Paper prepared for the Second Euroacademia International Conference

Identities and Identifications: Politicized Uses of Collective Identities

Florence, 18 – 19 October 2014

This paper is a draft

Please do not cite

Testing the Reliability of 2011 Census-Albania Using Benford's Law

PhD Candidate Alma Përmeti
Ministry of Interior Affairs of Albania

PhD Candidate Aurela Biçaku
University "Aleksander Moisiu" Durres, Albania

Prof. Dr. Fejzi Kolaneci
University of New York Tirana

Abstract

The main purpose of the study is to test the hypothesis whether the second digit of demographic data 2011 Census-Al obeys Benford's law. We consider 2827 data divided into five groups. The source of official data is INSTAT.

The results of this study include:

1. The mean of second digit for all five groups of the data converge to Benford's law, at the confidence level 95%.
2. The data for group 1 obeys Benford's law at the confidence level 95%.
3. The data for group 2 obeys Benford's law at the confidencde level 95%.
4. The data for group 3 contradicts Benford's law at the confidence level 99.5%.
5. The data for group 4 obeys Benford's law at the confidencde level 95%.
6. The data for group 5 contradicts Benford's law at the confidence level 95%.
7. The data for groups 1,2,3,4,5 together contradict Benford's law at the confidence 90%. Therefore, these 2827 official data obtained from 2011 Census-Al are suspectable for manipulation.

Key words: 2011 census, Albania, Benford's law, second digit, chi-squared test, manipulation

1. Introduction

In the present study we develop a statistical analysis, based on Benford's law, for the second digit of the demographic data 2011 Census-Albania. Let us explain briefly Benford's law. The universality is a remarkable phenomenon in Modern Probability Theory: many seemingly unrelated probability distributions of random variables which involve (contain) large numbers of unknown parameters, can end up converging to a universal law that may only depend on a small handful of parameters. One of the most famous examples of the universality phenomenon is the Kolmogorov's Central Limit Theory.

Analogous universality phenomena also show up in empirical distributions – the probability distribution of a random variable X from a large population of "real-world" objects. Examples include Benford's law and Zipf's law.

Benford's law governs the asymptotic probability distribution of many (but not all) real-life sources of data (random variables) X which satisfy the following assumptions:

1. The set of possible values of X is $R^+ = (0, +\infty)$.
2. X range over many different orders of magnitude.
3. X arise from a complicated combination of largely random independent factors, with different random samples of X selected from different independent factors.
4. The data have not been artificially rounded, truncated, or otherwise manipulated.

Like the Central Limit Theory, Benford's Law is an empirically observable phenomenon.

Benford's probability distribution is a second generation distribution, a complicated combination of other probability distributions.

If random variables are selected at random, and the samples are obtained from each of these random variables, then the combined samplings will converge to Benford's distribution, even though the individual distributions may not closely follow the Benford's law, see Hill (1995, 1998) and Tao (2009). The key is in the combining of data from different sources. Benford's distribution is a distribution of distributions. Janvresse and la Rue (2004) advanced the similar probabilistic explanation for the appearance of Benford's law in everyday-life numbers, when we consider mixtures of uniform distributions. Benford's law reflects a profound harmonic truth of nature.

Definition 1
A set of positive real numbers is said to satisfy Benford's law if the leading digit d occurs with probability

$$P(d) = \log_{10}(1 + \frac{1}{d}) \, for d = 1,2, \dots ,9$$

According to Benford's law, the leading digit d follows the following probability distribution:

d	1	2	3	4	5	6	7	8	9
$P(d)$	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.049

Generalization of Benford's law to digits beyond the first:

$$P(second significant digit = d) = \sum_{k=1}^{9} \log_{10}(1 + \frac{1}{10k + d}), for d = 0,1,2, \dots 9$$

$$P(third significant digit = d) = \sum_{k=10}^{99} \log_{10}(1 + \frac{1}{10k + d}), for d = 0,1,2, \dots 9$$

$$P(n^{th} significant digit = d) = \sum_{k=1}^{10^{n-1}-1} \log_{10}(1 + \frac{1}{10k + d}), for d = 0,1,2, \dots 9 \; and n = 2,3,4,5, \dots$$

In particular, the probability of the random event that a positive real number starts with the string of digits n is calculated by the formula:

$$P(string of digits n) = \log_{10}(1 + \frac{1}{n})$$

For example, the probability that a positive real number starts with digits 3, 1, 4 is equal to

$$\log_{10}(1 + \frac{1}{314}) \approx 0.0014$$

Corollary 1

The significant digits are dependent (and not independent as one might expect).

For example, the unconditional probability that the second digit is 2 is 0.10882, but the conditional probability that the second digit is 2, given the first digit is 1, is $\log_{10}\left(1 + \frac{1}{12}\right) : 0.30103 = 0.11548$.

Table 1 shows the expected relative frequencies for all digits 0 through 9 in each of the first four places in any positive real number, based on Benford's law.

Table 1. Expected relative frequencies based on Benford's law

Digit	1st place	2nd place	3rd place	4th place
0		0.11968	0.10178	0.10018
1	0.30103	0.11389	0.10138	0.10014
2	0.17609	0.10882	0.10097	0.10010
3	0.12494	0.10433	0.10057	0.10006
4	0.09691	0.10031	0.10018	0.10002
5	0.07918	0.09668	0.09979	0.09998
6	0.06695	0.09337	0.09940	0.09994
7	0.05799	0.09035	0.09902	0.09990
8	0.05115	0.08757	0.09864	0.09986
9	0.04576	0.08500	0.09827	0.09982

Source: Nigrini (1996)

Definition 2

The base 10 mantissa of an arbitrary positive real number x is the unique real number $u \in [0.1, 1)$ such that $x = u \times 10^n$ for some integer $n \in Z$.

Examples

$$mantissa(356) = 0.356,$$

$$mantissa(3560) = 0.356,$$

$$mantissa(35.6) = 0.356,$$

$$mantissa(0.356) = 0.356,$$

$$mantissa(2) = 0.2,$$

$$mantissa(20) = 0.2, etc$$

The general form of the Benford's law:

$$Prob(mantissa of any positive real number x \leq u) = \log_{10}(10u), \forall u \in [0.1, 1)$$

Benford's law is stated here for base 10, which is what we are most familiar with, but the Benford's law holds for any base, after replacing all the occurrences of 10 in the above law with the new base, of course.

The Benford's law tends to break down if the assumptions 1-4 are dropped. For instance, if the random variable Xconcentrates around its mean μ (as opposed to being spread over many orders of magnitude), then the normal distribution tends to be an appropriate mathematical model, as indicated by the Kolmogorov's Central Limit Theorem. The independence property of the factors (see assumption 3) is crucial. If, for instance, population X growth always slowed down for some (inexplicable) reason to a crawl whenever the first digit of the population X was 6, then there would be a noticeable deviation from the Benford's law in digits 6 and 7, due to this bottleneck (bottle-neck).

Roughly speaking, Benford's law asserts that the bulk probability density distribution of $\log_{10} X$ is locally uniform at unit scale.

Benford's law enjoys scale invariance: this law should be independent of the unit chosen (by changing the unit of measurement); for example, using metric system of units versus English system of units. However, according to Knuth and Gnedenko, there is no scale invariant probability measure on the Borel subsets of R^+. Therefore, the Borel sigma-algebra (the smallest sigma-algebra containing all open intervals) is not the appropriate domain for the Benford's law. An appropriate probability domain $\mathcal{A}$ for the Benford's law is defined rigorously by Hill (1995), p323.

Definition 3
The appropriate domain $\mathcal{A}$ for the probability is the smallest (minimal) collection of subsets of the positive real numbers, which contains all sets of the form

$$\bigcup_{n=-\infty}^{\infty} (a,b) \times 10^n \, , fora > 0, b > 0$$

and which is closed under complements and countable unions.

Main properties of the appropriate domain $\mathcal{A}$

1. Every non-empty set in $\mathcal{A}$ is infinite, with accumulating points at 0 and $+\infty$.
2. $\mathcal{A}$ is closed under scalar multiplication: $\forall a > 0 \ and \forall S \in \mathcal{A} \Longrightarrow \mathrm{a} \times \mathrm{S} \in \mathcal{A}$.
3. $\mathcal{A}$ is self-similar in the sense that if $S \in \mathcal{A} and n \in Z$, then $10^n \times S = S$.

The definition of the appropriate domain $\mathcal{A}$ for the probability is the first step toward making rigorous sense of the Benford's law.

Theorem 1 (Hill, 1995)
On the appropriate probability domain $\mathcal{A}$, scale invariance implies Benford's law.

Theorem 2 (Tao, 2009)
If the random variable X obeys Benford's law and U is an arbitrary positive random variable independent of X, then the product $Y = X \times U$ obeys Benford's law, even if U did not obey this law.

Absorptive property
If a random variable Y is a product of n independent factors $X_1, X_2, \dots, X_n$ and if a single factor ($X_1 or X_2\ or \dots X_n$) obeys Benford's law, then the whole product

$$Y = X_1 \times X_2 \times \dots \times X_n$$

obeys Benford's law.

Benford's law is the unique probability distribution with this absorptive property. If there is another law with absorptive property, what would happen if one multiplied a random variable with that law with an independent random variable with Benford's law?

Diaconis and Freedman (1979), p363, offer convincing evidence that Benford's law manipulated the round-off errors to obtain an even better fit. But even the unmanipulated data seems a remarkably good fit, and Benford's law has become widely accepted. Examples of the random variables that obey Benford's law include:

- the populations of 235 countries and regions of the world in 2013 (using CIA world factbook),
- the populations of the cities of USA in 2013,
- the surface areas of the world states,
- the mass of astronomic objects,
- the specific head of thousands of chemical compounds,
- the surface areas of 335 rivers,
- the net worth of individuals of the USA in 2013,
- the net worth of corporates of the USA in 2013,
- the square-root table of natural numbers, etc.

The 1990 census populations of the 3141 counties in the USA follow Benford's law very closely, see Nigrini and Wood (1995).

T. Hill (1995, 1998) noted that Benford's law is applied to census statistics, stock market data and certain accounting data. For instance, the series of one-day return on the Dow-Jones Industrial Average Index (DJIA) and the Standard and Poor's Index (S&P) reasonably agrees with Benford's law.

Benford's law has been promoted as providing the auditor with a tool that is effective and simple for the detecting fraud in a population of accounting data, see Durtschi, Hillson and Pacini (2004).

The contemporary bibliographic database on Benford's law includes more than 180 scientific studies, see

http://www.benfordonline.net

In the present study we apply Benford's law to investigate (for investigating) results of the 2011 Albanian population and housing census, see 2011 Census-Al.

The rest of the paper is organised as follows:

- Section 2 contains the five groups of the data set.
- Section 3 provides the investigation of expectation for second digit.

- Section 4 presents the hypothesis testing, whether the data for each group follow Benford's law.
- Section 5 concludes the paper.

2. Statistical analysis of the data

The data set includes five different groups for second significant digit of the demographic data 2011 Census-Albania. The source of the official data is INSTAT (Albanian Institute of Statistics).

The first group of the data contains resident population by age group, urban or rural area, sex, martial status of the municipality/commune, see 2011 Census-Al, pp 47-53.

The second group of the data contains resident population by sex, age and country of birth (Albania, Greece, Italy, USA, Kosovo, Turkey, Canada, Macedonia, other countries) and place of resident on 1 April 2001, see 2011 Census-Al, pp 54-57.

The third group of the data contains resident female population 15 years and over by urban or rural areas, age group, number of children ever-born alive and number of children still alive, see 2011 Census-Al, pp 58-61.

The fourth group of the data contains resident institutional population by age group and sex, ownership of institution, type of collective living quarter and sex, see 2011 Census-Al, pp 62-64.

The fifth group of the data contains resident Albanian citizens ever residing abroad who returned after 1 January 2011 by age group, sex and year of returning, country of previous residence and year of returning, see 2011 Census-Al, pp 65-69.

Group 1 contains tables 1.1.1-1.1.3

Group 2 contains tables 1.1.4-1.1.5

Group 3 contains tables 1.1.6-1.1.7

Group 4 contains tables 1.1.8-1.1.9

Group 5 contains tables 1.1.10-1.1.11

Using SPSS, IBM version 20, we develop the statistical analysis of the data (second significant digit), see Field (2009).

We calculate the statistical parameters for the first group of the data.

Sample size	670
Sample mean	4.5682
95% confidence interval for mean	(3.9743 , 5.1621)

Variance	7.857
Standard deviation	2.8031
Coefficient of variation	61.36%
Minimum	0
Maximum	9
Range	9
Skewness	0.046
Kurtosis	-1.347

The statistical parameters for the second group of the data.

Sample size	651
Sample mean	4.0568
95% confidence interval for mean	(3.4512 , 4.6624)
Variance	8.169
Standard deviation	2.85817
Coefficient of variation	70.45%
Minimum	0
Maximum	9
Range	9
Skewness	0.159
Kurtosis	-1.237

The statistical parameters of the third group of data

Sample size	616
Sample mean	3.8977
95% confidence interval for mean	(3.3036 , 4.4919)
Variance	7.863
Standard deviation	2.8041
Coefficient of variation	71.94%
Minimum	0
Maximum	9
Range	9
Skewness	0.312
Kurtosis	-1.133

The statistical parameters for the fourth group of the data

Sample size	88
Sample mean	4.1023
95% confidence interval for mean	(3.5240 , 4.6806)
Variance	7.449
Standard deviation	2.72932
Coefficient of variation	66.53%
Minimum	0

Maximum	9
Range	9
Skewness	0.171
Kurtosis	-0.998

The statistical parameters for the fifth group of the data

Sample size	802
Sample mean	3.8068
95% confidence interval for mean	(3.2752 , 4.3384)
Variance	6.296
Standard deviation	2.5091
Coefficient of variation	65.91%
Minimum	0
Maximum	9
Range	9
Skewness	0.434
Kurtosis	-0.652

3. Benford's analysis for the second digit of the data

Assume that the second digit of the demographic data 2011 Census-Al follows the probability distribution implied by Benford's law. Under this assumption, calculate the expectation of the second digit μ=4.18739

The 95% confidence interval for mean of each group of data is calculated in section 2:

(3.9743 , 5.1621) for the first group

(3.4512 , 4.6624) for the second group

(3.3036 , 4.4919) for the third group

(3.5240 , 4.6806) for the fourth group

(3.2752 , 4.3384) for the fifth group

Answer

$\mu \approx 4.18739 \in$ (3. 9743 , 5.1621), $\mu \in$(3.4512 , 4.6624), $\mu \in$(3.3036 , 4.4919), $\mu \in$(3.5240 , 4.6806), $\mu \in$(3.2752 , 4.3384), at the confidence level 95%.

In other words, the mean of second digit for all five groups of the data selected from 2011 Census-Al are in very good accord with Benford's law (at the confidence level 95%).

Table 2 contains the observed frequencies of second digit for each group 1,2,3,4,5 of the demographic data obtained from 2011 Census-Al.

Table 2. The observed frequencies of second digit

Digit	Group 1	Group 2	Group 3	Group 4	Group 5	Total
1	73	92	50	7	97	319
2	79	69	87	7	87	329
3	70	70	83	15	100	338
4	64	71	58	12	76	281
5	56	50	72	10	99	287
6	61	68	55	4	65	253
7	74	47	44	8	70	243
8	72	55	40	9	58	234
9	56	50	58	5	76	245
0	65	79	69	11	74	298
Total	670	651	616	88	802	2827

Table 3 contains the expected frequencies of second digit for each group of the demographic data obtained from 2011 Census-Al, according to Benford's law.

Table 3. The expected frequencies of second digit according to Benford's law

Digit	Group 1	Group 2	Group 3	Group 4	Group 5	Total
1	76.3	74.1	70.2	10	91.3	321.9
2	72.9	70.8	67	9.6	87.3	307.6
3	69.9	67.9	64.3	9.2	83.7	295
4	67.2	65.3	61.8	8.8	80.4	283.5
5	64.8	63	59.6	8.5	77.5	273.4
6	62.6	60.8	57.4	8.2	74.9	263.9
7	60.5	58.8	55.7	8	72.5	255.5
8	58.7	57	53.9	7.7	70.2	247.5
9	56.9	55.4	52.4	7.5	68.2	240.4
0	80.2	77.9	73.7	10.5	96	338.3
Total	670	651	616	88	802	2827

Consider each group 1,2,3,4,5 separately. O_k denotes the observed frequencies of the second digit and e_k denotes the expected frequencies of the second digit according to Benford's law, for $k=0,1,2,\ldots,9$ in each specified group of demographic data 2011 Census-Al.

Test the hypothesis

H_0: The data enjoys Benford's law.

H_1: The data does not enjoy Benford's law.

The appropriate test statistics is Pearson's chi-squared test:

$\chi^2 = \sum_{k=0}^{9} \frac{(e_k - o_k)^2}{o_k}$ with degrees of freedom df=9,

where e_k denotes the expected frequency according to Benford's law and o_k denotes the observed frequency, see Ramachandran and Tsokos (2009).

Calculate the observed value of the test statistics for each group separately.

Group 1

$$\chi^2 = \frac{3.3^2}{76.3} + \frac{6.1^2}{72.9} + \frac{0.1^2}{69.9} + \frac{3.2^2}{67.2} + \frac{8.8^2}{64.8} + \frac{1.6^2}{62.6} + \frac{13.5^2}{60.5} + \frac{13.3^2}{58.7} + \frac{0.9^2}{56.9} + \frac{15.2^2}{80.2} = 10.95$$

Group 2

$$\chi^2 = \frac{17.9^2}{74.1} + \frac{1.8^2}{70.8} + \frac{2.1^2}{67.9} + \frac{5.7^2}{65.3} + \frac{13^2}{63} + \frac{7.2^2}{60.8} + \frac{11.8^2}{58.8} + \frac{2^2}{57} + \frac{5.4^2}{55.4} + \frac{1.1^2}{77.9} = 11.48$$

Group 3

$$\chi^2 = \frac{20.2^2}{70.2} + \frac{20^2}{67} + \frac{18.7^2}{64.3} + \frac{3.8^2}{61.8} + \frac{12.4^2}{59.6} + \frac{2.4^2}{57.4} + \frac{11.7^2}{55.7} + \frac{13.9^2}{53.9} + \frac{5.6^2}{52.4} + \frac{4.7^2}{73.7} = 27.09$$

Group 4

$$\chi^2 = \frac{3^2}{10} + \frac{2.6^2}{9.6} + \frac{5.8^2}{9.2} + \frac{3.2^2}{8.8} + \frac{1.5^2}{8.5} + \frac{4.2^2}{8.2} + \frac{0^2}{8} + \frac{1.3^2}{7.7} + \frac{2.5^2}{7.5} + \frac{0.5^2}{10.5} = 10.68$$

Group 5

$$\chi^2 = \frac{5.7^2}{91.3} + \frac{0.3^2}{87.3} + \frac{16.3^2}{83.7} + \frac{4.4^2}{80.4} + \frac{21.5^2}{77.5} + \frac{9.9^2}{74.9} + \frac{2.5^2}{72.5} + \frac{12.2^2}{70.2} + \frac{7.8^2}{68.2} + \frac{22^2}{96} = 19.55$$

Find the critical value $\chi_c^2=\chi_\alpha^2(\text{df})=\chi_\alpha^2(9)$, where α denotes the significance level and df=10-1=9 denotes degrees of freedom:

$\chi_{0.10}^2(9)=14.684$, $\chi_{0.05}^2(9)=16.919$, $\chi_{0.025}^2(9)=19.023$, $\chi_{0.01}^2(9)=21.666$, $\chi_{0.005}^2(9)=23.589$,

see Ramachandran and Tsokos (2009), p 766.

Decision Rule for Group 1:

$\chi^2 = 10.95 < \chi_c^2 = 16.92$ $\longrightarrow$ accept the null hypothesis H_0 at the confidence level 95%. In other words, the data obeys Benford's law at the confidence level 95%.

Decision Rule for Group 2:

χ^2 =11.48<χ_c^2=16.919 $\longrightarrow$ accept the null hypothesis H_0 at the confidence level 95%.

Decision Rule for Group 3:

The observed value χ^2 =27.09>χ_c^2=23.589 $\longrightarrow$ reject the null hypothesis H_0 at the confidence level 99%. In other words, the data diverges from Benford's law at the confidence level 99.5%.

Decision Rule for group 4:

χ^2 =10.68<χ_c^2=16.919 $\longrightarrow$ accept the null hypothesis H_0 at the confidence level 95%.

Decision Rule for Group 5:

χ^2 =19.55>χ_c^2=16.919 $\longrightarrow$ reject the null hypothesis H_0 at the confidence level 95%. In other words, the data do not obey Benford's law at the confidence level 95%.

Consider the groups 1,2,3,4,5 together. Test the hypothesis whether the data set obeys (or contradicts) Benford's law.

Group 1 U Group 2 U Group 3 U Group 4 U Group 5

The observed value of the test statistics is:

$$\chi^2 = \frac{2.9^2}{321.9} + \frac{21.4^2}{307.6} + \frac{43^2}{295} + \frac{2.5^2}{283.5} + \frac{13.6^2}{273.4} + \frac{10.9^2}{263.9} + \frac{12.5^2}{255.5} + \frac{13.5^2}{247.5} + \frac{4.6^2}{240.4} + \frac{40.3^2}{338.3} = 15.15$$

The critical value of the test statistics is χ_c^2=$\chi_{0.10}^2$(9) = 14.68

Decision Rule

χ^2 =15.15>χ_c^2=14.684 $\longrightarrow$ reject the null hypothesis H_0 at the confidence level 90%. In other words, the demographic data 2011 Census-Al contradicts Benford's law at the confidence level 90%.

4. Conclusion

In the present study we develop a statistical analysis based on Benford's law for the second digit of demographic data 2011 Census-Albania. We consider five different groups of the data. The source of official data is The Albanian Institute of Statistics (INSTAT). The main objective of the study is to contribute to the debate whether the demographic data 2011 Census-Al obeys Benford's law. Benford's law governs the asymptotic probability distribution of many (but not all) real-life sources of data, which satisfy the assumptions 1,2,3 and 4 presented in section 1 of the study. Hill (1995) provided a measure-theoretical proof that data selected (obtained) from a random mix (mixture) of different random variables will ultimately converge to Benford's law, even though the individual probability distributions may not closely follow the Benford's law. If the random variables $X_1, X_2, \dots X_n$, are selected at random and the random samples are obtained from each of these random variables $X_1, X_2, \dots X_n$, then the combined random sample will converge to Benford's law.

Attempts at intuitive explanation of Benford's law have centered on ideas of scale invariance and base invariance.

The scale invariance argument says that any universal law of nature (especially Benford's law) should not depend upon the units of measurement. For example, if we were to convert the land areas of the world states from km^2 to $miles^2$, then the rescaling data should obey Benford's law. Similarly, any universal law of nature (like Benford's law) should not depend on the number bases 10 or 2 or 3 or…

Benford's law has intrigued scientists for over a century. Many scientists have come to the conclusion that Benford's law is a tantalizing and mysterious law of nature, for which the true explanation lies with the God (Fewster, 2009). The underlying reason why Benford's law occurs is, however, elusive. We believe that Benford's law is an universal law of nature. The probability of violating Benford's law in the known part of universe is relatively small (Raimi, 1976).

The populations of 235 countries and regions of the world in 2013 (using CIA world factbook) obey Benford's law. The populations of the cities of USA in 2013 obey Benford's law.

Some results of the present study include:

- The mean of second digit for all five groups of the data obtained (selected) from 2011 Census-Al are in very good accord with Benford's law (at the confidence level 95%).

Decision Rule for Group 1:

The data obeys Benford's law at the confidence level 95%.

Decision Rule for Group 2:

The data obeys Benford's law at the confidence level 95%.

Decision Rule for Group 3:

The data contradicts Benford's law at the confidence level 99.5%.

Decision Rule for Group 4 :

The data obeys Benford's law at the confidence level 95%.

Decision Rule for Group 5:

The data contradicts Benford's law at the confidence level 95%.

Decision Rule for all Groups 1,2,3,4,5 together:

The demographic data 2011 Census-Al contradicts Benford's law at the confidence level 90%. Therefore, these 2827 official data obtained from 2011 Census-Al are suspectable for manipulation.

References

2011 Census-Al, published by INSTAT, Tirana, Albania

Diaconis, P. and Freedman, D. (1979). On rounding percentages, J. Amer. Stat. Assoc. , 359-364.

Durtschi, C. , Hillson, W. and Pacini, C. (2004). The effective use of Benford's law to assist in detecting fraud in accounting data, Journal of Forensic Accounting 1524-5586/Vol. V, pp 17-34.

Fewster, R. M. (2009). Simple explanation of Benford's law, The American Statistican, February, Vol. 63, No. 1, pp 26-32.

Field, A. (2009). Discovering Statistics Using SPSS (3rd.edition), London: Sage.

Hill, T. P. (1995). A statistical derivation of the significant-digit law, Statistical Science, Vol. 10, No. 4, pp 354-363.

Hill, T. P. (1995). The significant-digit phenomenon, The American Mathematical Monthly, Vol. 102, No. 4, pp 322-327.

Hill, T. P. (1998). The first-digit phenomenon, The American Scientist, July-August, 86:358.

Janvresse, E. and Th. De la Rue (2004). From uniform distribution to Benford's law, Journal of Applied Probability, 41, 1203-1210.

Nigrini, M. J. (1996). Taxpayer compliance application of Benford's law. Journal of the American Taxation Association. 18 (1) : 72-92.

Nigrini, M. J. and Wood, W (1995). Assessing the integrity of tabulated demographic data, Preprint, University of Cincinnati and Saint Mary's University.

Raimi, R. K. (1976). The first digit problem. The American Mathematical Monthly, 83, 521-538.

Ramachandran, K. M. and Tsokos, C. P. (2009). Mathematical Statistics with Applications, Academic Press, Elsevier, Boston, USA.

Tao, Terence (2009). Benford's law, Zipf's law, and the Pareto distribution (http://terrytao.wordpress.com/2009/07/03)