than on higher ones. For instance, if we have a process that doubles the numbers, moving from 1 to 2 takes the same time as moving from 1000 to 2000 or from 1000000000 to 2000000000. This phenomenon helps explain why real-world datasets often follow Benford's Law and why lower digits appear as the leading digits more frequently.

3. **Scale-invariance explanation**: This condition was first time imagined by S. Newcomb, who wrote in his paper that the behaviour of the numbers should be indepen-

dent of the unit used (Newcomb, 1881). The idea is that if there exists a fixed distribution law that is universal and applies to different datasets, this law must surely be independent of the system of units chosen. Since measurement systems were invented by humans, and measurement in either the metric system or the imperial system essentially involves multiplication by a constant, that rule should apply to all existing measurement systems. For example, whether we use euros or dollars, it does not matter for our reasoning. This line of reasoning was well explained also by Pinkham in 1961 (Pinkham, 1961), who appropriately dissected this rule with a mathematical formulation, showing the invariance property. Pinkham demonstrated that being Benford is not only necessary but also sufficient for X to have scale-invariant significant digits. The theorem attributed to Pinkham (1961) and also reported by Nigrini (2012) is: " If the numbers x_1, x_2, x_3,..., x_N in a data field conform to Benford's law, any new field formed by multiplying the x_i values by a nonzero constant c will also conform to Benford's law."

In addition to these more formal tests and definitions provided by Pinkham (1961), we can verify the *scale-invariance* explanation by using a simple test. We can simply compare the proportion of numbers beginning with digits from 1 to 10 before and after the change of the scale. Recall that scaling is nothing more than a multiplication of numbers by a constant. If both datasets, the one before and the one after the rescaling, follow Benford law, we have a hint that the dataset could follow Benford's distribution.

We can use this rule also for data that does not have a unit of measurement like population counts or vote counts. We can simply multiply our number by a constant, let's say π, to verify if this condition is also respected in the dataset that apparently has no unit of measurement.

We also know that the combination of different datasets or the combination of different distributions is a condition that leads to a database consistent with Benford's law. As Hill wrote in his book (Berger and Hill, 2015): "Not only do iterations of most linear functions follow Benford's law exactly, but iterations of most functions close to linear also follow Benford's law exactly … Similarly, if random samples from different randomly-selected probability distributions are combined, the resulting meta-sample also typically converges to Benford's law"

2.4.1. *Nigrini's conditions*

In addition to the condition that we wrote above, Professor Mark Nigrini set out three requirements that the data must meet in order to conform to Benford's law (Nigrini, 2012):

1. The data must represent the sizes of facts or events. Examples of such data include populations, sizes, lengths of rivers, market values of companies, revenues or costs, and many others.

2. There should be no built-in minimum or maximum values in the dataset. A built-in minimum of zero is acceptable. For instance, the height of a person does not follow this condition because we know for certain that the height of a person cannot be lower than a certain value and cannot exceed another value. Another example would be to consider bank fees. If we think of a bank that charges a fixed fee of 2 euros per month plus a variable part depending on the amount of transactions, we know that even if the customer of the bank does not make a transaction, the fee charged is 2 euros. A data set derived from this bank would have an excess of numbers beginning with the digit 2.

3. The data set should not represent numbers used as identification numbers or labels. These numbers are assigned to things for identification purposes and do not follow a Benford distribution; instead, they exhibit digit patterns with specific meanings for the person who developed the sequence. Examples of labels or identification numbers include phone numbers, university roll numbers, encrypted keys, bank account numbers, and many others.

4. The mean value should be less than the median value and the data should not be tightly clustered around an average value. This rule means that there are more small records than large records in Benford's datasets.

2.4.2. *Non-decimal bases*

Throughout this thesis, we will primarily focus on the decimal base. However, this is not the only possible interpretation of the law. The significant digit law can also be generalised for other bases b different from 10. As Raimi wrote (Raimi, 1976) every argument that applies to base 10 also applies to base b *mutatis mutandis*. We can substitute the base 10 with another arbitrary integer $b \geq 2$. In particular, the general form of (3) with respect to any such base is as follows:

$$\Pr(D_1^{(b)} = d_1, D_2^{(b)} = d_2, ..., D_m^{(b)} = d_m) = \log_b \left(1 + \left(\sum_{j=1}^{m} b^{m-j} d_j \right)^{-1} \right) \tag{4}$$

where $\log_b$ denotes the base-b logarithm, and $D_1^{(b)}$, $D_2^{(b)}$, etc., are the first, second, third, etc., significant digits in base b, respectively; so d_1, d_2, d_3, etc., are digits in $\{1, ..., b-1\}$, and for $j > 1$, d_j is an integer in $\{0, ..., b-1\}$ (Berger and Hill, 2015). Also in Kossovsky (2014) this generalisation is highlighted. Kossovsky define that Benford's law is valid for any base using a simple generalisation of the base formula. Of course, this extension is trivial for the base 2, since the probability of a binary number beginning with the digit 1 is obviously 100%, but it may be of interest for other bases, and may be a further check or method to see if our data deviates from Benford's law. However, the basic assumption of Benford's law remains valid: the proportions of possible values that a digit can take have decreasing probability of appearing, even if the base is changed.

2.5. Statistical test

In this subsection, we briefly present some useful statistical tests that will be used later in the document. These tests can be applied to our datasets to determine whether the data are coherent with the theoretical law. Firstly, a simple comparison between our data and Benford's law that returned a low correlation would be a warning sign, indicating the risk that the data might contain strange duplications or anomalies. However, to obtain statistical confirmation, we must apply some statistical tests.

The use of some statistical tests is quite recent. As Professor Nigrini said in his book (Nigrini, 2012), one of the first documented applications of statistical tests to that law was made by Thomas in 1989 (Thomas, 1989) who analysed 80000 earning numbers with Z statistic. Prior to Thomas (1989)'s work, the researcher never used tests to check the coherency because they either used some sort of professional judgment or, with some clear conform or nonconform situations, a statistical test was not needed. They also often worked with a small set of data when assessing the conformity case by case, and watching all data entries was possible.

Anyway, if we are searching for the lowest level of aggregation possible to better determine whereas the coherency is just a coincidence or not, we need the largest amount of data available and, as a consequence, some statistical tests that can help us in the analysis.

It is important to keep in mind that no dataset perfectly adheres to Benford's Law. This is because datasets are fundamentally finite, whereas Benford's Law is defined using an irrational distribution that extends over infinite ranges (Newcomb, 1881; Benford, 1938). In practice, datasets are naturally limited, either because they represent a sample of the population or are restricted to a specific time frame. Consequently, even data that generally follow Benford's law have inevitably some level of divergence.

However, even in the case of very large data sets with many rows and an important level of detail, some problems can occur. Statistical tests are generally very sensitive to the sample size N and their enormous power for larger N can make the result of the test inappropriate for understanding whether some data belong to Benford's distribution or not. In these cases, the tests are too rigid to assess goodness of fit. This rigidity affects even a tiny deviation in the digit count in larger datasets, which will be statistically significant. This is a problem of all statistical tests as noticed by Campanelli (2022): "we know that for very large sample size ($N > 1000$), all existing statistical tests are inappropriate for testing Benford's law due to its empirical nature".

Another issue is related to the definition of these tests and the related distributions. We know that the Chi-squared test and the Kolmogorov-Smirnov tests are based on the null hypothesis of a continuous distribution. For that reason, they are generally conservative for testing discrete distributions such as Benford's one (Noether, 1963).

In summary, the reasons why statistical tests may not return an adequate result consistent with graphical analysis are as follows. The fact of testing discrete data in a test constructed for a continuous distribution, the problem of the power of the test with very large sample sizes ($N > 1000$), and the fact that Benford's law is not a limiting distribution but is applied to a finite set of data. It is important to note that if N is large, the problem of discretization becomes less important. In other words, two of the issues mentioned are mutually exclusive: one makes the statistical hypothesis more conservative, while the other increases the likelihood of rejecting the null hypothesis (higher power problem).

We know that none of the classical distributions we commonly use, such as uniform, exponential, Pareto, normal, beta, binomial, or gamma distributions, are Benford (Berger and Hill, 2015). Therefore, we will use the Benford law formula as the definition of a new distribution. In the remainder of this document, we will refer to it as the Benford distribution. This distribution is defined as follows and satisfies two fundamental conditions:

Non-negativity: The probability associated with each possible outcome must always be greater than zero:

$$\Pr(D_i = d_i) > 0 \quad \forall d_i \in S_{D_i}$$

where S_{d_i} is the support of the random variable D_i, which can be the first, second digits, or both. Every event (every digit in the defined set) must have an associated probability greater than 0.

Normalization: The sum of the probabilities for all possible outcomes must be equal to 1:

$$\sum_{d_i \in S_{D_i}} \Pr(D_i = d_i) = 1.$$

The formulas for the probabilities of the digits are defined as follows:

1. **First digit (d_1)** The first significant digit, d_1, is the leading non-zero digit of a number. The set of possible values for d_1 is:

   $$d_1 \in \{1, 2, \ldots, 9\}.$$

   The probability of observing a specific first digit d_1 is given by:

   $$P(d_1) = \log_{10}\left(1 + \frac{1}{d_1}\right).$$

2. **Second digit (d_2)** The second significant digit, d_2, is the digit in the second position of a number. The set of possible values for d_2 is:

   $$d_2 \in \{0, 1, \ldots, 9\}.$$

   To compute d_2, a first digit d_1 must exist and belongs to the set:

   $$d_1 \in \{1, 2, \ldots, 9\}.$$

   The probability of observing a specific second digit d_2 is defined as:

   $$P(d_2) = \sum_{d_1=1}^{9} \log_{10}\left(1 + \frac{1}{10d_1 + d_2}\right).$$

3. **First two digits ($d_1 d_2$)** The first two significant digits, $d_1 d_2$, are the leading two digits of a number. The set of possible values is:

$$d_{12} \in \{10, 11, \ldots, 99\}.$$

The probability of observing a specific combination of the first two digits $d_1 d_2$ is:

$$P(d_{12}) = \log_{10}\left(1 + \frac{1}{d_{12}}\right).$$

We can formally define a hypothesis test to compare Benford's distribution with empirical data. We define the null hypothesis as the hypothesis in which we do not have statistical evidence of significant differences between Benford's distribution and our dataset. In this hypothesis, data may be generated from the Benford's distribution. Otherwise, if we reject the null hypothesis at a certain level of significance, we can say that our data and Benford's distributions are statistically different.

The hypotheses for the test are defined as follows:

- **Null hypothesis** (H_0): The distribution of the leading digits of our dataset conforms to Benford's distribution.

$$H_0 : \text{The dataset conforms to Benford's distribution.} \tag{5}$$

This hypothesis cannot be rejected if the p-value of the test satisfies:

$$p\text{-value} > \alpha$$

where α is the significance level (e.g., $\alpha = 0.001, 0.01, 0.05$).

- **Alternative hypothesis (H_1)**: The distribution of the leading digits of our dataset does not conform to Benford's distribution.

$$H_1 : \text{The dataset does not conform to Benford's distribution.} \tag{6}$$

We reject the null hypothesis (H_0) and so accept the alternative hypothesis (H_1) if:

$$p\text{-value} \leq \alpha$$

.

The purpose of this hypothesis test is to check whether there are anomalies or tampering in the data through the rejection of the null hypothesis H_0 as we assume that the electoral datasets conform to the Benford distribution. The focus of the test is on minimising the type I error $alpha$ and we want to determine whether there is sufficient evidence to reject the null hypothesis. It is important to note that failing to reject the null hypothesis does not mean accepting or proving it to be true. It indicates that there is insufficient empirical evidence to conclude that the dataset does not conform to Benford's distribution. In other words, in the statistical tests that follow, we can only fail to reject the hypothesis that our data are coherent with Benford's law or reject it, but we cannot confirm its validity by accepting it. The construction of this test is based on the assumption that electoral data follow Benford's law, as confirmed by Mebane Jr's multiple works, (Pericchi and Torres, 2004) and many others. Our objective is therefore to understand and identify whether indications of fraud, tampering or error can be found in an election dataset, which is supposed to follow Benford's law.

2.5.1. Z Statistic

This test is used to verify if a digit's proportion in a data set differs significantly from the expectation of Benford's law. This test is performed digit by digit, and his usage was suggested by Nigrini (2012); Kossovsky (2014). The formula adapted from Nigrini (2012) is:

$$Z = \frac{|DP - BP| - \frac{1}{2n}}{\sqrt{\frac{BP(1-BP)}{n}}} \tag{7}$$

where $|\cdot|$ stands for the absolute function and:

- DP: Denotes the Dataset Proportion, which represents the proportion of the digits observed in our dataset.
- BP: Denotes the Benford's law proportion, which is the theoretical proportion derived from Benford's law and serves as the reference for comparison. We take the numbers of the formula (1) for the first digits and of Equation (3) for two or more digits.

- n: Represents the number of records. This variable represents the number of elements that we have in our dataset.

The numerator has an absolute function, as in the Mean Absolute Deviation metric defined in Subsection 2.6.1, which means that every deviation counts in the same way without taking into account the negative or positive sign. In the numerator, the term $\frac{1}{2n}$ is a continuity correction term. This is used only if it is smaller than the difference between the Dataset Proportions (DP) and the Benford Proportions (BP), the first term in the numerator. The Z statistic becomes larger for greater differences between DP and BP, i.e. evidence against the null hypothesis.

The Z-statistic follows a standard normal distribution $\mathcal{N}(0,1)$ under the null hypothesis. This means that under the null hypothesis H_0 defined in (5), the Z statistic will be normally distributed with a mean of 0 and a variance of 1. Consequently, deviations from Benford's law can be evaluated using the quantiles of the standard normal distribution.

The results of Z statistic tell us if the Dataset Proportion deviates significantly from Benford's proportion. The deviation is statistically significant if, at a chosen level of α, we can reject the null hypothesis H_0 defined in (5).

If the p-value derived from the Z statistic is smaller than or equal to the chosen level of α, we can reject the null hypothesis H_0. This implies that our data do not belong to Benford's law. Otherwise, if the p-value is greater than α, we can not reject the null hypothesis, and so we cannot exclude that our dataset respects Benford's distribution.

In the Z statistic, we can encounter the excess power problem: for datasets with a lot of observations, even small differences or spikes are highlighted as significant and may invalidate our test (Campanelli, 2022; Kossovsky, 2014; Nigrini, 2012) as stated previously.

2.5.2. Chi-squared test

In the literature, the use of the χ^2 test is widely prevalent to analyse contingency tables. This test is also the most common for testing if a dataset is coherent with Benford's law (Campanelli, 2022). To measure the discrepancy between observed and expected frequencies for the first digits, the chi-squared test equals:

$$\chi^2_{1st} = \sum_{j=1}^{9} \frac{(n_j - Nr_j)^2}{Nr_j} \tag{8}$$

where $j \in \{1, \ldots, 9\}$ represents the value of the first digit, n_j is the observed frequency with digit j as the first digit, N is the total number of values with one or more digits, and r_j is the expected proportion of occurrences of digit j as the first digit, according to Benford's Law, i.e., formula (2).

The number of degrees of freedom is the maximum number that j can assume minus one. For the first digit, this value is 8. Under the null hypothesis, χ^2_{1st} test statistic follows a chi-squared (χ^2) distribution with $k-1$ degrees of freedom:

$$\chi^2_{1st} \sim \chi^2_8.$$

Rejecting H_0 suggests that the data do not follow Benford's Law, indicating potential anomalies or deviations in the dataset.

For analysing the second digit or higher-order digits, the formula defined in Equation (8) becomes:

$$\chi^2_{2nd} = \sum_{j=0}^{9} \frac{(n_j - Nr_j)^2}{Nr_j} \tag{9}$$

which measures the discrepancy between observed and expected second or higher digit frequencies. Here, $j \in \{0, \ldots, 9\}$ represents the value of the second or higher digit, n_j is the observed frequency of values with the digit j, N is the total number of values with two or more digits, and r_j is the expected frequency of the digit j according to Benford's law, i.e. formula (3).

Under the null hypothesis H_0, the test statistic χ^2_{2nd} follows a chi-squared distribution with $k-1$ degrees of freedom:

$$\chi^2_{2nd} \sim \chi^2_9.$$

Instead, for analysing the first-two digits, the test statistic follows the same structure but considers all combinations of the first two digits:

$$\chi^2_{first2} = \sum_{j=10}^{99} \frac{(n_j - Nr_j)^2}{Nr_j}. \tag{10}$$

In this case, the test statistic follows under the null hypothesis:

$$\chi^2_{first2} \sim \chi^2_{89},$$

where 89 represents the degrees of freedom, calculated as the total number of possible combinations of digits (90) minus 1.

The p-value associated with the χ^2 test statistic is computed using the cumulative distribution function (CDF) of the χ^2 distribution. The p-value is then compared with the significance level α chosen, which are 0.001, 0.01 and 0.05:

- If p-value $> \alpha$, we fail to reject the null hypothesis H_0 defined in (5), meaning that the dataset is coherent with Benford's law.
- If p-value $\leq \alpha$, we reject the null hypothesis H_0, concluding that the dataset does not conform to Benford's law.

The χ^2 test is, as previously written, the most widely used method for comparing actual data with expected data. However, we know from its definition that it is highly sensitive to large samples, with the drawback that even small deviations from the target distribution can be statistically significant. Like the Z statistic, this test suffers from an excess power problem. When the observations or the data in a dataset become a lot, let us say more than 5000, the chi-squared test will almost always reject the null hypothesis H_0, leading us to conclude that the data do not conform to Benford's Law. In his book Professor Mark Nigrini said that "the chi-squared test is too sensitive to small deviations from Benford's law for large data sets to be useful for a conformity assessment. The test is not really of much help in forensic analytics because usually we are dealing with large data sets"(Nigrini, 2012). Also Campanelli (2022) suggests that "the use of the χ^2 statistic for checking the conformance of a set of data to Benford's law is not completeley reliable and should be used only for "qualitative" analysis."

2.5.3. Kolmogorov-Smirnov test

The use of the Kolmogorov-Smirnov (K-S) test in fraud detection and in verifying the coherence of a dataset with Benford's law is well defined in the literature (Nigrini, 2012). This is a nonparametric method that is used to compare two distributions and to see if one differs from the other. This test is based on the cumulative density function and, as in the chi-squared test, allows us to determine whether observed data does not conform to Benford's law.

The K-S test uses the maximum absolute difference between the cumulative distribution function of Benford's law and the cumulative distribution function of our electoral data. Like in the previous cases, the null hypothesis H_0 is that the two distributions are the same, rejecting that means that our data do not follow Benford's distribution. The K-S test is applied to the significant digits we want to verify. In our specific case, the test is applied to the first significant digit, the second significant digit, and the first two significant digits. The largest absolute value of these differences is also called the *supremum*. Formally, the test statistic is defined as:

$$KS = \sup_x \left| F_{\text{obs}}(x) - F_{\text{Benford}}(x) \right|, \tag{11}$$

where:

- $F_{\text{obs}}(x)$ is the cumulative distribution function (CDF) of the observed data.
- $F_{\text{Benford}}(x)$ is the cumulative distribution function (CDF) of Benford's law.
- $\sup_x$ represents the supremum (largest absolute value) of the differences on all possible values of x.

Under the null hypothesis H_0, the test statistic K-S follows the Kolmogorov distribution. For a dataset of size N, the test statistic is scaled by the square root of N:

$$\sqrt{N} \cdot KS \sim K, \quad \text{as } N \to \infty \tag{12}$$

where K represents the Kolmogorov distribution.

The p-value for the Kolmogorov-Smirnov test is calculated using the Kolmogorov distribution under H_0. The p-value is then compared with the significance level α to decide whether to reject the null hypothesis:

- If p-value $> \alpha$, we fail to reject H_0, meaning that the data give no evidence of not following Benford's law.
- If p-value $\leq \alpha$, we reject H_0, concluding that the data do not follow Benford's law.

This test, like the z statistic and the chi-squared, also suffers from the problems defined before (Campanelli, 2025). Specifically, the K-S test assumes that the underlying cumulative distribution function (CDF) is continuous; however, we have discrete data and a discrete distribution. So, the discrete distributions are treated as if they were continuous, and then the K-S test is applied to these cumulative values. This solution introduces some level of inaccuracy (Noether, 1963). Another problem from which the K-S test suffers, while remaining a useful tool, is the excess power problem. When the number of rows N becomes large, even small deviations from Benford's law are flagged as significant, even if the visual representation looks acceptable. As Professor Nigrini stated in Nigrini (2012): "The inclusion of N in the formula makes the K-S test inappropriate in real-world forensic settings."

2.5.4. Saville Regression Measure

This test was suggested by A. Saville in 2006 to detect data errors or fraud in Johannesburg Stock Exchange (JSE) listed companies (Saville, 2006). Saville's regression is included in the section on statistical tests because with a suitable extension it could be adapted so that the variations of the points in our dataset with respect to those in Benford's law are tested statistically.

This regression was applied to the first digit, but it can easily be extended to the second or first-two digits. It is also independent of data size N. The assumption behind Saville's algorithm is that a simple regression analysis can be employed to assess if there are significant deviations between empirical data and Benford's law. Regression line, for this purpose, is estimated on the form:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \tag{13}$$

where Y_i (dependent variable) is the relative frequency of the significant digit i of electoral data, β_0 is the intercept, β_1 is the slope and X_i (independent variable) is the relative frequency of the significant digit i of Benford's law. ϵ_i are the uncorrelated random errors distributed with a Normal distribution with mean 0 and variance σ^2. The error terms are independent and identically distributed (i.i.d.). A perfect correlation between the data and Benford's law occurs when $\beta_0 = 0$ and $\beta_1 = 1$, so we can define this situation as our null hypothesis H_0.

$$H_0 : \beta_0 = 0 \quad \text{and} \quad \beta_1 = 1 \tag{14}$$

In fact, in his paper, Saville suggested that one focus only on the slope and intercept output to make a decision. He used a t-test with a 5% confidence level for slope and intercept values. The suggestion is to focus only on the dissimilarity between the regression line and the bisector, the ideal Benford line $Y = X$. The relationship between the points and the bisector was not explored. However, this method is contested by Kossovsky (2014), which, with several practical applications, has shown how the distribution of points around the regression line is also crucial for understanding whether a dataset is consistent or not with Benford's law. Specifically, Alex Kossovsky highlighted how some datasets have $\beta_0 = 0$ and $\beta_1 = 1$ but with important deviations between Benford's law and the empirical data. These differences can be highlighted only with the graph observation: points of relative frequencies are spread around the graph and far from the regression line and the bisector (that, with this, values of β_0 and β_1 are overlapped).

Therefore, proximity to Benford cannot be measured only in terms of β_0 and β_1. Slope and intercept alone are not sufficient to conclude that the Saville test indicates that our dataset is close to Benford's distribution. The variation of points around their regression line and the bisector (ideal Benford line) also have to be taken into consideration.

In any case, the absence of N prevents us from using the Saville regression as a compliance test. The use of N is discussed and for this reason we preferred to avoid it. The partial result obtained by observing β_0 and β_1 leads us to combine the Saville regression with appropriate metrics, which are explained in the following, to get a more reliable conclusion.

2.5.5. Bonferroni's correction

When we compute the three statistical tests that we have already described, we perform the statistical analysis on the same sample. For that reason, our tests must be corrected to take into account the fact that we are using the same data to test our hypothesis. This problem is called the multiple test issue, highlighted by Goeman and Solari (2014) as many others.

Obviously, every hypothesis test has some errors. We say that the type I error is the risk of falsely rejecting a hypothesis that is true, and that the type II error is the risk of not rejecting a hypothesis that is false. In this document our attention is focused on the type I error, the risk of rejecting the hypothesis that our data are consistent with the Benford distribution (H_0) (5), even if it is true. A type I error means that we are signaling a dataset as suspicious even if it follows Benford's law. In every hypothesis test, an acceptable level oh alpha (α) is chosen; in our case, we focus our attention on three levels: 0.001, 0.01, and 0.05. Our goal is to test at least three hypotheses simultaneously from the same data, which are:

- $H_0^{(1)}$: The distribution of the first digit follows Benford's law.
- $H_0^{(2)}$: The distribution of the second digit follows Benford's law.
- $H_0^{(12)}$: The distribution of the first two digits follows Benford's law.

The simultaneous test of three null hypotheses leads, as mentioned above, to certain problems. As Goeman and Solari argued in their paper, "Because each test again has a probability of producing a type I error, performing a large number of hypothesis tests virtually guarantees the presence of type I errors among the findings".

For that reason, we need to introduce a correction that allows us to control Type I errors when more than one hypothesis is performed simultaneously on the same sample. The objective of multiple tests is to reduce the error probability of type I back to alpha even for multiple hypotheses.

A formalisation of the problem encountered is needed. Starting with the Chi-squared test and the K-S test, we have a multiple testing problem with a simple structure because these tests evaluate the entire probability distribution at once, rather than comparing individual values separately. We have a collection $H = \{H_0^{(1)}, H_0^{(2)}, H_0^{(12)}\}$ of null hypotheses, one for each aspect of the dataset that we want to assess for conformity with Benford's Law. In this scenario, the null hypotheses represent our starting assumption, which "we do not want to reject". An unknown number $m_1 \in \{0, 1, 2, 3\}$ of these null hypotheses may be false, while $m_0 = m - m_1 \in \{0, 1, 2, 3\}$ are true where m is the total number of hypotheses, i.e., $m = 3$. If we have p-values p_1, p_2, p_3 for the hypotheses $H_0^{(1)}, H_0^{(2)}, H_0^{(12)}$, the objective is to ensure that the decision to not reject these hypotheses is valid and properly corrected for multiple tests. We start from a collection of S test statistics, one for each hypothesis tested, with the corresponding p-values p_1, p_2, p_3. These p-values are called raw p-values or not corrected, as they do not account for the fact that multiple tests are being performed simultaneously.

When multiple hypotheses are tested, the Family-wise Error Rate (FWER) quantifies the probability of making at least one Type I error (false positive) among all tests. Mathematically, it is defined as:

$$\text{FWER} = P(\text{at least one Type I error}) = P(V > 0)$$

where V is the number of Type I errors. This means that the FWER measures the likelihood of incorrectly rejecting at least one true null hypothesis within the family of hypotheses that are being tested.

One way to control this issue is the Bonferroni correction, which adjusts the significance level to control the FWER. In the Bonferroni correction, instead of testing each hypothesis against the original significance level (α), the threshold is adjusted to take into account the number of tests performed (m). The new adjusted significance level α is calculated as:

$$\alpha_{\text{adjusted}} = \frac{\alpha}{m}.$$

For example, if we are testing $m = 3$ hypotheses on the same sample (first, second, and first-two digits) and the chosen significance level is $\alpha = 0.05$, the adjusted significance level becomes:

$$\alpha_{\text{adjusted}} = \frac{0.05}{3} \approx 0.0167$$

Without loss of generality, denote $H_0^{(12)}$ as $H_0^{(3)}$. Each hypothesis is then tested against this adjusted threshold. Specifically, for each hypothesis $H_0^{(i)}$, we reject the null hypothesis if:

$$p_i \leq \alpha_{\text{adjusted}}$$

where p_i is the p-value of the i-th hypothesis and $i \in \{1, 2, 3\}$.

The same type of reasoning must be taken into account in the case of Z-statistics. In this context, we have many more null hypotheses that produce different Z scores. For the first digit, the Z test evaluates 9 null hypotheses, 10 null hypotheses for the second digit, and 90 null hypotheses for the first-two digits.

All of these hypotheses are evaluated on the same data sample, and for that reason, the total number of null hypotheses (m) that must be considered is 109. Again, we want to fix the problem of multiple testing by adjusting the increased probability of Type I errors due to multiple comparisons with a correction: Bonferroni's correction. Taking as an example the significance level of $\alpha = 0.05$, the adjusted significance level becomes:

$$\alpha_{\text{adjusted}} = \frac{0.05}{109} \approx 0.000459$$

Thus, each null hypothesis is tested against this adjusted threshold. Specifically, for each hypothesis $H_0^{(i)}$, we reject the null hypothesis if:

$$p_i \leq \alpha_{\text{adjusted}}$$

where p_i is the p-value associated with the i-th hypothesis and $i \in \{1, 2, ..., 109\}$. This ensures that the probability of making at least one Type I error across all 109 tests remains controlled under the desired significance level.

This type of adjustment is the simplest, oldest and most well known multiple testing method and ensures that the total probability of making at least one Type I error (i.e., FWER) in all tests does not exceed the original significance level (α). However, it is known to be conservative, particularly when the number of tests (m) is large, leading to an increased risk of false negatives (Type II errors) while effectively controlling the FWER (Goeman and Solari, 2014). There are also other types of FWER control methods such as Holm-Bonferroni, Hochberg, and Hommel. These methods are generally less conservative than Bonferroni but still aim to control the FWER. All of these, in general, have the characteristic of being quite stringent, especially when the number of tests (m) is large, so they can lead to many false negatives. Otherwise, alternative approaches, such as controlling the False Discovery Rate (FDR), may be more suitable if the aim is to reduce or balance the rate of false negatives.

2.6. Metrics

In this subsection, three useful metrics are presented to evaluate whether our distribution is coherent or not with Benford's law. In the literature, there are established threshold or pivot values for these three metrics (Grendar et al., 2007; Nigrini, 2012; Kossovsky, 2014), which we use as reference benchmarks to compare with our results.

2.6.1. Mean Absolute Deviation

This simple measure helps us to understand whether a dataset is coherent with Benford's law. The use of this is suggested by Professor Nigrini in his book (Nigrini, 2012). It is important to note that Mean Absolute Deviation (MAD) ignores the number of rows in a data set. This is particularly useful because, in natural datasets, we rarely have access to complete data on a phenomenon and must instead rely on a sample. Specifically, in the case of electoral data, we do not have the smallest aggregation possible, i.e., the polling station level, but only some aggregation for municipalities or statistical zones.

The MAD is calculated using the formula below:

$$\frac{\sum_{i=1}^{K} |DP_i - BP_i|}{K} \tag{15}$$

where:

- DP: Denotes the Dataset Proportion.
- BP: Denotes the Benford's law Proportion.
- K: Represents the number of digits to compare. For example, it is equal to 9 for the first digit, 10 for the second digit, and 90 for the first two digits.

The numerator has an absolute function, which means that every deviation counts in the same way without taking into consideration the negative or positive sign. The denominator is needed to compute the mean because we want to know the average deviation between all the digits. In this measure, a low level of error indicates that our data sample closely approximates the value of Benford's distribution, and so the forecast can be relied on. If, instead, the MAD is higher, we have, on average, a large difference between the Dataset Proportion and the Benford Proportion. In that case, the distribution may not belong to the Benford rule, or some problems may have been encountered. A useful indication of what the critical reference values might be comes from the literature. To decide whether our data set follows Benford's law, we use the following values defined by Nigrini (2012) and represented in Table 2:

2.6.2. Mean of the first significant digit (FSD)

Another metric to measure the coherence of a data set with Benford's law was developed by (Grendar et al., 2007), and it is called the mean of the First Significant Digit (FSD). This measure is defined as follows for the first digit:

$$\bar{d} = \sum_{j=1}^{9} d_j p_j \tag{16}$$

where d_j represents each possible digit (1 through 9 for the first significant digit) and p_j is the probability of observing the digit d_j.

Digits	Range	Conclusion
First Digits	0.000 to 0.006	Close conformity
	0.006 to 0.012	Acceptable conformity
	0.012 to 0.015	Marginally acceptable conformity
	Above 0.015	Nonconformity
Second Digits	0.000 to 0.008	Close conformity
	0.008 to 0.010	Acceptable conformity
	0.010 to 0.012	Marginally acceptable conformity
	Above 0.012	Nonconformity
First-Two Digits	0.0000 to 0.0012	Close conformity
	0.0012 to 0.0018	Acceptable conformity
	0.0018 to 0.0022	Marginally acceptable conformity
	Above 0.0022	Nonconformity

Table 2: Critical Values for Mean Absolute Deviation results

The formula is used to compute the average value of the first significant digits, weighted by their probabilities in a dataset. We know that the theoretical mean of FSD under Benford's law is approximately 3.44 and therefore this is our benchmark. Significant deviations from the Benford mean (3.44) may indicate tampering, human error, or an alternative underlying distribution.

For this reason, we can use the values of the FSD to gain insights into our empirical distribution.

- If $\bar{d} \approx 3.44$, the data are likely to align with Benford's law.
- If $\bar{d} > 3.44$, the statistic indicates an over-representation of larger digits (5 or greater). This can occur due to misinformation or systematic errors.
- If $\bar{d} < 3.44$, the statistic indicates an over-representation of smaller digits (1 or 2). This might result from data truncation or specific data-generation processes.

Using this formula, the FSD mean is 5 when we compare a uniform distribution with the definition of Benford's law. This is because the weight of the frequencies is exactly divided into nine digits with a centre on the digit 5. For values of FSD mean values less than 5, the formula indicates that the distribution is asymmetric, with a higher weight on lower values, like the Benford distribution. For the FSD means between 3 and 4, there is a high correlation between Benford and the empirical proportions of our data set. The correlation would be equal to one if the FSD statistic is equal to 3.44, the precise value of the FSD statistic with Benford's data. Many empirical data sets have FSD means between 3 and 4 and this helps to explain why many unrelated datasets appear to exhibit distributions similar to Benford. When, quite infrequently, the FSD exceeds 5, the distribution becomes more spread out and skews to the left.

The same formula can be generalised for the second or higher digit. For the second significant digit (SSD), the formula is the following.

$$\bar{d}_2 = \sum_{j=0}^{9} d_j p_j. \tag{17}$$

For SSD, the expectation (and thus the result of applying the formula to Benford's distribution) is 4.187. The uniform distribution has an SSD equal to 4.5. For that reason :

- If $\bar{d} \approx 4.187$: the data are likely to align with Benford's law.
- If $\bar{d} > 4.187$: the statistic indicates an over-representation of larger digits (5 or greater). If the values are also larger than 4.5, we have a distribution left-skewed. We know that the Benford distribution is right-skewed.
- If $\bar{d} < 4.187$: Suggests an overrepresentation of smaller digits. We have a distribution more skewed to the right than the Benford one, so our distribution assigns higher proportions to lower digits (such as 0, 1, or 2)

It is important to note that the summation extends from $j = 0$ to $j = 9$, as 0 is a valid second significant digit.

Using the formulas above, written in Grendar et al. (2007), we extended and rewrote them to obtain a formula for the first-two significant digits that we know from Nigrini (2012) are more useful to carry out fraud detection. The formula is:

$$\bar{d}_2 = \sum_{j=10}^{99} d_j p_j. \tag{18}$$

Our value of interest here, the values of the First Two significant digits of Benford's law is 38.59. For a uniform distribution, the mean of the First Two significant digits is equal to 54.5. We can compare, as in the case of FSD and SSD, the value obtained by applying the metric to our data set with these pivot values:

- If $\bar{d} \approx 38.59$: the data are likely to align with Benford's law.
- If $38.59 < \bar{d} < 54.5$: the statistic indicates an over-representation of larger digits with respect to the Benford distribution. However, data are still right-skewed with respect to a uniform, like in Benford's distribution, but less asymmetrical.
- If $\bar{d} > 54.5$: the mean of first-two significant digits indicates a left-skewed distribution, with more weight in the right half of the distribution than in the left half. This situation is opposite to what we expect with a data set that is coherent with Benford's law.
- If $\bar{d} < 38.59$: the data are very right-skewed, more than a Benford distribution. The first digits, with lower numbers, have much more weight than what we expect from a Benford distribution.

Other extensions of this formula can be computed, but we focus our attention only on the first, second, and first-two significant digits.

2.6.3. Sum Squares Deviations

Another useful metric was developed in Kossovsky (2014) and is called sum of Squares Deviations. It is useful to quantify the differences between the actual proportions and the Benford proportion because, unlike the other metrics, it emphasises large deviations (outliers). Like other metrics, this measure does not account for the size of the dataset and therefore cannot serve as a statistical indicator. Without statistical values or hypotheses to test, it cannot determine compliance with the law but can only provide additional insight when considered with other measures discussed in this document. The sum of Squared Deviations is defined as:

$$\text{sum Squares Deviation (SD)} = \sum_{i=1}^{K} (DP_i \times 100 - BP_i \times 100)^2 \tag{19}$$

where:

- DP: Denotes the Dataset Proportion.

- BP: Denotes the Benford's law Proportion.
- K: Represents the number of digits to compare. For example, it is equal to 9 for the first digit, 10 for the second digit, and 90 for the first two digits.

To get a measure of the goodness of the result, we have to compare it with some cutoffs. We take as a reference those written by Kossovsky (2014).

SD	Benford	Acceptably close	Marginally Benford	Non-Benford
First Digit	< 2	$2 - 25$	$25 - 100$	> 100
Second Digit	< 2	$2 - 10$	$10 - 50$	> 50
First-Two Digits	< 2	$2 - 10$	$10 - 50$	> 50

Table 3: Critical Values for sum Squares Deviations (SD)

The formula of this metric is quite similar to the definition of a Chi-Squared statistic (Subsection 2.5.2), except for the fact that the number of rows N is absent and that we do not have to divide by the expected Benford proportion. Furthermore, the lack of N in the definition of the formula leads us to the same results and conclusions regardless of the number of observations in our data.

3 Analysis of electoral data

It is well-established in the literature that electoral data are consistent with Benford's law (Pericchi and Torres, 2004; Mebane Jr, 2010). For that reason, a detailed analysis can be useful in identifying anomalies in candidate vote counts. Such anomalies may be attributed to voter strategies or may serve as indicators of potential problems. These problems could range from simple calculation errors by poll scrutineers to unusual voting patterns that may suggest electoral fraud. The initial connection between vote counts and Benford's law was suggested by Pericchi and Torres (2004), but much work has been done by Mebane Jr (2010). It is possible to distinguish between two types of fraud in electoral data:

1. Planned distortion: Intention of the scrutineer or of someone to distort or change the vote results for some utilities. The intention could be to hide some results, reduce the votes obtained by a party, or increase the votes on a list in a fraudulent way. In that case, it should be possible that those who tamper with data take into consideration Benford's law, making impossible the detection of tampering through this statistical law.

2. Unplanned distortion: When occur some errors in calculation or some problems are found with vote results. This can be caused by lack of capacity of scrutineers, by some problems in aggregating counts, or by other nonfraudolent behaviours.

The use of Benford's law on electoral data is discussed in various studies. Pericchi and Torres (2011) argued that applying Benford's law to the second digits of vote counts provides a reliable standard for detecting election fraud, although there are opposing views. One notable critique comes from Miller in Chapter 9 of his book (Miller, 2015): "The claim by Pericchi and Torres that failure of vote counts' second digits to match the distribution implied by Benford's law provides a sufficient standard for diagnosing election fraud is almost certainly false, at least when precinct or polling station vote counts are examined". In addition, Mebane Jr (2010) argued that digit tests not only help diagnose strategic voting, but are also sensitive to other political factors, such as voter mobilisation or voting strategies. Cantú and Saiegh (2011) also found that Benford's law approximately describes the first digit in some cases of Argentine elections, confirming the fact that the use of this law applied to electoral data is reliable.

3.1. Variable of interest

Some deviations may be due to the shape of electoral districts or certain voter strategies influenced by the electoral system. Coalitions, blocked lists, alliances, electoral thresholds, and split-ticket voting can encourage voters to change their preferences or adopt specific strategies. Therefore, it seems reasonable to focus the analysis on the number of electors as the pivotal variable. However, this variable may also present challenges. The maximum number of possible votes is sometimes unknown, or in countries like the USA, it can be a random variable, especially in states like Maine and Wisconsin, where voter registration is possible even on election day.

For this reason, we will begin with a preliminary analysis of our dataset, followed by an analysis of the first, second, and first-two digits. We will choose a variable from our datasets to base this analysis on. In the analysis of the Italian electoral data, the variable we will use is called `VOTI_LISTA`, which indicates the number of votes that each electoral list or party received in the election. Other variables could also be chosen, such as the number of eligible voters in each constituency, the number of voters who actually voted in each constituency, and many others. In the case of the 2020 US elections, however, we decided to choose as a variable, instead of the vote for the party, the vote for the candidate.

3.2. Italian political elections 2018

We decided to take into consideration the political elections held in Italy in 2018. We used the official data provided by the Italian Ministry of the Interior, downloaded from the website *Eligendo* (`https://elezionistorico.interno.gov.it/eligendo/opendata.php`, last accessed: 17 October 2024). We decided to examine the case of Italian elections as we did not find many analyses with Italian data in the literature. Furthermore, all the Italian data are well stored and ordered in the above-mentioned website *Eligendo* and there is a good level of data detail.

In 2018, in Italy, the electoral system in force was called Rosatellum Bis due to the name of the proposer of the law, Ettore Rosato. In brief, this was a mixed system composed of two parts: the majoritarian and the proportional one. This is because approximately 37% of the seats were allocated using the first-past-the-post voting, and approximately 63% of the seats were assigned with the proportional system. However, the voter could only cast his or her vote on one ballot, and the allocation between the two parts of the electoral system took place later, when the ballot box was closed. Italy has a bicameral parliament, and this election involved all two chambers with slightly different rules. There were two voting papers, one for each chamber, but it was a single one for the first-past-the-post and the proportional systems.

We use the data about only one chamber of the Italian parliament, Camera dei Deputati: transliterated Chamber of Deputies. Due to the already mentioned electoral system, called Rosatellum Bis, we choose the Chamber of Deputies for reasons of simplicity. This branch of parliament has the lowest level of aggregation in terms of the electoral college, has lower electoral thresholds in the distribution of the proportional part of the law, and the age of the electors has to be at least 18 years old, unlike the other chamber, Senato, in which the active electorate has to have at least 25 years old.

The level of aggregation within the data is on a municipal basis, called *Comune*. The aggregated variable inside the dataset is called `COMUNI`. The number of voters at the municipal level is very uneven. There are very small municipalities with a few dozen voters and municipalities, like Rome, with millions of voters. For that reason, inside the system of the Ministery of Interior, the biggest municipalities are divided into statistical zones that correspond to a plurinominal electoral college. To give an example of this variability, we can highlight a very small municipality like Moncenisio, with 24 voters and the largest uninominal constituency in Rome, called Collegio 03 - ZONA CASTEL GIUBILEO, with 163964 voters. We don't have much more detail for that electoral constituency, for example, the polling station level. We exclude from our analysis the votes of Italians living outside of Italy. The decision to not consider these votes comes from the consideration that the vote distribution system for Italians abroad is different and that the constituencies are drawn according to the areas of the world where Italians live. In addition, the coalitions in these foreign constituencies are sometimes different from those made for Italy, and thus vote counting becomes difficult to consider for our purposes. Note that there were 4177725 possible voters from foreign countries in 2018 for the Chamber of Deputies and that they count for 12 seats, 2% of the then total number of seats in the Chamber of Deputies.

Regarding instead the lower electoral threshold in this chamber of the Italian parliament, this condition helps to exclude some types of electoral strategies. Voters are more likely to choose their preferred electoral party instead of choosing another party, worried that their vote may be irrelevant going to the coalition or, in the worst case, lost. In the other chamber, called Senato, a language minority list that has not reached at least 20% votes in a region is excluded from the allocation of the proportional part. In addition, the age of the electors helps us to have a more varied range of voters, numbers of rows, and types of parties. All of these conditions help us not to distort our analysis.

3.2.1. Preliminary statistical analysis

We decide to analyse the variable `VOTI_LISTA` contained in our data set as a pivot variable for Benford's analysis. Using this variable, we have many more numbers to rely on when performing the Benford analysis. Another variable, such as, for example, the number of voters, would reduce the number of rows available for our analysis because we would only have a number of entries equal to the number of constituencies, much less than the number of parties per municipality. Secondarily, the analysis of this variable is much more interesting because incoherence with the law can suggest a fraud in vote counts.

The variable `VOTI_LISTA`, however, considers votes for each party without taking into account the different candidates. This is a fundamental specification for understanding what data are considered. Due to the electoral law (Rosatellum bis), the voters cannot write their preference for the candidates. Candidates are elected in the order of the blocked list, and the electors can choose only a party connected to the candidate or a uninominal candidate. In detail, the elector can express his vote in three different ways, and this affects how our variable `VOTI_LISTA` is counted:

- by choosing the symbol of a list: in this case, the vote for the list or party chosen by the elector is considered;
- by choosing the symbol of a list and the name of the linked candidate in an uninominal constituency: in this case the vote for the list or party chosen by the elector is considered without caring about the name of the candidate;
- by choosing only the name of the candidate in the uninominal constituency: in this case, we do not consider it as a vote valid for `VOTI_LISTA` because any list is indicated. In reality, this vote is extended to the list supporting the candidate or, if there is more than one list, the vote is divided in a proportional way between them.

The proportion is computed according to the votes that each list has obtained in that constituency.

For that reason, in our simplification and our variable `VOTI_LISTA`, some thousand votes are not considered because they are expressed without the list or the preference of the party. The distribution of votes is a fairly complex activity for different reasons. Some candidates have many lists that support them, there are electoral thresholds for lists and parties inside the list, and there may also be some parties that sometimes run in a coalition and sometimes run by themselves. This is the example of Südtiroler Volkspartei, who in some uninominal constituencies is listed as a single party, and in others is inside a coalition. This is why, keeping things simpler, we use only the first two ways of expressing votes, and we do not consider the distribution of votes between lists when only the candidate was voted.

After that necessary specification, all the passages needed to reach a variable that can be analysed and compared to the Benford law are listed below:

1. We use the *group by* method by the name of the constituency and municipality, and we sum the votes of lists or parties with the same name. The name of the constituency is also necessary because in 2018 there were some municipalities with the same name, which were Calliano, Castro, Livo, Peglio, Samone, San Teodoro and Valverde. In other words, we compute a group by the rows `CIRCOSCRIZIONE`, `COMUNE` and `LISTA`, summing list votes contained in `VOTI_LISTA`. In doing so, it is important to note that 9 municipalities have more than a single statistical zone. Those correspond to the biggest cities of Italy and are Torino, Milano, Genova, Bologna, Firenze, Roma, Napoli, Bari and Palermo. For those cities, we are aggregating data from different statistical zones, but we keep them inside the dataset because aggregation makes sense. We lost some detail, but this is also useful to have data spread; that is one of the conditions written in Chapter 2.4;

2. We removed all the lists or parties in each municipality with zero votes. This is a fundamental operation for applying Benford's law. We need a data set in which all rows have one number in the variable `VOTI_LISTA`, and this number must have at least one digit. We remove 11198 rows that correspond to parties with zero votes in certain municipalities. We recall that the first significant digit, in order to be

considered as such, cannot be 0: it has to be an integer number from 1 to 9.

Before analysing data for any Benford's law test, it is useful to compute some exploratory analysis, like histograms, descriptive statistic and quartiles that helps us to gain information to understand if we can suppose that a Benford's distribution is suitable for our dataset. These basic statistical measures are also calculated to see the distribution of the data and to understand whether our variable `VOTI_LISTA` is suitable for a Benford analysis.

	VOTI_LISTA
count	122149.000
mean	258.557
std	2391.581
min	1.000
25%	6.000
50%	22.000
75%	110.000
max	423382.000

Table 4: Descriptive Statistics of VOTI_LISTA

From Table 4, we can get some information to understand if, at least for the rule that we have defined, this data set is a good candidate to be Benford. We can see that we have a lot of rows, in particular 122149, so our data set has a good number of variables on which to perform our analysis. There are no types of lower or upper bounds, except for the number 0 and this is coherent with the Nigrini's condition. We do not have any bounds because our variable `VOTI_LISTA` is a counter. Actually we do not have a well-spread distribution: we can see from the quartiles that most of the weight is in the left part of the distribution and so in the lower numbers.

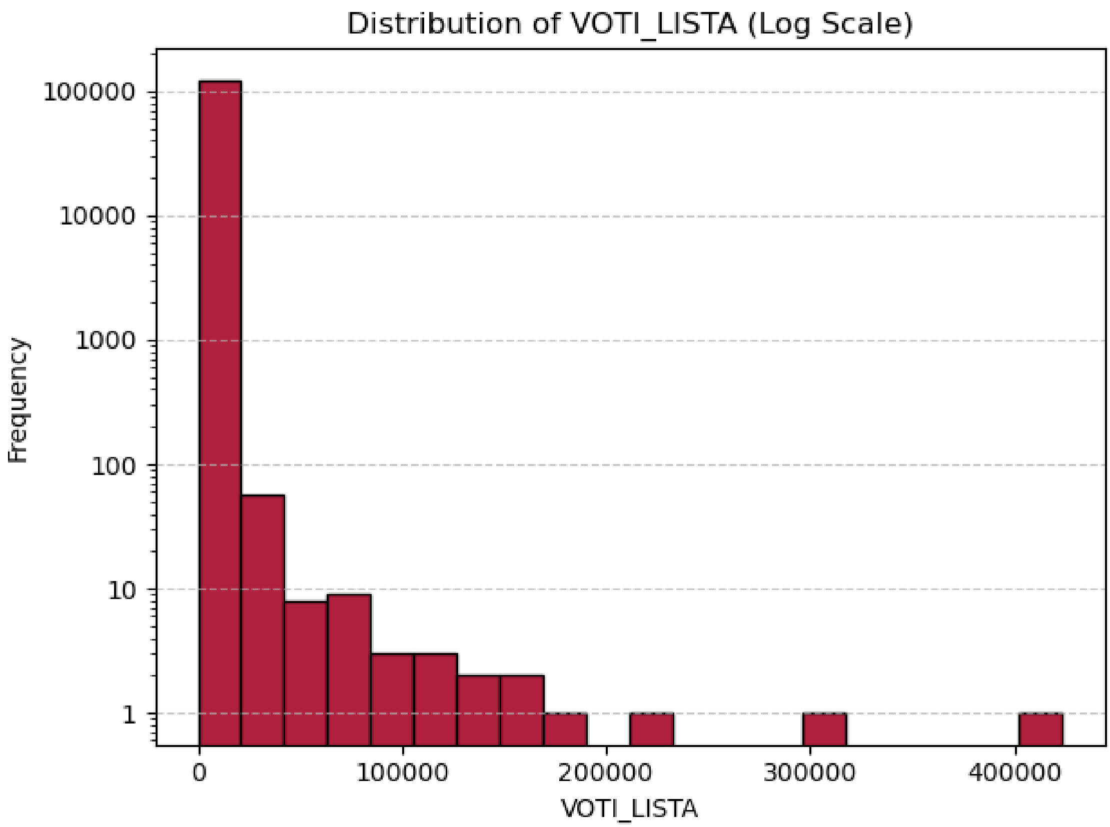

Figure 2: Histogram of the variable `VOTI_LISTA` with logarithmic scale

The histogram 2 suggests a greater concentration on small values of our variable of interest `VOTI_LISTA` and confirms the intuition made from Table 4. This graph uses the logarithmic scale on the y-axis and divides our data into 20 columns. We note that small values have the clear majority in terms of frequency in the graphical representation. In reality, we have many values equal to 1: 8633 rows. This means that there are 8633 cases in which a party collected only one vote in a specific municipality. On the other hand, we have very few municipalities in which the number of votes gathered by an electoral party is high. The maximum value of 423382 votes corresponds to the number of votes totalled by *MOVIMENTO 5 STELLE* in the municipality of Rome, summing all statistical zones of Rome. Note again that aggregating the data at the municipalities level, we lose some information since we have big cities that are divided into several constituencies, like Rome, Milan, Naples, Turin.

One interesting hint that our dataset can be coherent with the Benford's law comes from the conditions that professor Nigrini wrote. As written above, the data set appears to be consistent with those rules. In particular, recalling the conditions defined in Subsection 2.4, conditions 1, 2 and 3 are perfectly respected. As written in condition 1, this dataset indeed represents the size of events, the number of votes taken by a certain list or party in a specific municipality. As the second condition said, we also do not have any lower or upper bound in the dataset, except for the number 1 (because we excluded parties with 0 votes). Even condition number three is respected; our pivot value is not an identification number or a label.

3.2.2. First digit analysis

Now that we have our data set ready for analysis, we can use the variable as a pivot to see if there are some anomalies in the vote counts or some problems in the variable `VOTI_LISTA`. The assumption underlying this and future reasoning on this dataset is that the data from the 2018 Italian Chamber of Deputies elections also follow Benford's law.

In this subsection, we focus our analysis only on the first digit. We recall that the first-digit test compares the actual first digit frequency distribution of data that come from a dataset, in this case the Italian election of 2018, with that developed by the Benford distribution. It is an extremely high-level test and will only identify obvious anomalies (i.e., it will only point you in the right direction). This test can be useful with a data set with a few records, but is usually too high and rough to be of much use (Nigrini, 2012). If a very poor fit is detected in the first-digit test, it means that our data, for some reason, are not following Benford's distribution. This indicates the possible presence of duplications or very important anomalies in the data. Figure 3 shows the frequency distribution of our data as red bars and the Benford's probabilities for the first digit position as black dots.

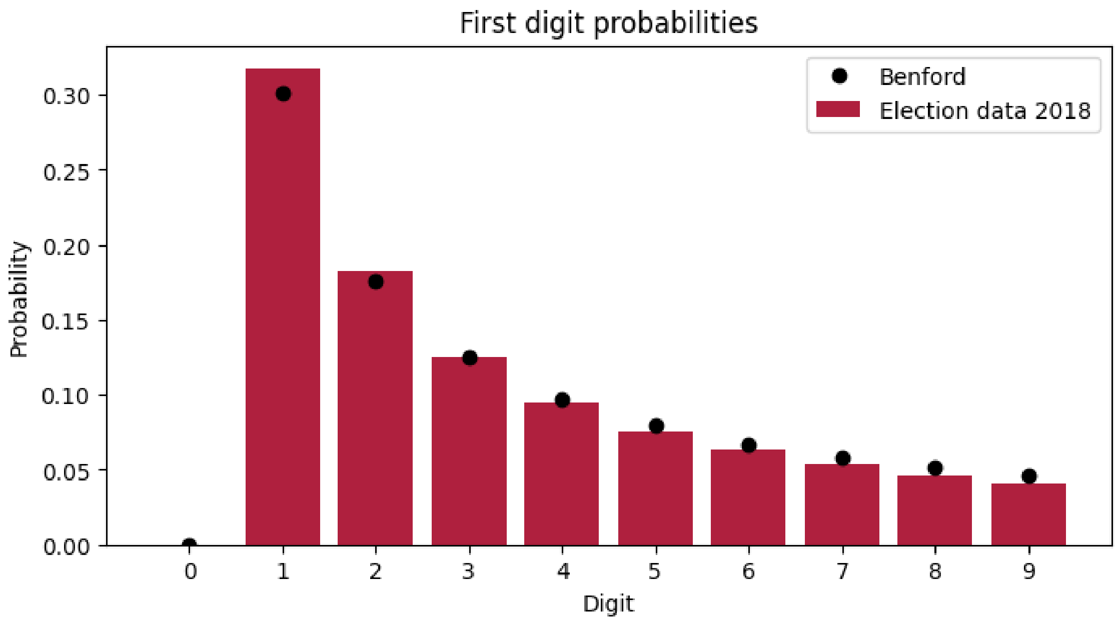

Figure 3: Comparison between election data frequency and Benford's probabilities for the first digit position

We can visually notice that the two distributions are quite similar. The differences seem to be quite small but are not equally distributed in the dataset. For the first lower digits, we have much more weight on data that come from the Chamber of Deputies election of 2018 with respect to Benford's one. This means that the frequency of observing a number of votes that starts with the digit 1 is higher in the data provided by the Italian Ministry of Interior than in the Benford's law dataset that follows a logarithmic law. After this preliminary analysis, we want to verify if at a certain level of significance it is possible to state that data that come from Italian election are not coherent with *First Digit law* from a statistical point of view. To do this, we compute below some statistical tests defined in Subsection 2.5.

Following what we wrote in Subsection 2.5.1 we try to estimate the Z statistic test for the first digit of our dataset. We use the formula (7) with the eventual continuity correction in case the second element of the numerator is smaller than the first term of the numerator. We decided to fix alpha (α) at 0.001, 0.01 and 0.05 that correspond to three different critical values of a normal distribution. In order to solve the multiple tests problem, we apply the Bonferroni's correction defined in Subsection 2.5.5. The number of null hypotheses tested simultaneously on the same dataset is 115. This number corresponds to the 9 Z-tests for the first digit, the 10 Z-tests for the second digit, the 90 Z-tests for the first-two digits, the 3 chi-squared tests and the 3 Kolmogorov-Smirnov tests. The three α levels obtained applying the correction are 0.0000087, 0.0000870 and 0.0004348.

Following the hypothesis test defined in 5, if the p-value is greater than α, we cannot exclude that our dataset follows Benford's law. On the other hand, if the p-value is less than or equal to α, we must reject H_0 and thus statically exclude that our dataset follows the distribution defined in 2.5. We know that we have 121959 entries and that our dataset is graphically similar to Benford's distribution (Figure 3).

First digit	Benford prob.	Election prob.	Difference	Z statistic	p-value
1	0.301	0.317	0.016	12.269	0.000
2	0.176	0.182	0.006	5.612	0.000
3	0.125	0.126	0.001	0.678	0.498
4	0.097	0.095	-0.002	1.808	0.071
5	0.079	0.076	-0.003	4.487	0.000
6	0.067	0.063	-0.004	5.003	0.000
7	0.058	0.054	-0.004	5.878	0.000
8	0.051	0.046	-0.005	8.425	0.000
9	0.046	0.041	-0.005	8.417	0.000

Table 5: Comparison between first digits of Benford's probabilities and first digits of Italian election data 2018 with Z statistic and related p-values

Looking at Table 5, we can see that the higher Z statistic logically corresponds to the highest differences between theoretical and empirical probabilities. Comparing the three corrected α levels with the p-values obtained, we can see that only for digits 3 and 4 we have a relatively high p-value that allows us to not reject the null hypothesis. Instead, for the other digits, the null hypothesis H_0 (5) is rejected and we can say that the distribution of those leading digits of our dataset does not conform to Benford's law. In conclusion, the Z statistic applied to the first digit does not suggest a coherence between our dataset and Benford's law for all the digits. The results need to be further investigated with other tests and metrics to see whether there are indeed problems with this dataset or whether it is only the result of the z-test that suggests that the data from the 2018 Italian Chamber of Deputies elections do not follow Benford's law, as far as the first digit is concerned.

However, it is important to notice that we do not have very high values for the Z statistic, and the result could be influenced by the problems highlighted in Section 2.5.

We then apply the χ^2 test defined in Subsection 2.5.2. We transform the probabilities into absolute frequencies multiplying proportions by the numbers of votes `VOTI_LISTA`. Doing this, we have a number of expected frequencies (the one that comes out from the Benford's distribution) and a number of actual frequencies. Also in this case we consider three levels of alpha $\alpha \in \{0.001, 0.01, 0.05\}$ and we compare it with the p-value. The three alpha levels are appropriately adjusted taking into account the 115 null hypotheses tested on the same dataset. Bonferroni correction defined in Subsection 2.5.5 is applied to control FWER. The levels of α are 0.0000087, 0.0000870 and 0.0004348. The chi-squared statistic (chi^2) equals 344.4165 with 8 degrees of freedom, which correspond to nine levels of our variable (digits from 1 to 9). The corresponding p-value is very low and is basically equal to 0. For that reason, in all three cases, we can reject the null hypothesis H_0 defined in 5. So, the test suggests that the difference between Benford's distribution and data from the Italian election of 2018 is statistically significant for all the α corrected levels we have set. This is an unexpected result that can suggest some problems or anomalies in our data. This result is consistent with the Z-statistic, but is probably due to the excess power problem, highlighted in Section 2.5. This effect was specifically encountered in the chi-squared statistic by Professor Nigrini in Nigrini (2012). In this book, Professor Nigrini wrote that even if from the empirical distribution graph we can see that the actual proportion of our data deviated only slightly from the expected proportion, i.e., Figure 3, the chi-squared test suffers from the excess power problem. "When the number of records becomes large, the calculated chi-squared will almost always be higher than the critical value, leading us to conclude that the data do not conform to Benford's law" Professor Nigrini suggested that this problem started being evident for data sets with more than 5000 records and we, in the electoral data that come from the Chamber of Deputies, have 122149 records.

Another statistical test that we have used to understand if data coming from the Italian election of 2018 of the Chamber of Deputies follows Benford's law is the Kolmogorov-Smirnov (K-S) test. We recall that this test uses the maximum difference between the two cumulative proportions of our data set and Benford's law. This difference is called the *supremum*. The p-value is approximately calculated with a number of simulations equal to 100. The computed p-value that we obtain, with a K-S statistic of 0.0229, is 0.

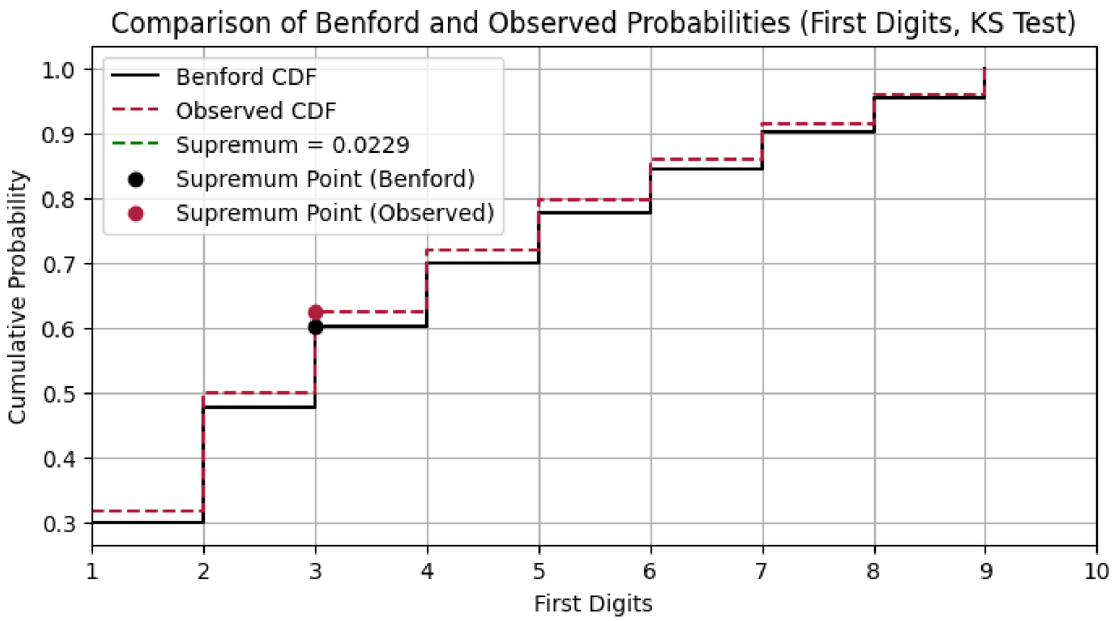

Figure 4: Comparison between cumulative proportion of Benford's law and cumulative proportion of electoral data of 2018 for the first digit, with the highlighting of the maximum difference, the supremum

As we can see in Figure 4, the supremum of 0.0229 corresponds to the digit 3. This value is visually not as high. The graph suggests that Benford's distribution and Italian electoral data behave very similarly and differ only a little as regards the cumulative distribution function. However, comparing the p-value with the three α levels yields an unexpected result. A virtually null level of the p-value leads us to reject the null hypothesis H_0 and to conclude that our data do not follow Benford's law regarding the first digit. Recall that the p-value is an approximation of the true value because both our function and the data are discrete. The K-S test requires the theoretical distribution to be continuous for its assumptions to hold, and for that reason, we introduce an approximation. In conclusion, it is possible to notice that even if there are only smaller deviations, the high number of N increases the denominator and reduces the number of critical values. With higher values of N, it is more difficult to not reject the null hypothesis H_0 and for that reason our data are flagged as suspicious.

The use of a simple regression, suggested in (Saville, 2006), can give us other insights to understand whether our data are consistent with Benford's law or not. Applying the formulation written in Subsection 2.5.4, we obtain an estimate of β_0 and β_1 for the first digit.

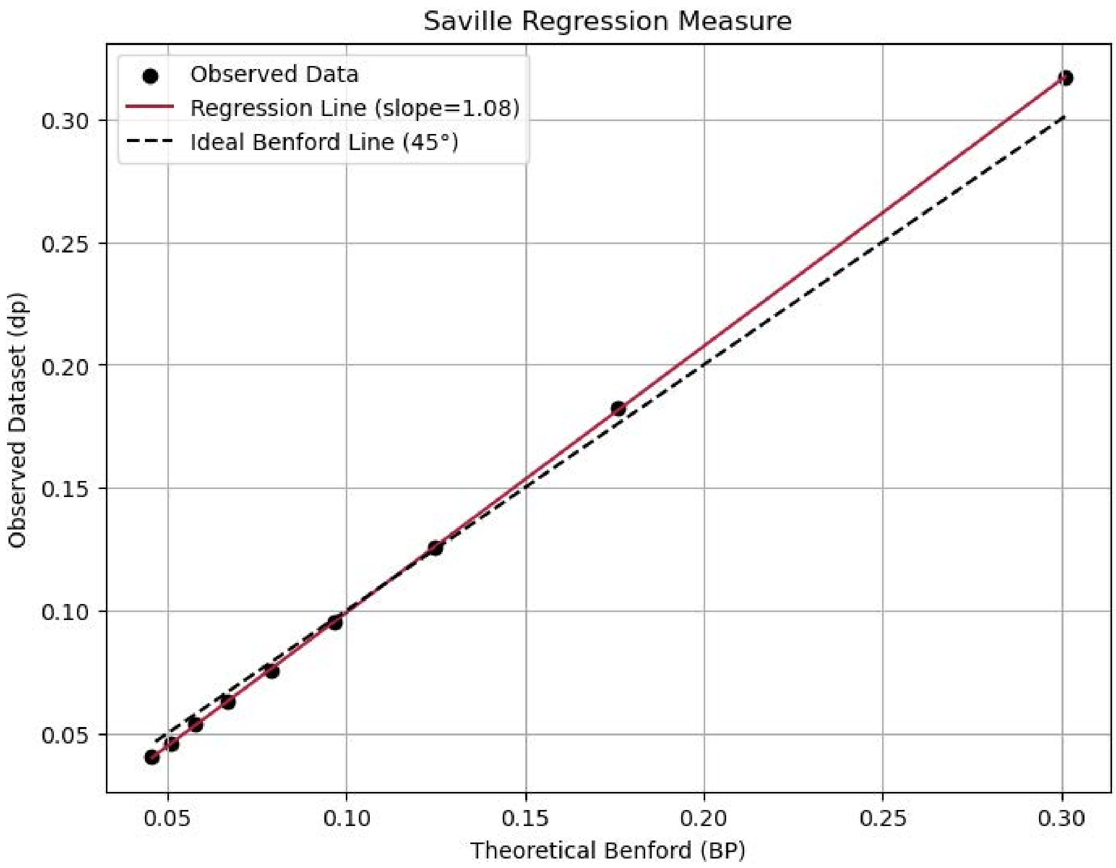

Figure 5: Saville Regression plot for the first digit of the variable `VOTI_LISTA` against Benford's distribution

From a graphical analysis, we can see that the two lines, the bisector and the line representing the regression for our data, are quite similar. The black points are aligned with the $45°$ line, suggesting that Italian electoral data of 2018 closely follow Benford's distribution. The slope value is 1.0849 (β_1). The value of β_0, the intercept is -0.0094. This confirms what is also evident from Figure 3: our data are slightly more skewed than the Benford distribution with higher frequencies on lower digits. The closeness of β_1 to 1 and of β_0 to 0 suggests that there are no significant anomalies or irregularities in our data contrary to the three statistical tests just calculated.

The MAD is particularly useful because, like in the Saville regression, we do not have to consider the number of rows that we have in our dataset. We take into consideration only the proportions and the number of digits that we want to compare. Using the formula and the rules written in Subsection 2.6.1, the result of 0.00508 is obtained for the first digit. Following the indications in Table 2, since our result is between 0 and 0.006, we can conclude that this metric indicates a close conformity of the data with Benford's law. In order to facilitate the comprehension, we recall from Subsection 2.6.1 that the MAD of a Fibonacci's sequence is almost 0, exactly $9.2065e-05$ for 10000 numbers: a number of 10000 rows is chosen due to computation limits of computers. Since it is verified that Fibonacci's sequence follows Benford's rule, refer to Subsection 2.3 and (Miller, 2015, chapter 1) for more details, we have a coherent result. Instead, the MAD for a uniform function, with all probabilities equal to 11%, is 0.05972, very different from our result.

The use of the mean of the FSD summary is based on the research and assumptions of Grendar et al. (2007). In the case of the first digits, the one that goes from 1 to 9, at the very first left position in our 2018 electoral data from the Chamber of Deputies, we apply the formula (16). The result we obtain is 3.3101, very close to the mean of the FSD of Benford's law, calculated by following formula (2) that is 3.4402. Following the decision-making scheme written in Subsection 2.6.2, we can say that our data set is almost aligned with Benford's law, with an over-representation of smaller digits. This confirms what was visually represented in Figure 3 in which it is clearly visible that especially for digit 1, but also for digit 2, we have much more proportion on our data, represented by the red bars, than on Benford's law, represented by some black dots. This analysis seems to confirm that our data have no anomalies or errors and that they follow Benford's law as expected.

The sum of Squared Deviations (SD) is the last metric computed. The formula written in Subsection 2.6.3 is applied for the first digit obtaining the result of 3.9365. Following the table written in (Kossovsky, 2014), we can infer that our dataset is acceptably close to Benford's law. This because the result of the metric gives a value between 2 and 25. Again, the value of a uniform distribution is computed to make a comparison. In this case, the result is 543.42, very different from the result of our dataset that is much more near to a Benford's distribution than to a Uniform distribution. Therefore, the result of this metric is in line with what was calculated before. The other metrics, the visual representation of the two distributions (Figure 3) and the Saville regression (Figure 5) suggest the same thing, our data can belong to Benford's distribution and there are no anomalies or errors in data reporting. However, a further analysis of the second digits and the first two digits is necessary to exclude that this is a blunder and to deepen the results of statistical tests.

3.2.3. Second digit analysis

The use of the second digit as a marker to understand whether an election follows or does not follow the Benford rule is suggested by Pericchi and Torres (2011). This analysis is known to work very well, for example, in detecting problems in election counts, inventory counts, and daily sales numbers (Nigrini, 2012). In order to proceed with the analysis of the second digit, it is necessary to assume that the first digit also conforms to Benford's law, and from the result obtained from the metrics analysed so far, the dataset appears to be consistent.

We computed all the calculations that we did before but applied them to the second digit position. It is important to recall that here we have one more digit, that is, the number 0. This is because if the first digit exists, obviously, a second digit can assume values from 0 to 9. Using the second digit, we are reducing the numerosity of our dataset because we eliminate all the parties or electoral lists that do not have a second digit. These are all the parties that took votes from 1 to 9 in a certain municipality. This results in a reduction of the rows from 122149 to 79678. We are losing 42471 values and so our data set on which we can rely is reduced. As we did for the first digit, Figure 6 shows the comparison between the data and Benford's rule.

Also, in the second-digit case, we can see that the differences between our data and Benford's law are slight. However, we need some statistical tests to understand if we can accept it within the statistical error or not. Like in the previous case, we have more weight on the lowest digits than on the highest one.

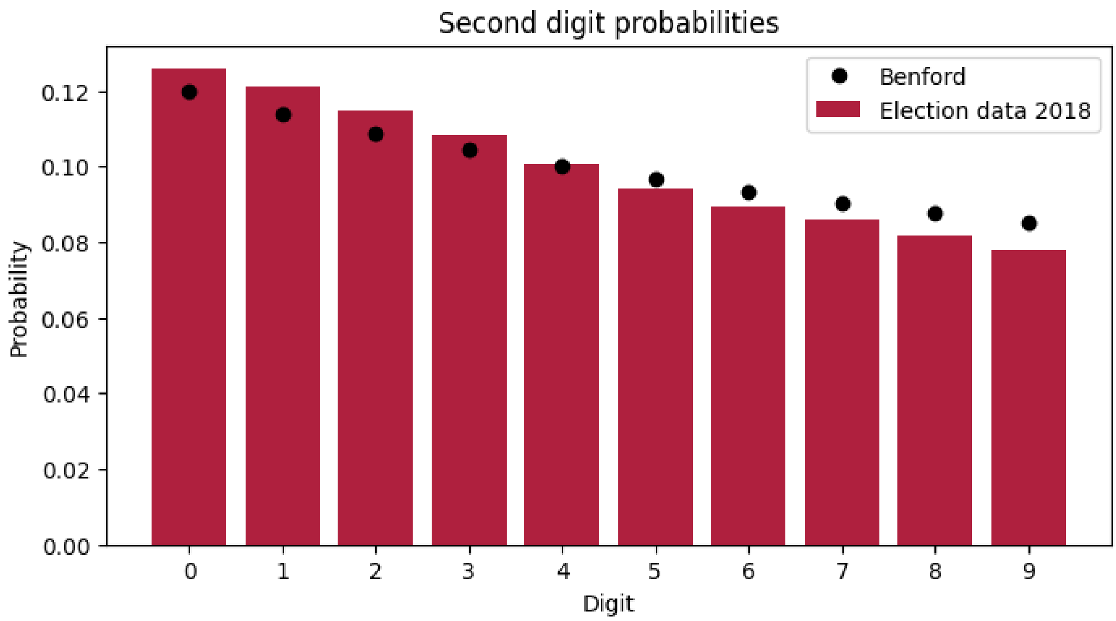

Figure 6: Comparison between election data frequency and Benford's probabilities for the second digit position.

Second digit	Benford prob.	Election data	Difference	Z statistic	p-value
0	0.120	0.126	0.006	5.268	0.000
1	0.114	0.121	0.007	6.390	0.000
2	0.109	0.115	0.006	5.185	0.000
3	0.104	0.108	0.004	3.543	0.000
4	0.100	0.101	0.000	0.238	0.812
5	0.097	0.094	-0.002	2.248	0.025
6	0.093	0.090	-0.004	3.633	0.000
7	0.090	0.086	-0.004	4.134	0.000
8	0.088	0.082	-0.006	5.739	0.000
9	0.085	0.078	-0.007	7.100	0.000

Table 6: Comparison between second digits of Benford's probabilities and second digits of Italian election data 2018 with Z statistic and p-value

We start our statistical analysis with a Z test. Looking at Table 6, we see that the results are similar to the first digit. Also, here, we do not have strange outliers, and the Z

statistic follows proportionally to the size of the difference between the two proportions. The formula also reflects the smaller number of values that we have, due to the elimination of rows with values from 1 to 9, but allows us to consider in the model also digits equal to 0. We compare the p-values obtained for each digit with the three corrected α that are 0.0000087, 0.0000870 and 0.0004348 using Bonferroni's correction defined in Subsection 2.5.5. The comparison between p-values and α values leads us to reject the null hypothesis H_0 for all digits except digits 4 and 5, concluding that this dataset is not coherent with Benford's law. Regarding digits 4 and 5 we have a low Z statistic and a p-value that suggest that we cannot reject H_0, the hypothesis that our dataset follows Benford's distribution. This result is stronger for the digit 4 than for the digit 5. Therefore, also in the case of the second digits, the Z statistic does not confirm that our data follow Benford's law for all digits. However, we recall that we do not have strange or very high values and that the test can be affected by some problem written in Subsection 2.5.

In addition to this test, we calculate the chi-squared test. We recall that to compute this statistic, we transformed the values of the actual and expected proportions into the actual and expected frequency. Here, differently from the case of the first digit, we have one more digit: digit 0. For that reason, we have one more degree of freedom (9). Keeping the same alpha levels corrected with Bonferroni's correction $(0.0000087, 0.0000870, 0.0004348)$, we compare these values of α with the p-value obtained from the test. The result of our chi-squared statistic is 204.5951 with a p-value equal to 0.000000. For that reason, we have to reject the null hypothesis, even if the graphical representation suggests a similar behaviour between our data and Benford's law. However, although this test suggests that there may be problems in our data, it is necessary to remember that we may have run into problems attributable to this type of test, already explained above in Subsection 2.5 and already verified in the case of the first digit.

Another statistical test we apply is the Kolmogorov-Smirnov (K-S) test for the second digit. Again, like in other cases where we use the second or higher digits, we have to take into account one more digit with respect to the first digit, the digit 0. The computational solving returns as a solution the number of $.0231$. This is the supremum, and it corresponds to the digit 4.

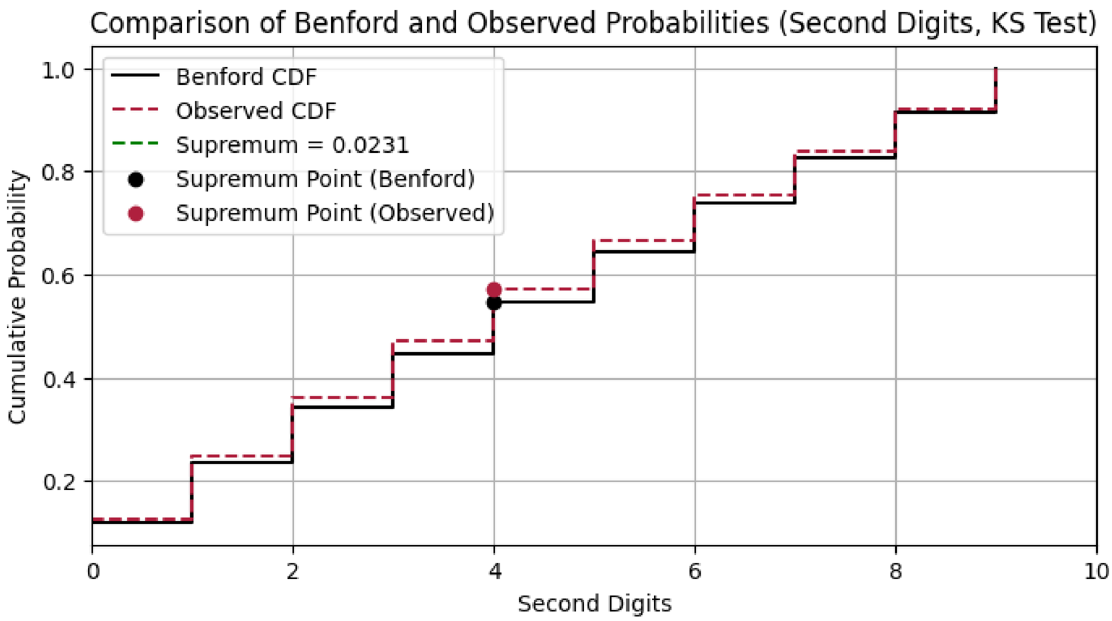

Figure 7: Comparison between cumulative proportion of Benford's law and cumulative proportion of electoral data of 2018 for the second digit, with the highlighting of the maximum difference, the supremum

In Figure 7, we can see a visual representation of the result, with the two cumulative proportions of Benford's law, in black, and the Italian election data, in red. We can notice that the two cumulative distribution functions seem quite aligned.

The result of the Kolmogorov-Smirnov statistical test, 0.0231, corresponds to a p-value of 0.000000. This p-value is an approximation performed on 100 numbers, to overcome the continuity problems of this test already defined in Subsection 2.5.3.

Like in the first-digit case, our p-value compared with all three levels of α corrected suggests to us to reject the null hypothesis H_0, excluding that, with regard to this test, our dataset is coherent with Benford's law. However, even in this case, we have a large number of entries, which probably fall under the higher power problem defined in Subsection 2.5. Therefore, the fact that this statistical test indicates that our data set may have been tampered with should not worry us too much.

The use of a regression suggested by Saville can be easily extended to the second digit (Saville, 2006). We apply a simple regression to our data, using the formulation written in Subsection 2.5.4 to obtain an estimated β_0 and β_1.

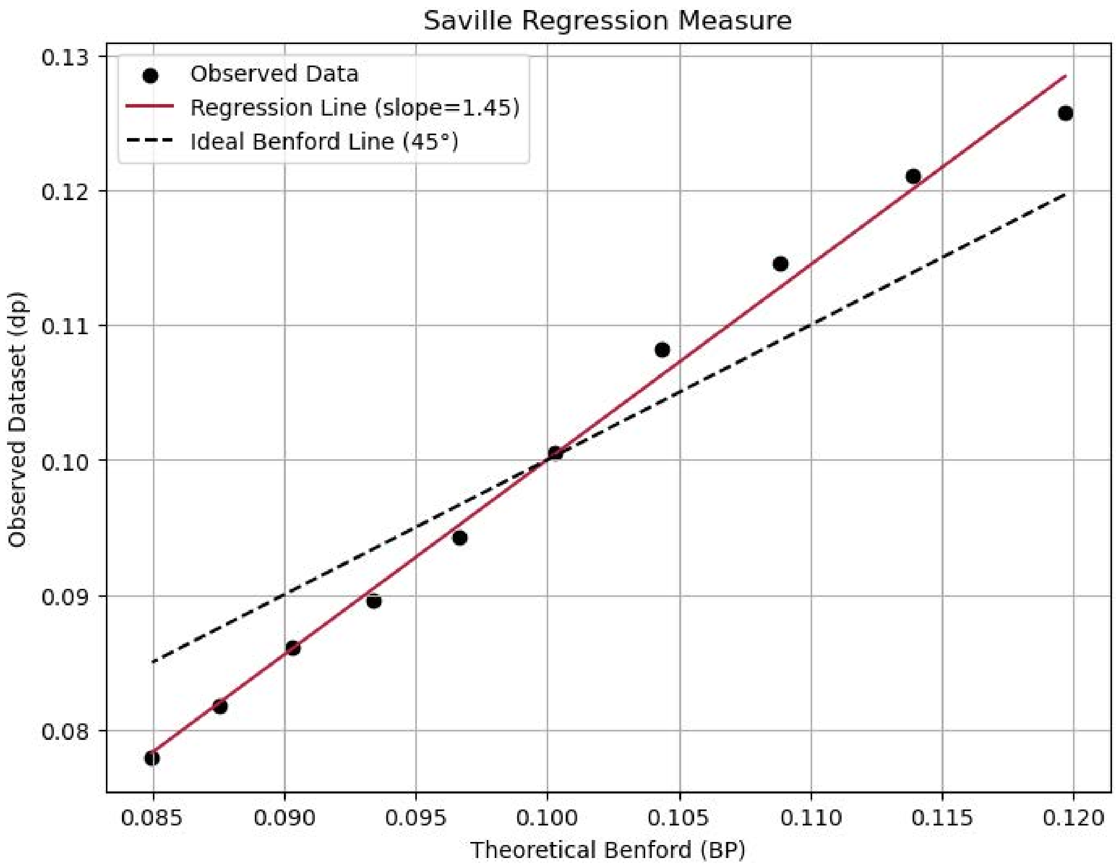

Figure 8: Saville Regression plot for the second digit of the variable `VOTI_LISTA` against Benford's distribution

Graphically, we can see that the black dots, located in combination of the theoretical and actual probabilities, are not exactly aligned with the bisector. Our data assign a higher probability to the lower digits (represented in the upper right part of the graph) and a lower probability to the higher digits. This is confirmed by the values of the results of β_0 and β_1 of the regression. The intercept β_0 is -0.0446, not far from our objective of 0. The slope β_1 instead is 1.4456 different from our objective of 1. In conclusion, this regression reveals that our dataset amplifies the differences between frequencies of lower and higher digits, resulting in a greater skewness. This conclusion is consistent with Figure 6 and Table 6. This test therefore leads us to doubt whether the second digit follows Benford's law. We must therefore apply some metrics like MAD, SSD and SD.

In the application of the MAD for the second digit, we also have to consider the digit 0. For that reason, in the formula (15) K is equal to 10. The result of this formula applied to our data is 0.004618, which suggests that our data deviate very little from Benford's law.

Again, like in the case of the first digit, this is an indication of close conformity, taking by reference what is written in Table 2. Also here, to facilitate comparison and understanding of the numbers, we report the MAD result for a Fibonacci sequence of 10000 numbers, that is, 0.00015. The result of the MAD for a uniform function, with all the probabilities equal to 10% is instead equal to 0.00941. This reflects the fact that for the second digit, the proportions for Benford's law are less skewed and so are more similar to a uniform function than the first digits.

Another metric is calculated, using the formula defined in Equation (17) suggested by (Grendar et al., 2007). This is called the mean of the Second Significant Digit (SSD). Our benchmark is again the value calculated following the formula of Benford's distribution extended for the second digit, i.e., Equation (3), that is 4.1874. The value obtained by analysing the SSD of our 79678 rows is instead 4.0456. The two values are very close, so our data are very coherent with Benford's distribution. Following the decision-making scheme defined in Subsection 2.6.2 it is possible to state that, also in this case (like for the FSD), our smaller digits are represented more than the lower digits of the Benford law. This is consistent with Figure 6, in which we can clearly see this behaviour.

Lastly, we compute the sum of Squared Deviations (SD). For the application of this metric, we use the formula written in Subsection 2.6.3, applied to the second digit. The result is 2.5594, a value that, following the table written in (Kossovsky, 2014), allows us to say that our dataset is acceptable close to Benford's law. The result is consistent with the other two metrics, although the Saville regression (Figure 8) shows that our data set assigns a higher frequency to the lower figures. The result of a uniform distribution that can be used for a comparison is in this case 12.0447, very different from our value.

In conclusion, we can state that, using these metrics, our suspicions that the data from the Italian 2018 Chamber of Deputies elections may have been tampered with are completely disproved. The results obtained from the metrics are diametrically opposed to the results obtained with statistical tests.

3.2.4. First-two digits analysis

The use of the first two digits together was suggested by Professor Nigrini (Nigrini, 2012), who formally defined it in his book and used it for forensic accounting, auditing, and fraud detection. He said that the first two-digit test has to be preferred to the other Benford's law tests "because it captures more information than the first and second digit tests combined". We take the definition of Professor Nigrini in Nigrini (2012):

$$\text{First two digits}(x) = |\text{Significand}(a)| \tag{20}$$

$$\text{where} \quad x = a \times 10^n \quad (10 \leq a < 100,\ n \text{ integer})$$

The significand is the first left part of a number, which is always an integer value. It is important to note that this definition restricts our analysis to a smaller set of data because we are analysing the 90 digits from 10 to 99. All rows with a number from 0 to 9 are excluded, like in the second digit analysis.

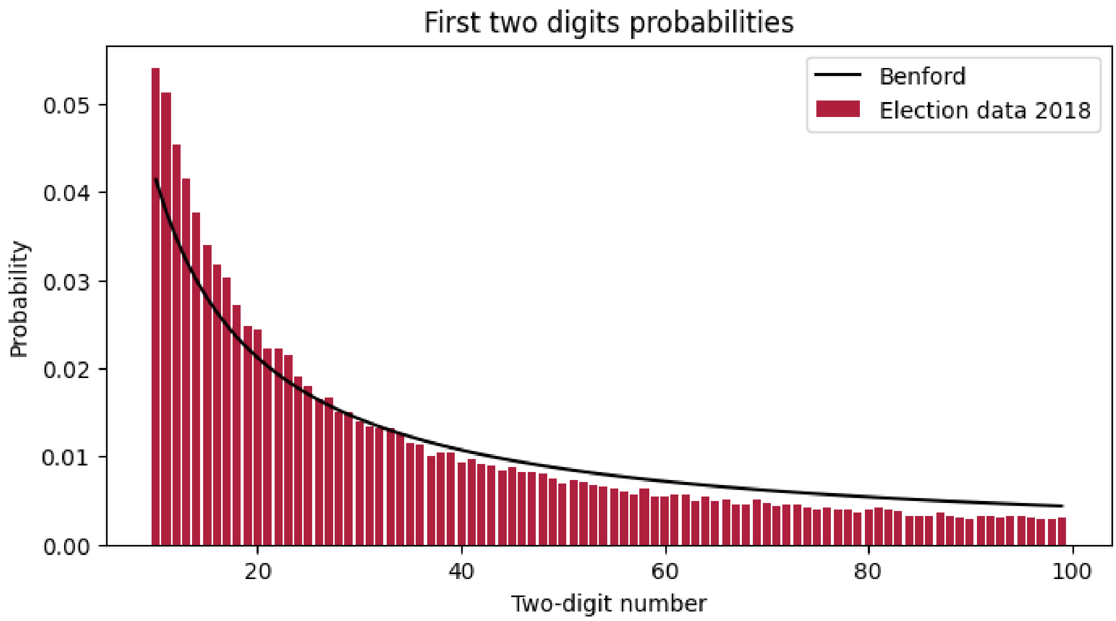

Figure 9: Comparison between election data frequency and Benford's probabilities for the first-two digits position

Looking at Figure 9 we can note the behaviour of our distribution (red bars) compared with the Benford's distribution line (black line) and we can see that they have quite similar behaviour. Also, in this case, our data have a more prominent skewness than the Benford one. There is more weight on the lower digit of Italian election data for 2018 than on the Benford's law. This asymmetry is mirrored in the larger digits. When the digits get higher, the difference becomes proportionally higher.

The Z statistic is then computed on all the 90 digits present in the first-two digit test. There are 79678 rows with at least two digits and are therefore eligible for the First-Two digits test.

Firs-two digits	Benford prob.	Data prob.	Difference	Z statistic	p-value
11	0.038	0.051	0.013	19.816	0.000
10	0.041	0.054	0.013	17.792	0.000
12	0.035	0.045	0.011	16.260	0.000
13	0.032	0.041	0.009	14.876	0.000
14	0.030	0.038	0.008	12.781	0.000
33	0.013	0.013	0.000	0.516	0.606
32	0.013	0.013	-0.000	0.380	0.704
34	0.013	0.013	0.000	0.299	0.765
28	0.015	0.015	-0.000	0.283	0.777
26	0.016	0.016	0.000	0.183	0.855

Table 7: Comparison between first-two digits of Benford's probabilities and first-two digits of Italian election data 2018 with Z statistic and p-value. In the table are represented the five highest differences and the five lowest differences with the correspondent digit and Z statistic.

We can notice from the data, resumed in Table 9, that the differences between our data and Benford's distribution are higher at the beginning, in the very first digit. In the table, we can see that the digits from 10 to 14 are those with the highest Z statistic. We also note that the differences become very low after a few digits and become a little bit higher in the last part of the distribution. The five digits with a lower Z statistic are reported in Table 9 and are digits 26, 28, 34, 32, and 33. Since several null hypotheses H_0 are tested on the same data, we apply the Bonferroni's correction. Comparing the p-values against the three levels of α, we know that we have 35 values out of 90 for which we cannot reject the null H_0 at the 0.0000087 corrected significance level, 28 at the 0.0000870 corrected significance level and 24 at the corrected significance level of 0.0004348. Our results for the first-two digits are not well defined. We have some digits for which comparison with the Z critical value suggests rejecting the null hypothesis H_0 and other values that suggest that we should not reject H_0. The results do not suggest a complete adherence of the data to Benford's law; there may be some errors or tampering in the data.

We now compute the χ^2 test statistic. The increase in the number of possible values that our variable can assume leads to an increase in the critical value of our test. In particular, we know that our digit can assume 90 values, from 10 to 99. The degrees of freedom now are $k - 1$ and so 89. The result of our Chi-squared statistic is 3600.4739 with a corrisponding p-value of 0.000000. Keeping fixed the same levels of α that we used before and using the Bonferroni's correction we have to reject the null hypothesis H_0. From this test we have to conclude that our data do not come from Benford's distribution. However, like in the case of the first and second digit, we have to remark that this test suffers from the excess power problem because we have here 79576 entries, values much higher than 5000, the maximum value suggested by Mark Nigrini in Nigrini (2012). For that reason and because of what we have written above in Subsection 2.5, we can not really rely on this text.

Another test that we use is the Kolmogorov-Smirnov test. The application of this test to the first-two digits was suggested by Professor Nigrini in Nigrini (2012). We applied this test to all the digits from 10 to 99, obtaining a supremum value of 0.0917 in correspondence with the first-two digits equal to 29. This means that the highest difference between the two cumulative density functions is when the first-two digits are 29.

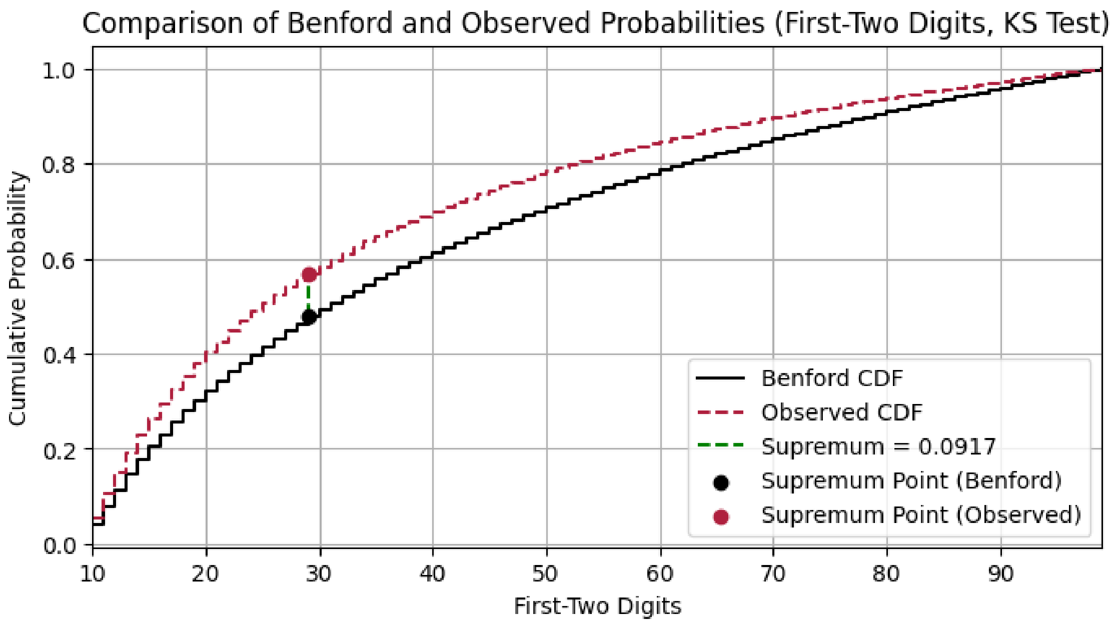

Figure 10: Comparison between cumulative proportion of Benford's law and cumulative proportion of electoral data of 2018 for the first-two digits, with the highlighting of the maximum difference, the supremum. This value is represented with a green line that connects the two points in the two cumulative distribution functions.

Figure 10 supports our numerical analysis and also helps us visualise the behaviour of the two cumulative distribution functions. Now, to see if the two curves are statistically similar, we have to compare the p-value of our K-S statistic against the three corrected alpha levels, which are 0.0000087, 0.0000870 and 0.0004348. Recalling again that with this test, we have to face some problems related to the definition of the K-S statistic defined in Subsection 2.5.3, we obtain a p-value equal to 0.000000. Clearly, again we have to reject the null hypothesis H_0. The K-S test for the first-two digits therefore suggests that our data do not belong to Benford's distribution and this result is an indication of some problems in our data. Again, since the number of N is high, we have to keep in mind the excess power problem defined in Subsection 2.5.

The use of a regression, suggested in (Saville, 2006), can also be applied to the first-two digits. We extended formula written in Subsection 2.5.4 to the 90 first-two digits to obtain a value of β_0 and β_1.

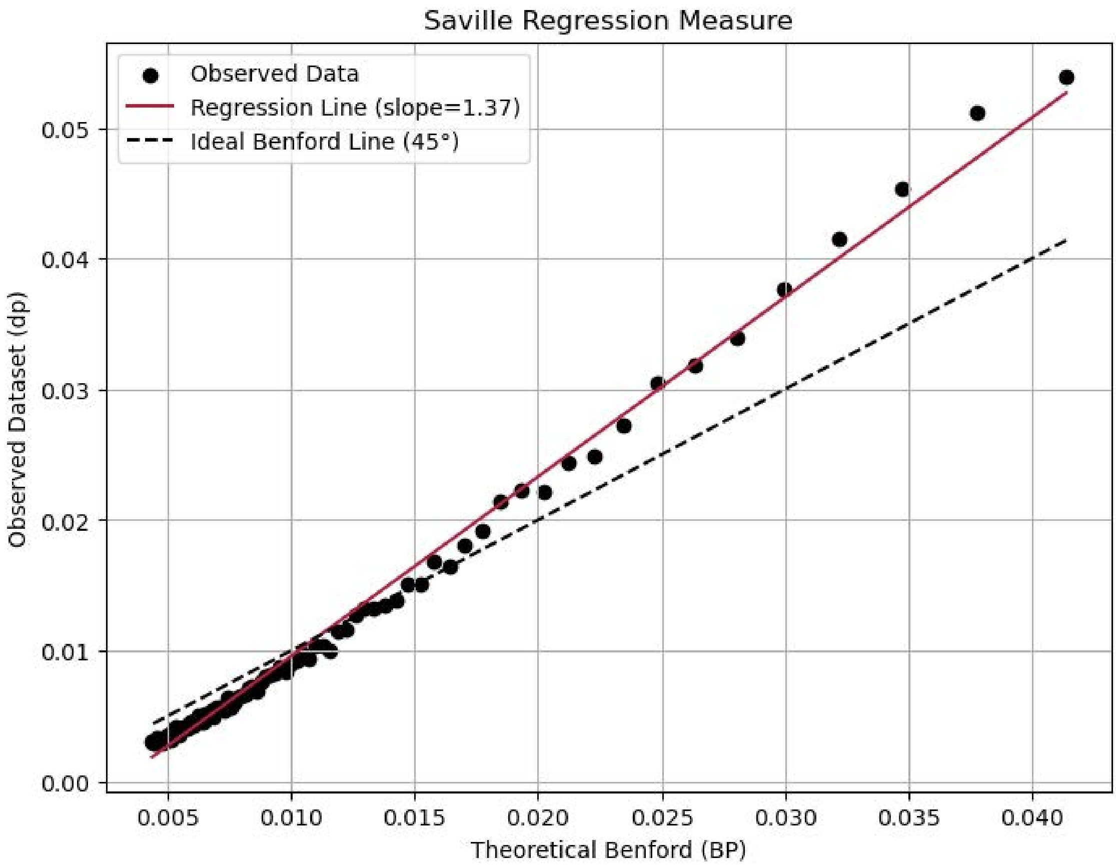

Figure 11: Saville Regression plot for the first-two digit of the variable `VOTI_LISTA` against Benford's distribution

Again, like in previous cases, from Figure 11 it seems that our distribution is more skewed than the Benford one. At the bottom left of the graph, there are many black points that correspond to the higher digits that have lower probabilities to appear. Our data set and Benford's distribution seem quite coherent, but the regression line is not very aligned with the bisector. In fact, the slope value β_1 is 1.3728, different from 1. The intercept value β_0 is -0.0041, quite aligned with 0, the value of the Benford line. For that reason, the Seville regression applied to the first two digits suggests that our data have a stronger skewness than Benford's law predicts. The divergence is concentrated on the lower digits (digits with the highest relative frequencies), as already highlighted in Figure 9 and Table 7.

The result of the MAD metric for the first-two digits, from 10 to 99, is 0.00205. This value, between 0.0018 and 0.0022 0.0022 suggests "Marginally acceptable conformity", following Table 2. So, this metric suggests that, as hypothesised from the graph and preliminary data observation, our data are coherent with Benford's law. Again, Mean Absolute Deviations of Fibonacci's sequence and Uniform distribution are reported. The first is almost equal to 0 ($8.2106e-05$), and the second is equal to 0.006, suggesting a clear non-conformity. This metric is widely used in Nigrini (2012) for forensic accounting, auditing, and fraud detection and is more recommendable than the same metric applied to the first or second digit. In summary, we obtained that the data from the Italian election of the Chamber of Deputies are coherent with the Benford's law, using the Mean Absolute Deviation on the first-two digit. Now, to assess the strength of this result, we compute other metrics.

The mean of the First-Two Significant Digit (FTSD) is an extension of what is written in Grendar et al. (2007). Applying the formula defined in Equation (18), we obtain a value of 33.5873 for our dataset, different from 38.5898 of Benford's law. This difference is also visible in Figure 9, where we can clearly see that our distribution assigns much more weight to the digits from 10 to 20 than to the digits from 69 to 99. This intuition is also confirmed by the decision-making scheme described in Subsection 2.6.2. Our data are very right-skewed, more than a Benford's distribution. This result seems to confirm that our data are skewed to the right, like Benford's one, but suggests a problem in the weight of skewness. However, it is not possible to rely on this metric because the FTSD is not supported by the literature. Only FSD and SSD are written and defined in Grendar et al. (2007).

The use of the sum of square deviations (SD) is supported by the literature (Kossovsky, 2014) and is therefore more reliable. Using the formula written in Subsection 2.6.3, we obtain the result of 8.8314. This value is again, as in the first and second digits, indicating an acceptably close conformity to Benford's law and is very different from the value of this metric applied to the uniform distribution that is 58.5872. However, we can notice that this value is higher than 0 because, as already written above, our data are here very right-skewed, more than the Benford's distribution. Nevertheless, in this case as in the previous metrics, the metrics suggest a belonging of our dataset in Benford's law.

3.3. 2020 US presidential election

To understand the goodness of the statistical tests and metrics, we use another dataset with a smaller number of rows. This data comes from the 2020 US presidential election, won by

Joseph R. Biden on 3 November 2020. We decided to take the 2020 data because that are the most recent data available on the Federal Election Commission website of the United States of America. Data are downloaded from the website *FEC* (`https://www.fec.gov/introduction-campaign-finance/election-results-and-voting-information/federal-elections-2020/`, last accessed: 2 January 2025). The election of the President and Vice President of the United States is an indirect election in which citizens of the United States who were registered to vote can vote for electors, who then formally elected the President and Vice President through the Electoral College.

Each state has a number of electors equal to the sum of its US Senators (2) and Representatives in the House. The total number of electors is 538, so a candidate needed at least 270 electors to win the presidency. The electoral system in effect in most states in 2020 was a winner-takes-all system, in which the candidate that receives the most votes won all of its electoral votes and electors. In Maine and Nebraska, instead, the electoral system was different, following a district-based allocation of the electors.

3.3.1. Preliminary statistical analysis

We decided to analyse only the votes for the candidate president, contained in sheet 9 called "2020 Pres General Results" of the xlsx file. Our variable of interest is called `GENERAL RESULTS` and contains the votes for each candidate grouped by state. We remove all the rows without a name of the candidate president in order to exclude from our datasets the total states votes: rows that sum the votes for all the candidates in a state. We also drop all rows with an empty or zero value in the variable `GENERAL RESULTS`. Furthermore, for the state of New York, we have grouped by the name of the presidential candidate. This is because in these states there are two parties that have cross-endorsed other parties. In particular, the Conservative Party of New York State cross-endorsed the Republican ticket, nominating Donald Trump for president and Mike Pence for vice president, and the Working families Party cross-endorsed the Democratic ticket, nominating Joseph R. Biden for president and Kamala Harris for vice president. We therefore joined these parties in New York State because they had the same presidential and vice-presidential candidates. Taking into account then only the votes of the presidential candidates, we aggregated different running mates. In the United States of America, for a candidate, it is possible to designate different vice-presidential candidates depending on the state. This is due to different agreements and different laws that govern the access of the ballot. Taking into consideration only the name and surname of the candidate presidents, this aspect is not taken into account.

We ended up with a dataset of 553 rows for 51 states (District of Columbia is also considered) with 96 candidates. Note that not all the candidates at the presidency of the United States for America are candidated in all the 51 states; only Donald J. Trump, Joseph R. Biden and Jo Jorgensen are.

	GENERAL RESULTS
count	553.00
mean	286145.38
std	923113.38
min	1.00
25%	25.00
50%	1035.00
75%	22656.00
max	11110639.00

Table 8: Descriptive Statistics of GENERAL RESULTS

The smallest value we have is 1, which is present in 32 lines and corresponds to candidates who only scored one vote in a certain state. The maximum value of 11110639 corresponds to the votes totalled by Joseph R. Biden in California. From Table 8 it is also possible to infer that our data respect the conditions 1, 2 and 3 defined by Professor Nigrini and written in Subsection 2.4. These conditions are respected because, like in the case of the Italian elections, data that we want to analyse are representing in some sense a size of an event. Our pivot variable `GENERAL RESULTS` is a counter, with no minimum or maximum values except for the number 0, which allows us to apply Benford's law. However, in contrast to the Italian elections, we have far fewer rows in the dataset: 553. This allows us to understand whether or not our dataset belongs to Benford's law, limiting the disturbance of the phenomenon called the excess power problem. We know that this problem, already defined in Section 2.5, occurs in datasets with more than 1000 rows. To summarise, we want to check whether there may be errors in this dataset or whether data tampering may have occurred, assuming that the US electoral data, like the other electoral data, follow the

Benford distribution for first, second and first-two digits.

3.3.2. *First digit analysis*

Our analysis begins with the observation of the first digit. It is important to recall that the first digit can be a number from 1 to 9 and therefore the number 0 must be excluded from our dataset.

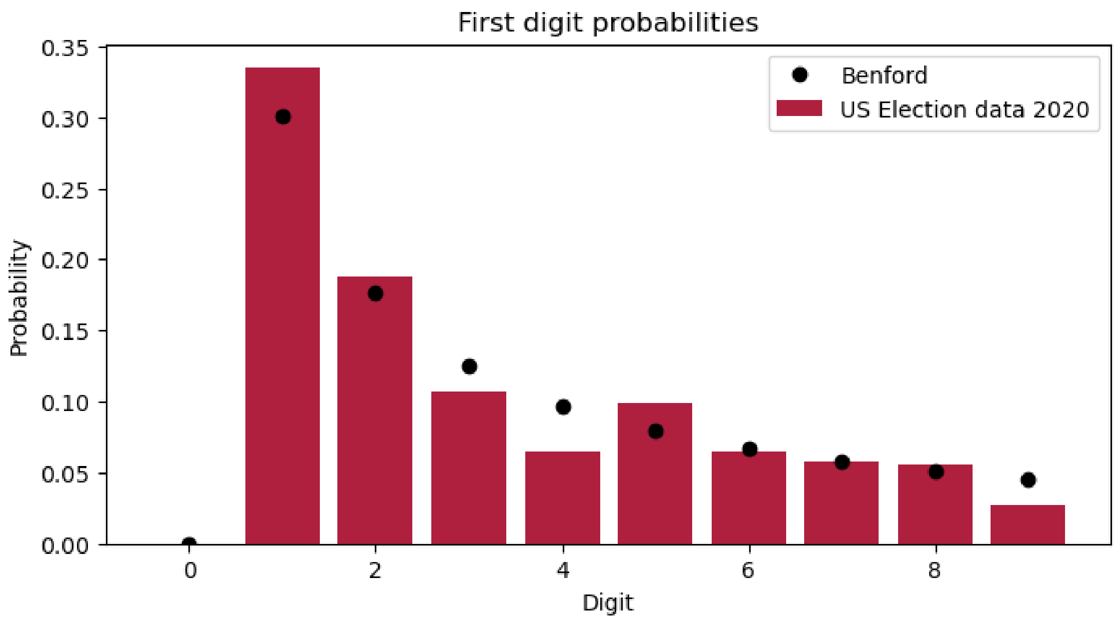

Figure 12: Comparison between US election data frequency and Benford's probabilities for the first digit position

From Figure 12 it is possible to compare the behaviour of the probabilities of the US dataset with the theoretical probabilities of Benford's law. We can infer that our dataset has a skewed distribution as regards the first digits. The probabilities are decreasing from the smallest digits (1 and 2) to the highest one (8 and 9) but there are some exceptions: we can see from the Figure 12 that the decreasing function is not respected for all digits. In addition, for some digits, the difference between the probability of our data and the theoretical one is visually evident and clear. We therefore need the support of some numbers that allow us to understand quantitatively how far our data deviates from the assumed probabilities. We apply the statistical tests defined above in Section 2.6.

Digit	Benford	U.S. data	Difference	Z P-values	χ^2 p-value	K-S p-value
1	0.301	0.335	0.034	0.095	0.029	0.203
2	0.176	0.188	0.012	0.494	0.029	0.203
3	0.125	0.107	-0.018	0.217	0.029	0.203
4	0.097	0.065	-0.032	0.014	0.029	0.203
5	0.079	0.099	0.020	0.092	0.029	0.203
6	0.067	0.065	-0.002	0.929	0.029	0.203
7	0.058	0.058	0.000	0.990	0.029	0.203
8	0.051	0.056	0.005	0.669	0.029	0.203
9	0.046	0.027	-0.019	0.046	0.029	0.203

Table 9: Comparison between first digit of Benford's probabilities and first digit of United States election data of 2020 with p-values of Z test, chi-squared test and Kolmogorov-Smirnov test.

From Table 9 it is possible to confirm what is visually evident in Figure 12. The probabilities of a number to appear as a first digit in our dataset do not really follow Benford's law. There are digits where the theoretical and empirical probabilities are closer to 0. For the number 7, for example, the difference is negligible, and for the number 6 we have only a -0.2% smaller probability in our dataset than Benford's law. There are digits instead where the differences are greater. The number 1, for example, has a difference of 0.034. This corresponds to a higher probability of 3.4% of our dataset with respect to Benford's law to have a number that begins with a number 1. Another thing we can see from Table 9 is that the probabilities of the numbers to appear as the first digit are not strictly decreasing. The number 4 has a lower probability of appearing than the number 5 and is almost the same as the number 6.

After that preliminary analysis, our attention is focused on statistical tests. We fix the same levels of alpha as before and apply Bonferroni's correction to overcome the problem of multiple tests. Our levels of corrected alphas (α) are 0.0000087, 0.0000870, and 0.0004348. To begin, we perform the Z test for each digit, defined in Subsection 2.5.1. The values obtained and reported in the fifth column of Table 9 are quite convincing. We cannot reject the null hypothesis H_0 defined in 5 for all digits for all three levels of corrected α. We have the lowest p-values when the first digit is 4 or 9. In summary, the Z test suggests that we cannot reject the null hypothesis H_0 and therefore exclude the compliance between our data and Benford's law for almost all digits. For that reason, as far as Z-test regards, we do not have evidence of errors or anomalies in our data. Apparently, we are not experiencing the excess power problem that affects this kind of test. More tests and appropriate corrections should be evaluated.

The second test we apply is the chi-squared test (χ^2), defined in Subsection 2.5.2. This test, unlike the previous one, is not applied digit by digit but instead evaluates the joint probability of all digits in the first position, which can take values from 1 to 9. For that reason, the result reported in the sixth column of Table 9 is equal for the nine digits. We maintain the same corrected alpha (α) levels of the previous statistic, since we are computing the 115 test simultaneously. The result obtained by applying the formula (8) is 0.029. For that reason, we cannot reject the null hypothesis H_0, defined in 5: we cannot exclude that data from the election of the presidential United States of 2020 belong to the Benford distribution. No anomalies or errors are detected, but again, as in the case of the Z statistic, we do not have a very large dataset. We have only 553 rows and so we do not incur in excess power problem, verified for Chi-Squared tests in datasets with more than 5000 rows (Nigrini, 2012).

The last test we apply is the Kolmogorov-Smirnov test, and the value is reported in the last column of Table 9. This test, like the chi-squared, returns only one value that corresponds to the p-value associated with the supremum. The highest difference between the two cumulative distribution functions occurs at the value of the second digit equal to 2. Here, the supremum is 0.0455 and the p-value is 0.02027. Using the three corrected levels of α defined above, we cannot reject the null hypothesis H_0. Even if, as reported in Subsection 2.5.3, this test is not appropriate, we have a useful indication for our analysis. The indication is that the comparison between the p-value and the level of alpha suggests that we cannot exclude the fact that our data belong to Benford's distribution and therefore we do not have anomalies or evident errors that are detected with this test. Notice that the excess power problem here is not encountered because of the low number of rows that we have.

We also apply the Saville regression, defined above in Subsection 2.5.4 to have another hint and to understand if our data are consistent with Benford's distribution or not.

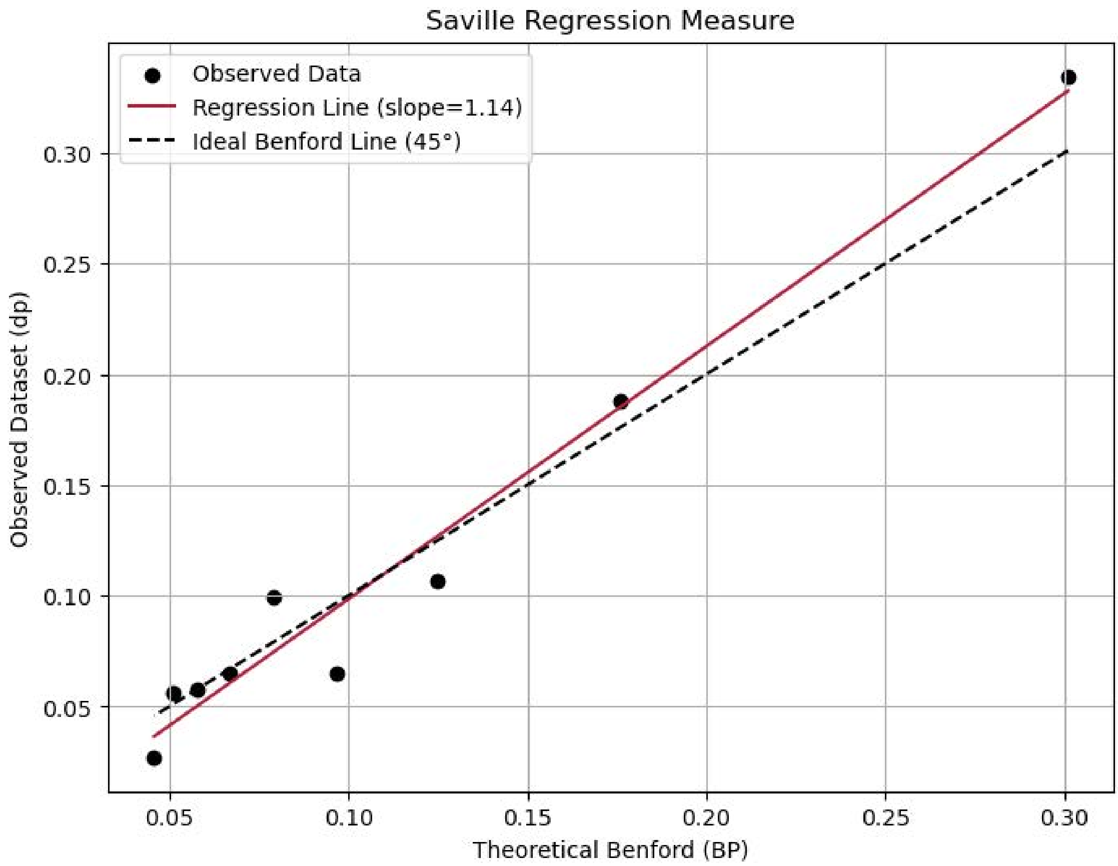

Figure 13: Saville Regression plot for the first digit of the variable `GENERAL RESULTS` against Benford's distribution

We apply the Equation (13) to obtain an estimate of the values of β_0 and β_1. From a graphical analysis, we can see that the bisector and the line representing the regression for our data are quite aligned. The closeness of the regression line to the $45°$ degree line suggests that US data behaviour is quite similar to the Benford's distribution. The slope value (β_1) is 1.1418 and the intercept value (β_0) is -0.0158. The value of β_1 is slightly higher than the ideal 1 and suggests that our data are more biased than the Benford's distribution. The intercept is instead very close to the ideal value 0 and so β_1 and β_0 indicate that our data seem to follow Benford's distribution, without important anomalies or irregularities.

Now we compute some metrics to understand whether the results of the statistical test are confirmed or not even without considering the number of entries of our dataset. The first metric that we want to apply to our dataset is the Mean Absolute Deviation (MAD), already defined above (Subsection 2.6.1). We apply Equation (15) to our 9 proportions and the result we obtain is 0.0157. This value obtained is higher than 0.015 and suggests a non-conformity of our data with Benford's law, following the table of critical values proposed by Professor Nigrini (Table 2). It is interesting to recall that the values of the MAD metric for the data coming from the Italian election were instead 0.00508, indicating a close conformity to Benford's law even if all the statistical tests suggested the opposite. In this case, the MAD metric suggests a non-compliance with Benford's law, probably because there are too few rows, although the number of rows only indirectly enters the function. A larger number of rows would probably allow us to refine the proportions. To determine whether the inconsistency of the results with statistical tests is specific to this metric, we can apply two additional metrics defined above: the mean of the First Significant Digit (FSD) and the sum of Squared Deviations (SD).

Regarding the mean of the First Significant Digits (FSD), the formula (16) is applied. We obtain a value of 3.27667. Recalling that the Benford's law FSD mean is 3.44024, we can compare it with our value to obtain some useful indications. Following the scheme written in Subsection 2.6.2, we can conclude that our statistic indicates an over-representation of smaller digits. However, this over-representation is not too much of a problem, because the value is close to the ideal one (3.44). The result of this metric is consistent with the Saville regression analysis, where the value of β_1 higher than 1 indicated that our data are skewer than Benford's law.

The last metric that we want to compute is the sum of Square Deviations (SD) defined in (Kossovsky, 2014). Applying the formula written in Subsection 2.6.3, we obtain a value of 33.9691. Using the evaluation scheme proposed by Kossovsky and reported in Table 3, we can conclude that the US data for the presidential election are marginally Benford. Again, this value is not as strong as the result obtained by the statistical tests already computed, but is not suggesting that our data are not following Benford's law: no evident errors or anomalies are highlighted.

Since we know that the application of Benford's law to the first digit is discussed in the literature (Nigrini, 2012), we can continue and analyse the second digit and the first-two to see if results of the first digit are confirmed.

3.3.3. *Second digit analysis*

Before analysing the second digit, we must remember that it has one more number in its range, the number 0. There are also fewer rows eligible for that kind of analysis because we have to eliminate all rows that do not have a second digit. These are the rows in which a candidate president has totalled 1 to 9 votes. In the end, our eligible dataset for the second-digit analysis now consists of 453 rows, 100 rows less than the one for the first-digit analysis.

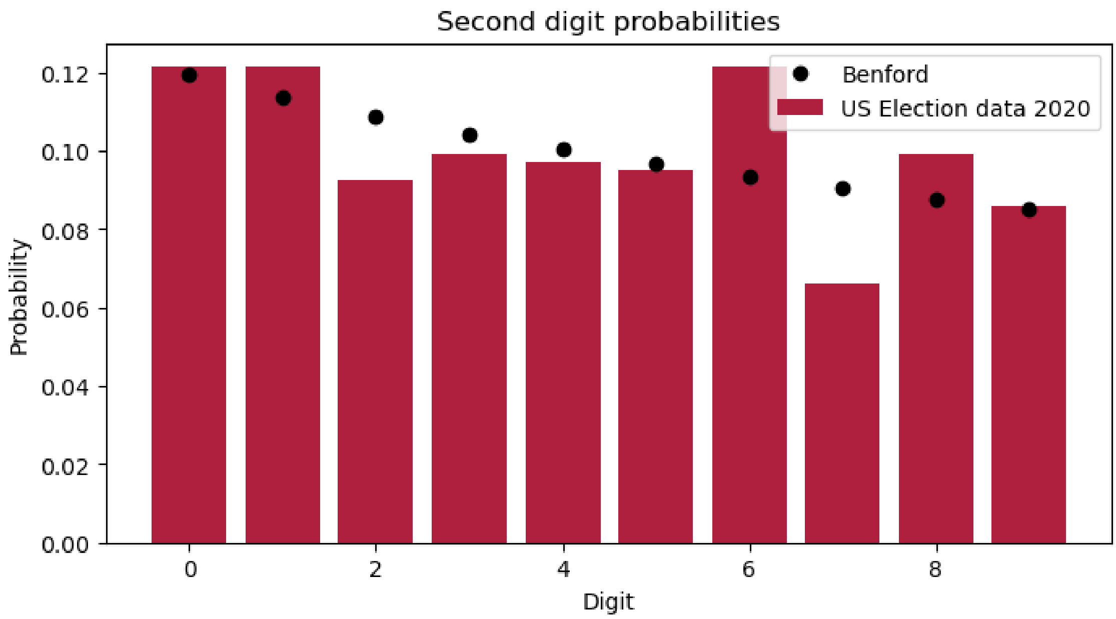

Figure 14: Comparison between US election data frequency and Benford's probabilities for the second digit position

From Figure 14 we can preliminarily analyse the behaviour of our data. The frequencies of our data are slightly decreasing from lower to higher digits, but there are some numbers for which the difference between the theoretical frequency and the empirical one is graphically significant. In order to see whether these differences noted graphically are relevant in statistical terms, we compute some statistical tests. The main assumption at the base of these tests is the one made above: we want to verify, assuming that our data follow Benford's distribution, if some errors or anomalies are encountered.

Looking at Table 10, it is possible to state that the larger differences between Benford theoretical probabilities and US data occur where the second digit is 6 or 7. At these

Digit	Benford	U.S. data	Difference	Z P-values	χ^2 p-value	K-S p-value
0	0.120	0.121	0.001	0.967	0.443	1.000
1	0.114	0.121	0.007	0.667	0.443	1.000
2	0.109	0.093	-0.016	0.305	0.443	1.000
3	0.104	0.099	-0.005	0.787	0.443	1.000
4	0.100	0.097	-0.003	0.883	0.443	1.000
5	0.097	0.095	-0.002	0.963	0.443	1.000
6	0.093	0.121	0.028	0.049	0.443	1.000
7	0.090	0.066	-0.024	0.087	0.443	1.000
8	0.088	0.099	0.011	0.422	0.443	1.000
9	0.085	0.086	0.001	0.933	0.443	1.000

Table 10: Comparison between second digit of Benford's probabilities and second digit of United States election data of 2020 with p-values of Z test, chi-squared test and Kolmogorov-Smirnov test.

values, the difference is expressed in terms of percentages equal to 2.8% in the case of number 6 and -2.4% in the case of number 7. Another observation we can deduce from Table 10 is that the frequencies of our data do not strictly decrease like the Benford's one. These two factors combined lead us to doubt whether our dataset follows Benford's law.

For that reason, we compute the Z statistic, already defined in Subsection 2.5.1. The p-value related to the result is reported in the fifth column. Keeping fixed the same levels of α as before ($\alpha = 0.001, 0.01, 0.05$) corrected with Bonferroni's correction, we cannot reject the null hypothesis H_0 for all digits. The results of this test are quite satisfactory and cannot rule out our initial hypothesis that there may be some sort of conformity between our data and Benford's law for the second digit. From the results of the Z-test it is possible to exclude that US data have been tampered or that there are some errors in reporting. Unfortunately, however, knowing that we do not have too many rows, we cannot really rely on this test because it is calculated separately for each possible value of the second digit.

The second statistic that we want to apply is the chi-squared test. The p-value related to the result of this test is reported in the sixth column of Table 10 and is equal for all ten possible values of the second digit. This is due to the formula of the chi-squared statistic, reported in (9). The p-value we obtain is equal to 0.443. We do not reject the null hypothesis H_0 for all the three levels of significance alpha (α) corrected with Bonferroni's correction. We recall that Bonferroni's correction is necessary because we are computing the 115 test hypothesis on the same dataset. In other words, the result of this test suggests that we cannot reject the hypothesis that our data follow Benford's distribution, and so no errors or anomalies are highlighted.

The Kolmogorov-Smirnov test is also applied, and its p-value related to the result is reported in the last column of the Table 10. Again, as with the chi-squared, the result is unique for all ten digits. The value obtained from the test is very close to 1, precisely 0.9996. This value corresponds to a supremum of 0.0168. This is the highest difference between the two cumulative distribution functions and it occurs when the value of the second digit is 5. This p-value makes us not reject the null hypothesis for all previously set alpha levels, concluding that, based on the K-S test, we cannot exclude the hypothesis that our data are consistent with Benford's law.

To obtain further useful information on how the values in the US dataset are distributed and whether they are more or less skewed than the Benford distribution, we use the Saville regression. We apply the formula written in (13) to obtain the values of β_0 and β_1.

From Figure 15 we can see that the points are spread around the graph: only some points are close to the bisector, the ideal $45°$ line. For these scattered points the graph is signaling a discrepancy between our data and Benford's distribution, already highlighted in Table 10. The value obtained for β_1 is 0.8871. This slope, less than the ideal value 1, indicates that the frequencies of our data increase less than the theoretical one. In some sense, our data frequencies are flatter. In terms of β_0, the Saville regression returns a value of 0.0113. However, in this case, the intercept suggests that the frequencies of our data set are shifted upward with respect to the theoretical values, but the value is very close to the ideal one, 0. For that reason, the regression line and the ideal Benford line are quite well aligned.

To better understand the behaviour of our data, we can also apply the three metrics defined above, recalling that we do not have to take into consideration the number of rows that we have.

The Mean Absolute Deviation (MAD) is the first metric applied to the second digit. Using the formula defined in (15) we substitute as K the number 10 that is the possible

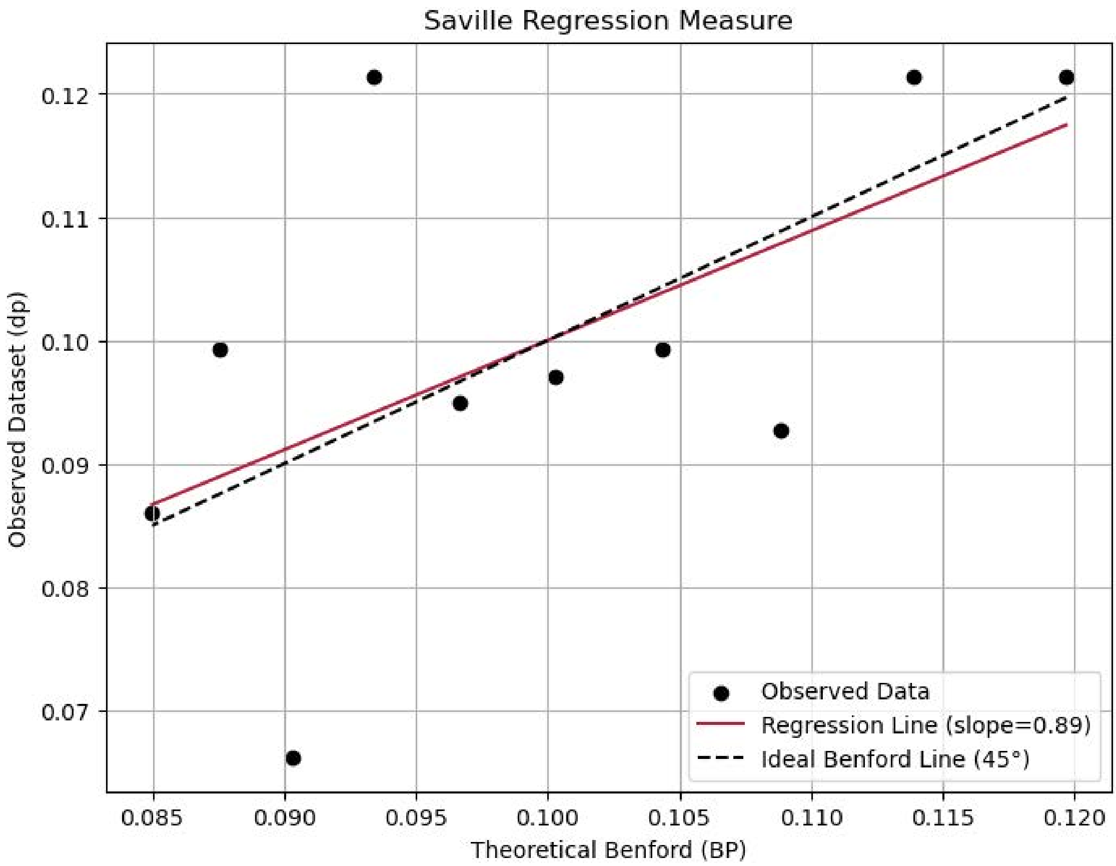

Figure 15: Saville Regression plot for the second digit of the variable `GENERAL RESULTS` against Benford's distribution

number of values that a second digit can assume. The MAD result obtained is 0.01003. Following the values defined by Nigrini and reported in Table 2, we can say that our dataset exhibits a marginally acceptable conformity to Benford's law.

As far as the mean of the Second Significant Digits is concerned, we can apply equation (17). The result obtained is 4.2296, larger than the reference value of Benford's law (4.187) but smaller than the value of a uniform distribution. The value of obtained from this metric is quite close to the ideal one, signaling a slight over-representation of larger digits but still a right-skewed distribution.

The last metric that we want to apply is the sum of Squares Differences. The result obtained is 18.65 that, following the decision scheme defined above in Subsection 2.6.3 suggests that our dataset is marginally Benford, with respect to the second digit. This value is not so close to 0, the ideal value, but we know that this metric considers outliers too much, and from Figure 15 we know that there are some values for which the difference with Benford's law is relevant.

In summary, the results obtained from the statistical tests do not suggest anomalies or errors in our dataset, since we know that we cannot reject the null hypothesis for almost all cases. However, the results obtained for the three metrics are a little more doubtful. They do not categorically rule out the fact that our dataset follows Benford's law, but neither do they confirm it with certainty.

3.3.4. First-two digit analysis

To complete the analysis and to observe whether any obvious manipulations have occurred, we need to compute the analysis also on the first-two digits. The number of rows eligible for this type of analysis is 453.

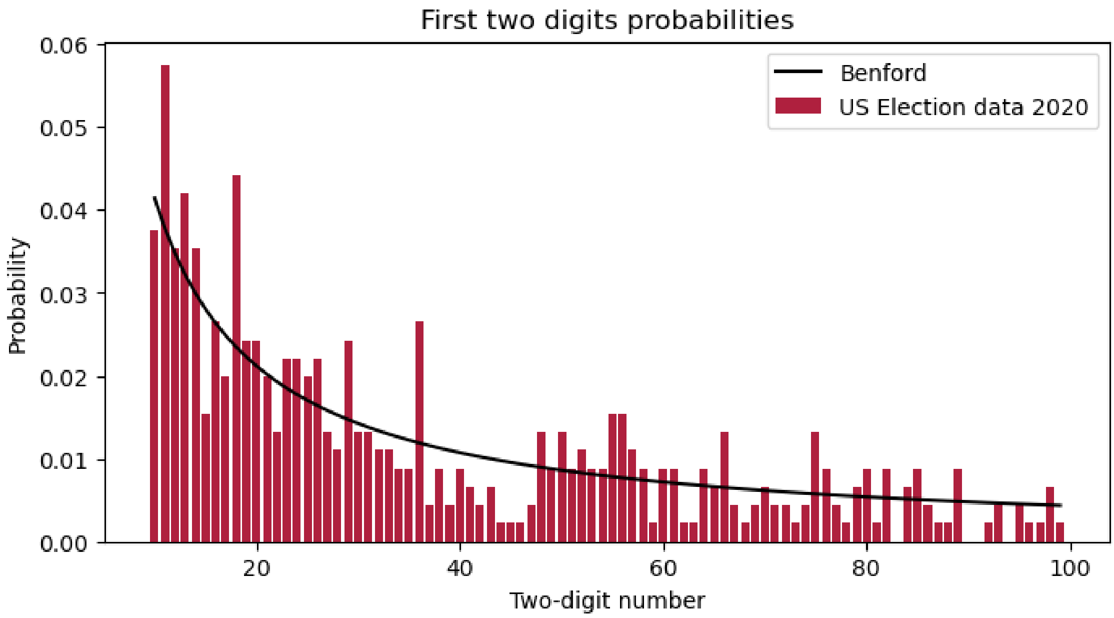

Figure 16: Comparison between Us election data frequency of 2020 and Benford's probabilities for the first-two digits position

From the graphical representation, we can immediately see that not all values are well allineated with the Benford distribution. The trend of the relative frequencies of the first-two digits of 2020 US election data is decreasing from lowest digits to the highest one, but is not very clear if the dataset is coherent with Benford's law.

Digits	Benford	U.S. data	Difference	Z P-values	χ^2 p-value	K-S p-value
18	0.023	0.044	0.021	0.006	0.279	0.138
36	0.012	0.026	0.014	0.008	0.279	0.138
11	0.038	0.057	0.019	0.039	0.279	0.138
75	0.006	0.013	0.007	0.072	0.279	0.138
56	0.008	0.015	0.007	0.104	0.279	0.138
55	0.008	0.015	0.007	0.115	0.279	0.138
29	0.015	0.024	0.009	0.135	0.279	0.138
66	0.007	0.013	0.006	0.138	0.279	0.138
15	0.028	0.015	-0.013	0.139	0.279	0.138
44	0.010	0.002	-0.008	0.163	0.279	0.138

Table 11: Comparison between first-two digits of Benford's probabilities and first-two digits of US election data 2020 with p-values of Z test, chi-squared test and Kolmogorov-Smirnov test. Only the 10 first-two digits with the lowest p-value of the Z statistic are represented.

In Table 11 we highlighted the ten digits with the highest Z p-values. These digits correspond to the ones with the larger differences between the US data and Benford's law and can also be noticed in Figure 16. Since the Z test is computed digit by digit, for the first-two digits, we have 90 tests. Keeping the same corrected levels of α already used in the document fixed, we cannot reject the null hypothesis H_0 (5) for all three levels. For this reason, as far as the Z test is concerned, it is impossible to exclude that the data of the US presidential election of 2020 follow Benford's law and so we do not have any type of evidence of errors or anomalies. Although it is important to recall that we have only 453 rows of values eligible for our analysis. The number of rows is not very large and probably not sufficient for such an analysis. For example, we know that there are some values with a relative frequency of 0, since there is no observation starting with these first two digits in the dataset of the 2020 US elections. This occurs when the value of the first-two digits is 83, 90, 91 and 94.

To have an idea on how the joint distribution complains or not with Benford's law we need to calculate some statistical tests on the entire possible values assumed by the variable `GENERAL RESULTS` of the US dataset. The chi-squared statistic is calculated using the formula (10). Knowing that the number of possible values assumed by the first-two digits is 90, we have 89 degrees of freedom. The p-value returned from the function is 0.279 which, for all three levels of alpha corrected with Bonferroni's correction ($\alpha = 0.0000087, 0.0000870, 0.0004348$), allows us to say that we cannot reject the null hypothesis H_0 that our data belong to Benford's distribution. The conclusions for the Kolmogorov-Smirnov test are the same. This test returns a p-value of 0.138, higher than those of the three alpha levels. This p-value is encountered when the first-two digits assume a value of 36. At this point, we have the highest difference between the two cumulative distribution functions. The two p-values, that of the chi-squared test and that of the K-S test, are shown in the last two columns of the table and are equal for all the possible values of the first two digits.

In addition it is possible to use Saville's regression for a better understanding of the general behaviour and the skewness of our data compared to a baseline, Benford's law.

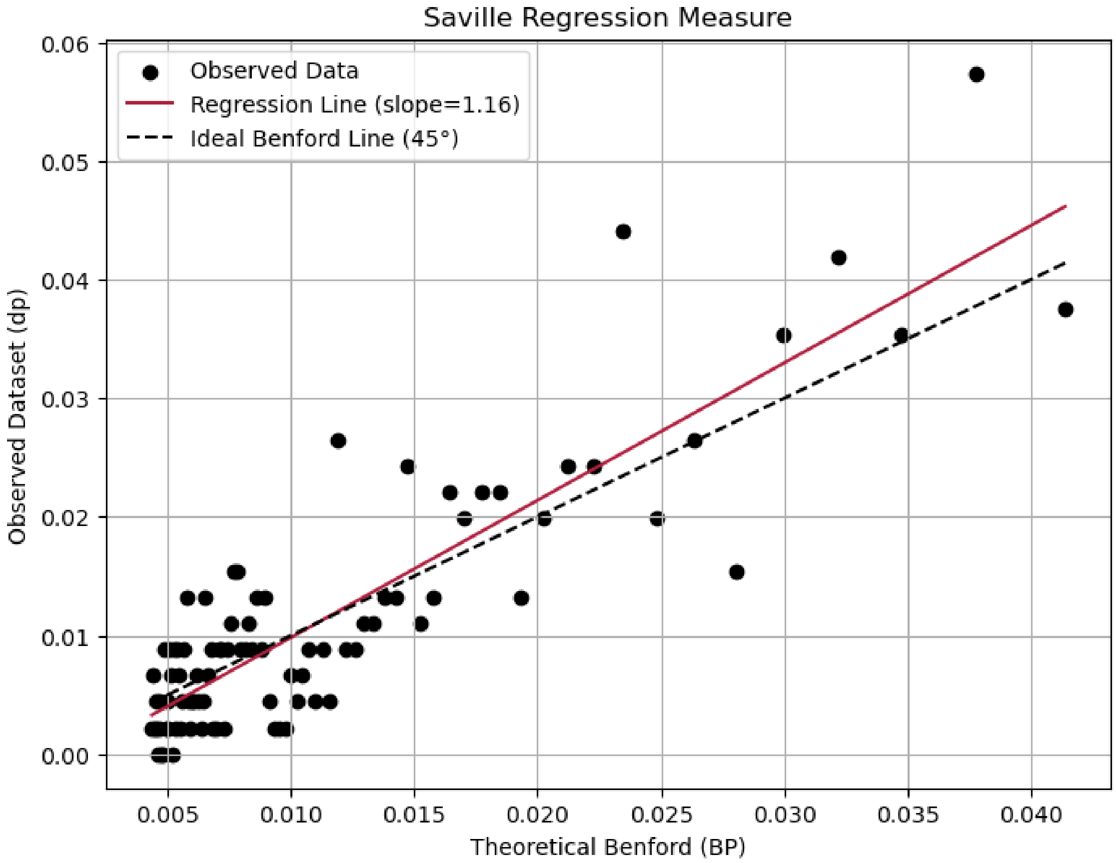

Figure 17: Saville Regression plot for the first-two digit of the variable `GENERAL RESULTS` of US data election of 2020 against Benford's distribution

The points in Figure 17 are not so far from the bisector, which represents the Benford's line. The regression line, indicated in red, is very close to the $45°$ line. The intercept of the regression line, based on data from the US presidential elections of 2020 is -0.0017. This value is very close to 0, our target, and does not indicate particular problems or strange behaviours. Unlike the intercept (β_0), the slope value β_1 is different from our ideal value of 1. The slope of the regression line is 1.16, a value that indicates that the observed values increase slightly faster than the theoretical one. The regression line is steeper than the ideal line and this implies that the observed dataset amplifies theoretical probabilities of the lowest values of the first-two digits. From the graph, we can also see that there are certain points for which the relative proportions of the dataset are 0. These points correspond to the value already mentioned and highlighted in the Z test comment and correspond to the points 83, 90, 91 and 94. In any case, the conclusion is that even the Saville regression cannot exclude that data of 2020 US election are coherent with Benford's law.

To partially circumvent the problem of few observations in our dataset, we can apply the metrics defined above that do not take the number of observations into account directly. The Mean Absolute Deviations applied to the first-two digits return a value of 0.00397 that indicates non-conformity with Benford's law, according to the p-values defined in Table 2. This metric therefore suggests that our data may have been manipulated or are incorrect.

The mean of the First-Two Significant Digits (FTSD) returns a value of 36.481. The indication from the result of this metric is that our data are more right-skewed than Benford's distribution. This because the value obtained is smaller than 38.59, the FTSD of Benford's law. However, the result is close to the ideal value, giving us the conclusion that we cannot exclude that our data can belong to Benford's distribution.

The last metric computed is the sum of Squared Deviations (SD). Applying the formula suggested by Kossovsky and defined in (19) we obtain a value of 26.03. Following the decision scheme written in Table 3, this metric indicates that our dataset is marginally Benford. Again, we cannot exclude the possibility that our data are coherent with the Benford's distribution.

In conclusion, taking into consideration the statistical tests and the metrics for the first, second and first-two digits, we do not have sufficient proofs to exclude that our data set does not belong to Benford's distribution. For this reason, it is not possible to claim that, through the combined use of statistical tests and metrics, the data from the 2020 US elections have any errors, anomalies or tampering. This conclusion is certainly supported more by statistical tests than by metrics. However, we know that metrics suffer from the fact that there is not much data available in our dataset.

4 Conclusions

This thesis highlights how the Benford's law can be used to determine if some fraud or anomalies had occurred. Some metrics and some statistical tests were applied to two different data sets: Italian election of the Camber of Deputies of 2018 and the US presidential election of 2020. These two datasets are very different in terms of detail and number of rows eligible for our analysis. In the Italian dataset there are 122149 rows, instead, in the US dataset rows are only 553, 0.453% of the total Italian rows. This different number of rows allows us to perform the same tests and metrics, but highlights different types of problem we may encounter when varying the number of rows in a dataset. In particular, we verified that, as far as Italian election regard, we encounter an excess power problem. In this dataset, even if the graphical analysis and the metrics are quite coherent with Benford's law, the general suggestion from statistical tests is that we can reject the null hypothesis H_0 that our data are coherent with Benford's distribution and so Benford's law. The three calculated statistical tests suggest that there may have been errors in data reporting, anomalies or tampering. In contrast, in the dataset that contains the US data of presidential elections of 2020, the suggestion of partial or limited compliance of our data with Benford's law comes from the metrics defined in this document. However, in this case, we have clear and straightforward results in the statistical test that suggest us to not reject the null hypothesis H_0 and to not exclude the possibility that our dataset could respect Benford's distribution. The conclusions of our analysis applied to two different datasets are to some extent the opposite. The results of the statistical tests and metrics are different but lead us to the same conclusion. From the data in our possession it cannot be ruled out that the datasets follow Benford's law and this result is derived either with metrics or with statistical tests. It is essential to apply both statistical tests and metrics together, as relying on only one could lead to misinterpretations and wrongly classify some Benford's data as manipulated.

4.1. Limitations and suggestions for further research

The problem of a large sample size can be overcome by undersampling the data or by reducing the granularity of the information in our possession. If, in the case of the Italian electoral data, instead of taking the data at the level of the municipality, we had taken them at the level of the statistical zone, we would have had fewer rows in our dataset. In contrast, our metrics would have been less refined and precise and we would have risked obtaining a non-compliant result. Another possible solution would have been to choose only one Italian region, but even then we would have lost some detail in the data and, moreover, our decision on which region to choose would have been completely arbitrary. Other methods to overcome the problem of larger sample size were discussed in Campanelli (2022), but are not developed here.

It is known that Benford's law can be generalised for other bases than the decimal base. Following what was written above in subsection 2.4.2, this generalisation is useful not only to understand how Benford's law works but also to check the conformity of a data set. For this reason, an extension to other numerical bases can provide a more complete picture of the behaviour of our data and any anomalies detected. In this thesis, for reasons of time and space, this extension is not calculated but can easily be applied with a minimum of effort.

Another extension not developed in this work is the application of the law to variables other than the two chosen for our analyses. Not only can variables related to voting be used for this type of analysis, but also variables related to the number of voters, the number of preferences, and other types of variables in our dataset can be used for fraud detection purpose (Mebane Jr, 2010).

Two other extensions of this paper can be related to metrics and tests and other datasets. Certainly, there are other statistical tests and metrics that can be applied to our cases and datasets. In this paper, only the most popular and well-supported ones in the literature have been cited and applied, but nothing prevents us from applying other tests or metrics.

The Section 3 that refers to application of Benford's law to electoral data can be extended to other election-related datasets. Furthermore, some datasets from different elections can be mixed together firstly to confirm the adherence of election data to Benford's law and secondly to create critical values for metrics to be reused and applied in other analyses.

References

Barabesi, L., Cerasa, A., Cerioli, A., and Perrotta, D. (2022). On characterizations and tests of benford's law. *Journal of the American Statistical Association*, 117(540):1887–1903.

Becker, P. W. (1982). Patterns in listings of failure-rate & mttf values and listings of other data. *IEEE Transactions on Reliability*, 31(2):132–134.

Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American philosophical society*, pages 551–572.

Berger, A. and Hill, T. P. (2015). *An introduction to Benford's law*. Princeton University Press.

Campanelli, L. (2022). Testing benford's law: from small to very large data sets. *Submitted to Spanish Journal of Statistics*.

Campanelli, L. (2025). Tuning up the kolmogorov–smirnov test for testing benford's law. *Communications in Statistics-Theory and Methods*, 54(3):739–746.

Cantú, F. and Saiegh, S. M. (2011). Fraudulent democracy? an analysis of argentina's infamous decade using supervised machine learning. *Political Analysis*, 19(4):409–433.

Corazza, M., Ellero, A., and Zorzi, A. (2018). The importance of being "one"(or benford's law). *Lettera Matematica*, 6:33–39.

Corazza, M., Ellero, A., Zorzi, A., et al. (2008). What sequences obey benford's law? *WORKING PAPER SERIES-DEPARTMENT OF APPLIED MATHEMATICS, UNIVERSITY OF VENICE*, 185:0–5.

Crocetti, E. and Randi, G. (2016). Using the benford's law as a first step to assess the quality of the cancer registry data. *Frontiers in public health*, 4:225.

Goeman, J. J. and Solari, A. (2014). Multiple hypothesis testing in genomics. *Statistics in medicine*, 33(11):1946–1978.

Grendar, M., Judge, G., and Schechter, L. (2007). An empirical non-parametric likelihood family of data-based benford-like distributions. *Physica A: Statistical Mechanics and its Applications*, 380:429–438.

Kossovsky, A. E. (2014). *Benford's law: theory, the general law of relative quantities, and forensic fraud detection applications*, volume 3. World Scientific.

Ley, E. (1996). On the peculiar distribution of the us stock indexes' digits. *The American Statistician*, 50(4):311–313.

Lioy, A. (2021). The blank ballot crisis: a multi-method study of fraud in the 2006 italian election. *Contemporary Italian Politics*, 13(3):352–381.

Mebane Jr, W. R. (2010). Fraud in the 2009 presidential election in iran? *Chance*, 23(1):6–15.

Miller, S. J. (2015). *Benford's law*. Princeton University Press.

Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. *American Journal of mathematics*, 4(1):39–40.

Nigrini, M. J. (1996). A taxpayer compliance application of benford's law. *The Journal of the American Taxation Association*, 18(1):72.

Nigrini, M. J. (2012). *Benford's Law: Applications for forensic accounting, auditing, and fraud detection*, volume 586. John Wiley & Sons.

Noether, G. E. (1963). Note on the kolmogorov statistic in the discrete case. *Metrika*, 7(1):115–116.

Pericchi, L. and Torres, D. (2011). Quick anomaly detection by the newcomb—benford law, with applications to electoral processes data from the usa, puerto rico and venezuela. *Statistical science*, pages 502–516.

Pericchi, L. R. and Torres, D. (2004). La ley de newcomb-benford y sus aplicaciones al referéndum revocatorio en venezuela. *Reporte Técnico no-definitivo 2a. versión: Octubre*, 1(2004):12.

Pinkham, R. S. (1961). On the distribution of first significant digits. *The Annals of Mathematical Statistics*, 32(4):1223–1230.

Raimi, R. A. (1976). The first digit problem. *The American Mathematical Monthly*, 83(7):521–538.

Saville, A. D. (2006). Using benford's law to detect data error and fraud: an examination of companies listed on the johannesburg stock exchange: economics. *South African journal of economic and management sciences*, 9(3):341–354.

Thomas, J. K. (1989). Unusual patterns in reported earnings. *Accounting Review*, pages 773–787.