Journal of Mathematical Sciences: Advances and Applications
Volume 52, 2018, Pages 69-81
Available at http://scientificadvances.co.in
DOI: http://dx.doi.org/10.18642/jmsaa_7100121981

BENFORD'S LAW AND WILCOXON TEST

ZORAN JASAK

NLB Banka d. d.
Sarajevo
Bosnia & Herzegovina
e-mail: zoran.jasak@nlb.ba

Abstract

Benford's law gives expected patterns for frequencies of digits in numerical data sets according to their position in numbers. Benford's law is known as a 'first digit law', 'digit analysis' or 'Benford-Newcomb phenomenon'. Based on Benford's law tests for leading and non-leading and appropriate groups of digits have been derived. Practical problem is how to test conformity to this law. In this paper, variant of Wilcoxon test, which can be used both for frequencies and sums of significands, is proposed.

1. Introduction

In 1881, Simon Newcomb in the article "Note on the Frequency of use of different digits in natural numbers" [1], stated that "the first significant digit is oftener 1 than any other digit, and the frequency diminishes up to 9". This idea came from Newcomb's observation of the use of library logarithm tables where he noticed that the first pages of these tables were actually dirtier than the last ones. From this he concluded that people are more likely to use numbers starting with

2010 Mathematics Subject Classification: 62E10, 62Q99.
Keywords and phrases: Benford's law, Wilcoxon test.
Received June 30, 2018; Revised August 6, 2018

smaller than with larger digits. This phenomenon was re-discovered by Frank Albert Benford, who in 1938 published a paper entitled "The Law of Anomalous Numbers" [2]. He investigated 20 different sets of natural numbers involving over 20,000 samples, from newspaper pages to home addresses. He gave mathematical formulation:

$$P[D = d] = \log_{10}(1 + 1 / d),$$

where d is leading digit, $d \in \{1, 2, \dots, 9\}$. This theory was then tested on a large number of different statistical data sets and proved to hold true for most of them, with similar laws derived for other order digits. It was also observed that the more mixed the data, the closer the distribution of digits was to the logarithmic one. This is now commonly called Benford's Law. Other mathematicians (S. Goudsmit, W. Furry, H. Hurwitz, R. Pinkham, R. Raimi, and especially T. Hill) gave the theoretical basis for this law. It's possible on website http://www.benfordonline.net to find numerous texts both on theoretical and practical aspects of Benford's law.

Since we have theoretical model next problem is to test do our data for having Benford property. As it's mentioned in [14], methods of testing are highly dependent on the area of application; different protocols are used when analyzing financial results, voting registers, or network intrusion records. In the same text wide list of tests for Benford's property is given: Chi-square, Mean Absolute Deviation, Mantissa Arch Test, Distorsion Factor, Kolmogoroff-Smirnov test, Freedman-Watson test, Chebyshev Distance test, Euclidean Distance test, Judge-Schechter Mean Deviation test, Joenssens J_p^2 test, Hotelling T^2 test. Mark Nigrini [13] (pp. 109-129) describes some other tests (z-test, tests bases on the logarithmic basis of Benford's law). In article by Lesperance et al. [10], likelihood ratio, Pearson's Chi-square tests for benford's law are presented and tests based on Cramer-Von Mises statistics as well, in context of simultaneous confidence intervals. Variant of Hosmer-Lemeshow test is proposed by author in [15].

2. Wilcoxon Test

2.1. Introduction

The Wilcoxon signed-rank test is a non-parametric statistical test used to compare two related samples, matched samples, or repeated measurements on a single sample to assess whether their population mean ranks differ (i.e., it is a paired difference test) [4]. It can be used as an alternative to the paired Student's t-test, t-test for matched pairs, or the t-test for dependent samples when the population cannot be assumed to be normally distributed. This test can be used to determine whether two samples were selected from populations having the same distribution. The test is named for Frank Wilcoxon (1892-1965) who, in a single paper, proposed rank-sum test for two independent samples too (Wilcoxon, 1945) [3].

If we have two samples both in size of N elements, we make pairs $\{x_{1,i}, x_{2,i}\}$, $i = 1, \dots, N$, to compare them in some way [4]. Null hypothesis H_0 is that differences between the pairs follow a symmetric distribution around zero and alternative H_1 hypothesis is that differences don't follow a symmetric distribution.

Testing procedure for two samples is next:

• Calculation of differences $|x_{2,i} - x_{1,i}|$ and values $\operatorname{sgn}(x_{2,i} - x_{1,i})$, $i = 1, \dots, N$, where sgn is the sign function.

• Exclusion of pairs with $|x_{2,i} - x_{1,i}| = 0$; remains N_r pairs as a reduced sample size.

• Ranking absolute values of differences, starting with the smallest as 1. Ties receive a rank equal to the average of the ranks they span. Let R_i denotes the rank of absolute difference $|x_{2,i} - x_{1,i}|$.

• Calculation of sums of positive and negative ranks:

$$S = \sum_{i01}^{N_r} [\operatorname{sgn}(x_{2,i} - x_{1,i}) \cdot R_i].$$

Let *S_pos* and *S_neg* are sums of positive and negative ranks, respectively. Wilcoxon *W* statistics is minimum of those two values:

$$W = \min\{S\_pos, S\_neg\}.$$

There are two critical values for appropriate significance level, tabulated in appropriate statistical tables [6]. If *W* is less than lower or higher than upper critical value null hypothesis is rejected. In this way we have two-tailed test.

If we want to test frequencies of leading digits in two samples by this test, first step is to sort both samples by leading digits, to form *N* pairs $\{x_{1,i}, x_{2,i}\}$, $i = 1, \dots, N$.

2.2. Testing of frequencies

Traditional approach to test Benford's property is to divide sample in 9 classes and count frequencies in each of them. Another approach is proposed here [5]. Basic concept is to have two samples of equal sizes, to arrange them in ascending order of leading digits or groups of digits, to calculate differences between digits in pairs and to calculate appropriate statistics. If one sample is made as sample with theoretical frequencies we test goodness-of-fit. If we use any other sample in equal size we compare two samples according to leading digits. In sequel, procedure for first digits is presented.

The first step is to have frequencies $m_{d0}, n_{d0}, d = 1, 9,$ of first digits in first and second sample, respectively. Zero as the second index denotes initial frequencies. Main condition is:

$$\sum_{d=1}^{9} m_{d0} = \sum_{d=1}^{9} n_{d0}.$$

If we have m_{10} and n_{10} numbers beginning by 1, then we have $k = \min\{m_{10}, n_{10}\}$ numbers beginning with the same digit, namely, k ties. In this case difference between leading digits in each of those k pairs is $r = 0$.

- If $m_{10} > n_{10}$, then $k = \min\{m_{10}, n_{10}\} = n_{10}$ and we have $m_{11} = m_{10} - k$ and $n_{11} = n_{10} - k = 0$.

- If $m_{10} < n_{10}$, then $k = \min\{m_{10}, n_{10}\} = m_{10}$ and we have $m_{11} = m_{10} - k = 0$ and $n_{11} = n_{10} - k$.

- If $m_{10} = n_{10}$, then $k = m_{10} = n_{10}$ and we proceed with next pair.

Suppose for the moment that is $m_{10} > n_{10}$; next step is compare m_{11} and n_{20} to calculate $k = \min\{m_{11}, n_{20}\}$. In this case we have k pairs with different leading digits, 1 and any other; if the second digit is 2, then difference between digits is $r = 2 - 1$. Next step is to calculate:

$$m_{12} = m_{11} - k, \quad n_{21} = n_{20} - k.$$

This procedure continues until we exceed all available pairs.

3. Numerical Examples

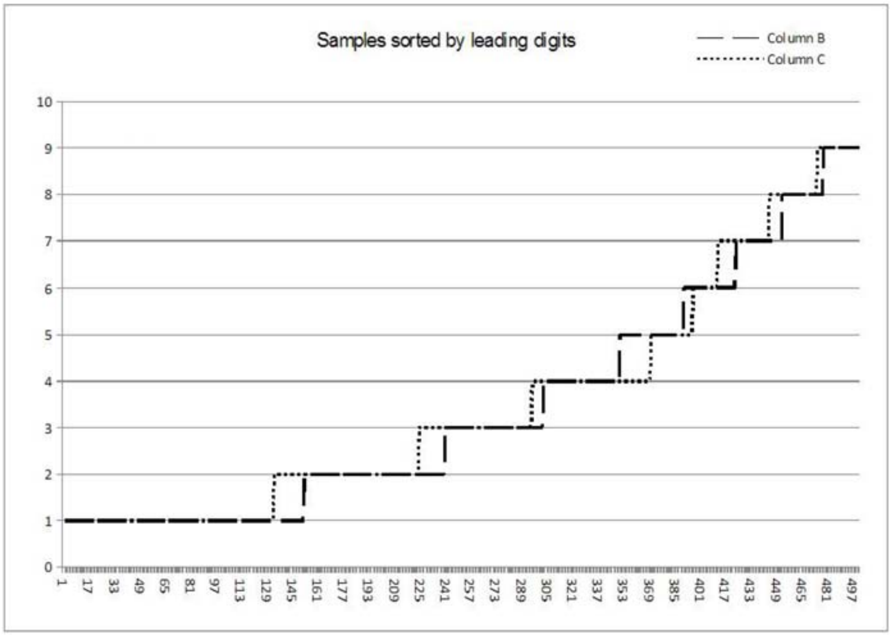
Whole procedure is illustrated by two samples with frequencies of leading digits presented in Table 1. All calculations are conducted by use of Excel, on $\alpha = 0.05$ significance level.

Assume in first case we have two samples in size of 500 elements each.

Table 1.

Digits	1	2	3	4	5	6	7	8	9
Theoretical frequencies	151	88	62	48	40	33	29	26	23
Sample frequencies	132	91	71	75	26	16	32	30	27

This corresponds to case when both samples are arranged in ascending order according to leading digit, forming n pairs, as we can see in Graph 1.



Graph 1. Leading digits of two samples. Dashed line presents theoretical and dotted line presents observed frequencies.

In this diagram x-axis represents ordinal number of every pair after sorting. Difference in leading digits is visible by vertical gap for every pair. Ties are visible as mixture of dashed and dotted line. Calculation is simulated in Table 2.

Table 2. Illustration of calculation procedure. Explanations are in text

i	j	*Mi_a*	*Nj_b*	*Mi_aP*	*Nj_bP*	k	*Mi_a*1	*Nj_b*1	r	r_n
1	1	151	132	151	132	132	19	0	0	0
1	2		91	19	91	19	0	72	1	19
2	2	88		88	72	72	16	0	0	0
2	3		71	16	71	16	0	55	1	16
3	3	62		62	55	55	7	0	0	0
3	4		75	7	75	7	0	68	1	7
4	4	48		48	68	48	0	20	0	0
5	4	40		40	20	20	20	0	– 1	– 20
5	5		26	20	26	20	0	6	0	0
6	5	33		33	6	6	27	0	– 1	– 6
6	6		16	27	16	16	11	0	0	0
6	7		32	11	32	11	0	21	1	11
7	7	29		29	21	21	8	0	0	0
7	8		30	8	30	8	0	22	1	8
8	8	26		26	22	22	4	0	0	0
8	9		27	4	27	4	0	23	1	4
9	9	23		23	23	23	0	0	0	0

In columns i and j are leading digits in first and second sample, respectively. At the same time, they are indexes for m_{ik} and n_{jk}. Frequencies *Mi_a* and *Nj_b* are theoretical and observed frequencies respectively, taken in some step of calculation. In columns *Mi_aP* and *Nj_bP* are m_{ik} and n_{jk}, depending on values from previous step.

In first step we have $M1\_a = m_{10} = 151$ and $N1\_b = n_{10} = 132$. From formal reasons, in first step is:

$$M1\_aP = M1\_a = 151, \quad N1\_bP = N1\_b = 132.$$

Column k denotes result of calculation:

$$k = \min\{M1\_aP,\ N1\_bP\} = \min\{m_{10}, n_{10}\} = \min\{151, 132\} = 132.$$

By this subsequence in length of k is defined. Columns Mi_a1 and Nj_b1 denote numbers we have after subtraction of k from m_{ik} and n_{jl} so we have:

$$M1\_a1 = m_{11} = M1\_aP - k = 151 - 132 = 19;$$

$$N1\_b1 = n_{11} = N1\_bP - k = 132 - 132 = 0.$$

Column r denotes difference between digits, namely, between indexes i and j. For the first row this difference is $j - i = 0$. Column r_n denotes product $k \cdot r$, sum of differences in current subsequence.

Total number of (positive and negative) differences is sum of absolute values in column r_n, which is basis for Wilcoxon test. Actually, we have k differences $r \neq 0$. Formally, sum of differences is:

$$S = \sum_i k_i \cdot |r_i|, \quad r_i \neq 0.$$

The same result we can get by use of cumulative frequencies, what is presented in Table 3.

Table 3. Calculation by use of cumulative sums

Digits	Fr1	Fr2	Sum1	Sum2	Diff	R_pos	R_neg
1	151	132	151	132	19	7	
2	88	91	239	223	16	6	
3	62	71	301	294	7	3	
4	48	75	349	369	– 20		8
5	40	26	389	395	– 6		2
6	33	16	422	411	11	5	
7	29	32	451	443	8	4	
8	26	30	477	473	4	1	
9	23	27	500	500	0		

In columns Fr1 and Fr2 are frequencies of digits in both samples. In this case frequencies Fr1 are calculated as expected theoretical values. In columns Sum1 and Sum2 are cumulative sums of frequencies for the first and second sample, respectively. In column Diff are differences of cumulative frequencies for every digit (by row). Those differences are identical to non-null values in column *r_n* in Table 3. In columns *R_pos* and *R_neg* are ranks for positive and negative differences, respectively. Critical values for two-tailed test for $n = 8$ and 5% significance level are 49 and 87. In this calculation we have $W = \min\{26, 10\} = 10,$ below low value (49) what implies that we need to reject null hypothesis.

Testing procedure is demonstrated on other sample in size of 11,486 elements, generated by program El-Muhsi, macro in Excel. Minimal sample value is 10, maximal value is 176,932.50, sample average is 3,606.00, standard deviation is 7,793.29, total sum of all values is 41,418,526.12. Observed and expected frequencies of leading digits are in Table 4.

Table 4. Observed and expected frequencies

Digits	1	2	3	4	5	6	7	8	9
Theoretical frequencies	3,458	2,023	1,435	1,113	909	769	666	588	526
Observed frequencies	4,047	1,747	1,222	997	921	721	623	639	569

Test procedure by use of Wilcoxon test of equivalent pairs for leading digits is in Table 5.

Table 5. Wilcoxon test for frequencies, leading digits

i	*j*	*Mi_a*	*Nj_b*	*Mi_aP*	*Nj_bP*	***k***	*Mi_a1*	*Nj_b1*	*r*	*r_n*	*Rank*	*R_Pos*	*R_Neg*
1	1	3,458	4,047	3,458	4,047	**3,458**	0	589	0	0			
2	1	2,023		2,023	589	**589**	1,434	0	– 1	– 589	8		8
2	2		1,747	1,434	1,747	**1,434**	0	313	0	0			
3	2	1,435		1,435	313	**313**	1,122	0	– 1	– 313	7		7
3	3		1,222	1,222	1,222	**1,222**	0	100	0	0			
4	3	1,113		1,113	100	**100**	1,013	0	– 1	– 100	6		6
4	4		997	1,013	997	**997**	16	0	0	0			
4	5		921	16	921	**16**	0	905	1	16	2	2	
5	5	909		909	905	**905**	4	0	0	0			
5	6		721	4	721	**4**	0	717	1	4	1	1	
6	6	769		769	717	**717**	52	0	0	0			
6	7		623	52	623	**52**	0	571	1	52	4	4	
7	7	666		666	571	**571**	95	0	0	0			
7	8		639	95	639	**95**	0	544	1	95	5	5	
8	8	588		588	544	**544**	44	0	0	0			
8	9		569	44	569	**44**	0	525	1	44	3	3	
9	9	525		525	525	**525**	0	0	0	0			

Frequencies *Mi_a* and *Nj_b* are observed and theoretical frequencies, respectively. Short calculation is given in Table 6.

Table 6.

Digits	Fr1	Fr2	Sum1	Sum2	Diff	R_pos	R_neg
1	3,458	4,047	3,458	4,047	– 589		8
2	2,023	1,747	5,481	5,794	– 313		7
3	1,435	1,222	6,916	7,016	– 100		6
4	1,113	997	8,029	8,013	16	2	
5	909	921	8,938	8,934	4	1	
6	769	721	9,707	9,655	52	4	
7	666	623	10,373	10,278	95	5	
8	588	639	10,961	10,917	44	3	
9	536	569	11,486	11,486	0		

Smaller sum of positive ranks (15) and of negative ranks (21) is 15 for positive ranks. Lower and upper critical table values for reduced sample size of $N_{r'} = 8$ on $\alpha = 0.05$ level of significance, for two-tailed test, are 49 and 87 [6]. Probability level for those values is 0.0249. It means that if value of sample statistic is smaller than 49 or bigger than 87 we need to reject hypothesis. In this case, we conclude that frequencies of leading digits in our sample are not equally distributed.

4. Conclusion

In this text variant of Wilcoxon test for frequencies of leading digits is presented. The same concept is possible to use for second, third, ... digit or any group of digits.

Main characteristic of this approach is that we don't have strong separation in 9 classes. Differences bigger than 1 mean that some frequencies are unusually small or big, what can imply rejection of null hypothesis.

Sum invariance testing is possible by the same procedure in way that we use ranks of differences between theoretical and sample sums of significands. According to sum invariance property of Benford's law, sums of significands are equal (in expectation). Theoretical sums of significands for leading digits can be calculated as product of average significand for every position and sample size. Theoretical sums of significands for second position are calculated as 9/10 of theoretical sum of significands for first position.

References

[1] Simon Newcomb, Note on the frequency of use of different digits in natural numbers, American Journal of Mathematics 4(1) (1881), 39-40.

DOI: http://dx.doi.org/10.2307/2369148

[2] Frank A. Benford, The law of anomalous numbers, Proceedings of the American Philosophical Society 78(4) (1938), 551-572.

[3] Frank Wilcoxon, Individual comparisons by ranking methods, Biometrics Bulletin 1(6) (1945), 80-83.

DOI: http://dx.doi.org/10.2307/3001968

[4] Wikipedia, Free Encyclopedia

http://en.wikipedia.org/wiki/Wilcoxon_signed-rank_test

[5] Zoran Jasak, Sum Invariance Testing and Some New Properties of Benford's Law, Doctorial Dissertation, University of Tuzla, Bosnia and Herzegovina, 2017.

[6] http://plantsys.elte.hu/oktatas/Biometria/tablazatok/Wilcoxon_table_ketmintas_probahoz.pdf

[7] http://www.statstutor.ac.uk/resources/uploaded/wilcoxonsignedranktest.pdf

[8] Jeff Baggett, Critical Values for the Wilcoxon Rank-Sum Test.

http://www.uwlax.edu/faculty/baggett/Math_145

[9] Daniel P. Pike, Testing for the Benford property, School of Mathematical Sciences, Rochester Institute of Technology.

[10] M. Lesperance, W. J. Reed, M. A. Stephens, C. Tsao and B. Wilton, Assessing conformance with Benford's law: Goodness-of-fit tests and simultaneous confidence intervals, PLOS ONE, March 28 (2016).

DOI: https://doi.org/10.1371/journal.pone.0151235

[11] Mark Linville, The problem of false negative results in the use of digit analysis, The Journal of Applied Business Research 24(1) (2008), 17-25.

DOI: https://doi.org/10.19030/jabr.v24i1.1363

[12] P. Kruger and S. Yadavalli, The power of one: Benford's law, South African Journal of Industrial Engineering 28(2) (2017), 1-13.

DOI: http://dx.doi.org/10.7166/28-2-1753

[13] Mark J. Nigrini, Forencis Analytic: Methods and Techniques for Forensic Accounting Investigations, John Wiley & Sons, Inc., 2011.

[14] Mikolaj Morzy, Tomasz Kajdanowicz and Boleslaw K. Szymanski, Benford's distribution in complex networks, Scientific Reports 6 Article 34917, 2016.

DOI: http://dx.doi.org/10.1038/srep34917

[15] Zoran Jasak, Benford's law and Hosmer-Lemeshow test, Journal of Mathematical Sciences: Advances and Applications 41 (2016), 57-73.

DOI: http://dx.doi.org/10.18642/jmsaa_7100121718

■