benford.pl

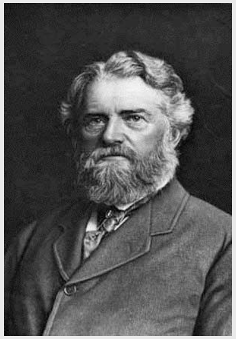
Application of the first digit law in credibility evaluation of the financial accounting data based on particular cases

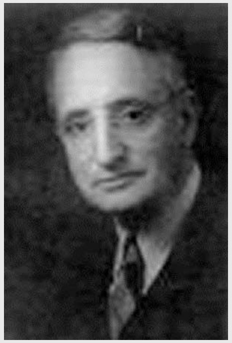
Marzena Farbaniec
Tadeusz Grabiński
Bartłomiej Zabłocki
Wacław Zając

Andrzej Frycz Modrzewski Cracow University

Discovery

Simon Newcomb: „Note on the Frequency of Use of the Different Digits in Natural Numbers", The American Journal of Mathematics 4, pp 39-40, 1881





Frank Benford: "The Law of Anomalous Numbers", Proceedings of the American Philosophical Society 78, pp 551-72, 1938

Both independently noticed that first pages of logarithms tables are much more worn and dirty than later pages. Their conclusion was that numbers beginning with 1 were looked up more often than those beginnig with 2, ..., 9.

Benford's Law:

- investigates the frequency distribution of digits in a given data set.

- says that the probability of the first digit being a "d" is:

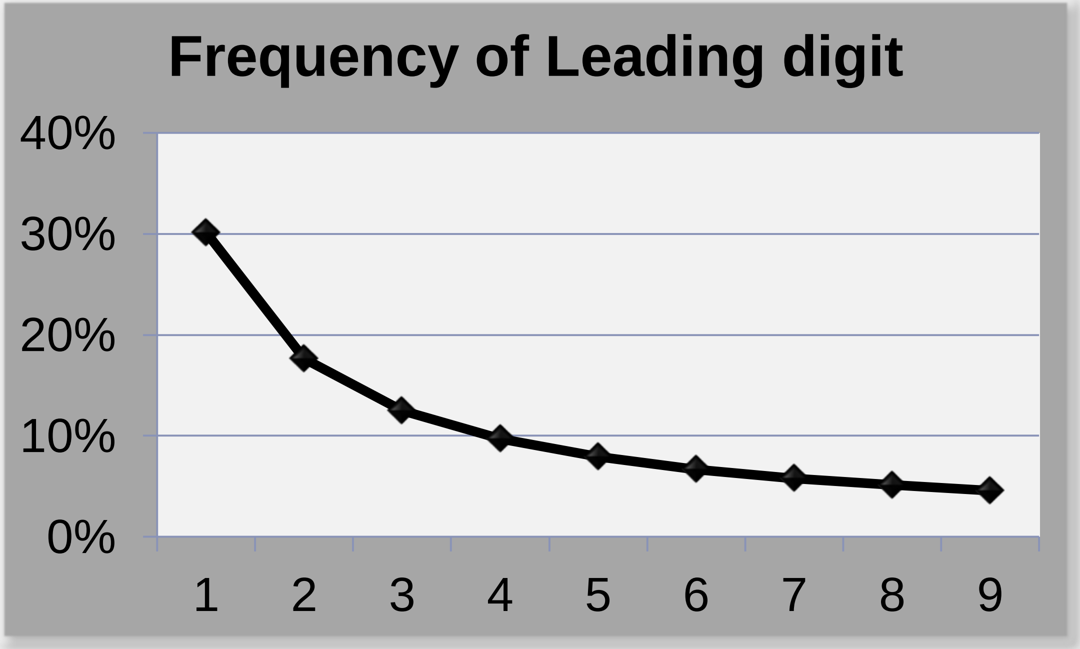
$$P(d) = \log_{10}\left(1 + \frac{1}{d}\right)$$

where: $(d = 1, 2, 3, 4, 5, 6, 7, 8, 9)$

$$\sum P(d) = 1$$

Leading digits frequencies

Leading digit	Frequency
1	30.10%
2	17.61%
3	12.49%
4	9.69%
5	7.92%
6	6.69%
7	5.80%
8	5.12%
9	4.58%
Total	100.00%



Other tests

- Probability *P{D}* that the number starts with the sequence *{D}:*

  F1, F2, F3: $$P_{\{D\}} = log_B(1 + \frac{1}{D})$$

- Probability that the second significant digit of a number in the decimal system is *k:*

  D2: $$P_k^{(2)} = \sum_{i=1}^{9} log\left(1 + \frac{1}{10i + k}\right), \qquad k \epsilon (0, 1, 2, 3, 4, 5, 6, 7, 8, 9)$$

- Probability that the third significant digit of a number in the decimal system is *k:*

  D3: $$P_k^{(3)} = \sum_{i=1}^{9} log\left(1 + \frac{1}{100i + k}\right), \qquad k \epsilon (0, 1, 2, 3, 4, 5, 6, 7, 8, 9)$$

- Probability that the last digit of a number in the decimal system is *k:*

  L1: $$P_k = \frac{1}{10}, \qquad k \epsilon (0, 1, 2, 3, 4, 5, 6, 7, 8, 9)$$

Conditions of Benford Sets

- All data must be recorded in the same physical unit;
- The data set may not contain any inherent minima or maxima;
- The data set may not include any assigned numbers, which do not arise from any “natural” process (e.g. bank accounts, telephone numbers);
- A data set should tend to have more small than large numbers, which also accords with the “natural” development process.

(Nigrini, 2000)

Characteristics of Benford Sets

- *Scale invariance:* multiplying all values in the Benford Set by any constant yields another Benford Set *(Pinkham 1961);*
- *Base invariance:* Benford's Law applies in all numerical systems not only to numerical systems with a basis of 10. *(Hill 1995b);*
- *Invariance in mathematical operations:* raising to powers, multiplication and division *(Hamming 1970; Schatte 1988; Boyle 1994)* as well as addition and subtraction *(Hamming 1970; Schatte 1988)*, of Benford Sets with one another lead to a new Benford Set.

Fraud in Arizona, 1993 (1)

Date	Amount
09.09.1992	1 927,48
	27 902,31
14.09.1992	86 241,90
	72 117,46
	81 321,75
	97 473,96
19.09.1992	93 249,11
	89 658,17
	87 776,89
	92 105,83
	79 949,16
	87 602,93
	96 879,27
	91 806,47
	84 991,67
	90 831,83
	93 766,67
	88 338,72
	94 639,49
	83 709,28
	96 412,21
	88 432,86
	71 552,16
	1 878 687,58

Wayn J. Nelson was found guilty trying to defraud almost $2 000 000.

• He started with smaller amounts and started to increase it;
• Most of the amounts were just below $100 000;

• The digit patterns were almost opposit to Benford's Law distribution – over 90% of the values had 9, 8, 7 as first digit;

• Non of the check amounts were duplicated;

• There were no round numbers.

Fraud in Arizona, 1993 (2)

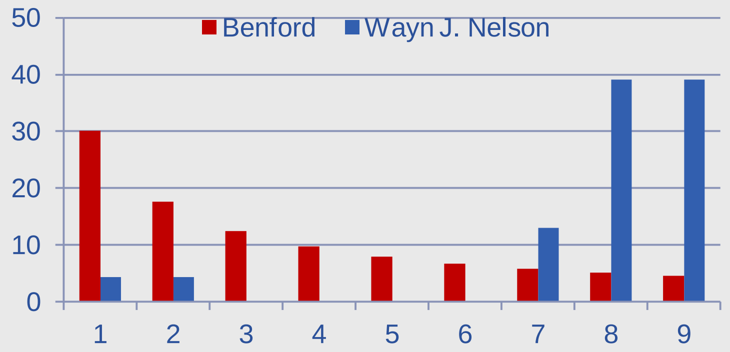
First two digits used twice:

- 87
- 88
- 93
- 96

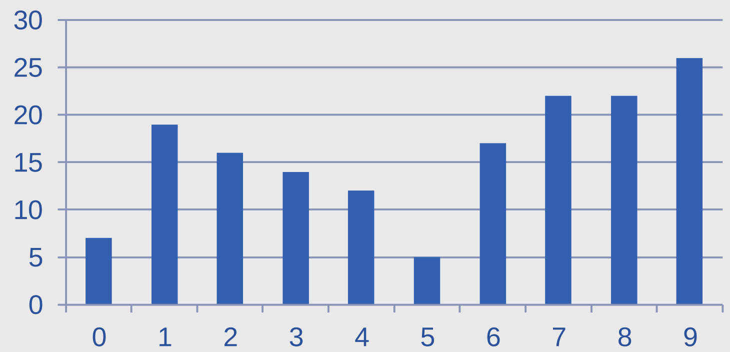
Last two digits used twice:

- 16
- 67
- 83

First digits (%):



Digits used by Nelson:



Measures of fit (1)

benford.pl

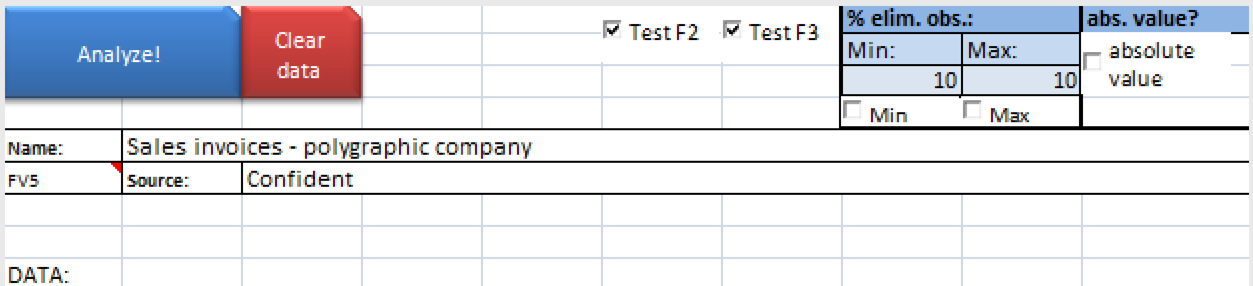
Test	Equation
Chi-square	$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} = \sum_{i=1}^{k} \frac{(n_i - \hat{n}_i)^2}{\hat{n}_i}$
Kolmogorov-Smirnov 1 (KS1)	$KS1 = D\sqrt{\frac{n^2}{2n}} \qquad D = \max_i \left\| f_i - \hat{f}_i \right\| \qquad (i = 1,\ldots,k)$
Kolmogorov-Smirnov 2 (KS2)	$KS2 = D\sqrt{n} \qquad D = \max_i \left\| f_i - \hat{f}_i \right\| \qquad (i = 1,\ldots,k)$
Kolmogorov-Smirnov 3 (KS3) Kuiper's test	$KS3 = V_N * [\sqrt{N} + 0{,}155 + 0{,}24N^{-1/2}] \qquad where \qquad (i = 1,\ldots,k)$ $V_N = D_N^+ + D_N^- \quad D_N^+ = \sup_i \left[f_i - \hat{f}_i \right] \quad D_N^- = \sup_i \left[\hat{f}_i - f_i \right] \quad N = \frac{n^2}{2n}$
Z-test	$z_i = \frac{p_i - \hat{p}_i}{\sqrt{\hat{p}_i(1 - \hat{p}_i)/n}} \qquad (i = 1,\ldots,k)$

Measures of fit (2)

Measures	Equation
M_1	$M_1 = \frac{100}{k}\sum_{i=1}^{k}\left\lvert\frac{c_i - \hat{c}_i}{\hat{c}_i}\right\rvert$
M_2	$M_2 = \frac{1}{k}\sqrt{\sum_{i=1}^{k}(c_i - \hat{c}_i)^2}$
M_3	$M_3 = \sqrt{\frac{\sum_{i=1}^{k}(c_i - \hat{c}_i)^2}{k}}$
M_4	$M_4 = \frac{100\sqrt{\sum_{i=1}^{k}(c_i - \hat{c}_i)^2}}{\sqrt{\sum_{i=1}^{k}\hat{c}_i^2}}$
M_5	$M_5 = \frac{100\sum_{i=1}^{k}\lvert n_i - \hat{n}_i\rvert}{n}$

Analyzing process

Benford Excel Analyzer tool:



- Data preparation;
- Option customization;
- Analyzing;
- Inference from the results.

FV3 dataset analysis result (1)

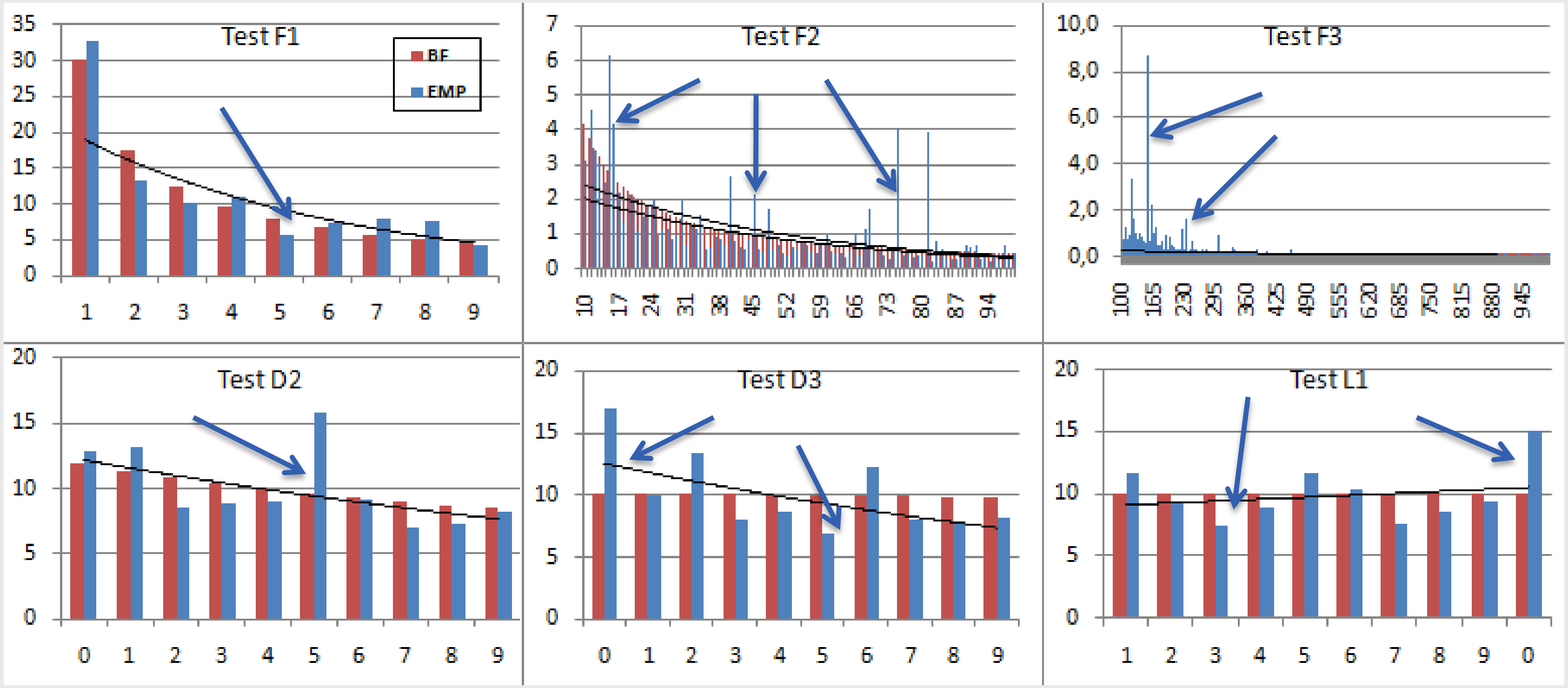
			α=0,05	**1,36**	**1,36**	**1,747**
			α=0,01	**1,63**	**1,63**	**2,001**
	chi	chi-0,05	p(chi)	KS1	KS2	KS3
F1	**341,8**	**15,5**	**0,000**	**3,13**	**4,43**	**4,67**
F2	**4703,8**	**112**	**0,000**	**3,66**	**5,18**	**6,04**
F3	**13830,7**	**969,9**	**0,000**	**15,98**	**22,60**	**15,98**
D2	**433,7**	**16,9**	**0,000**	**2,42**	**3,42**	**3,76**
D3	**320,7**	**16,9**	**0,000**	**4,24**	**6,00**	**4,24**
L1	**357,8**	**16,9**	**0,000**	**3,12**	**4,41**	**4,08**

FV3 dataset analysis result (2)

Nazwa:	**Sales invoices - polygraphic company**												
FV3	Source:	**Confident**											
Test z	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**max**	α=0,05	α=0,01
F1	**-**	**4,76**	**9,55**	**6,1**	**3,54**	**7,41**	**2,67**	**7,44**	**10,1**	**0,97**	**10,1**	8	**8**
D2	**2,49**	**4,81**	**6,64**	**4,4**	**2,74**	**18,01**	**0,74**	**5,96**	**4,42**	**0,9**	**18,01**	8	**7**
D3	**13,57**	**0,28**	**6,57**	**4,14**	**2,8**	**6,43**	**4,55**	**3,96**	**3,95**	**3,32**	**13,57**	9	**9**
L1	**4,54**	**1,89**	**7,62**	**3,09**	**4,88**	**0,98**	**6,77**	**3,94**	**1,81**	**14,71**	**14,71**	7	**7**
F2-10	**4,47**	**3,66**	**0,22**	**1,01**	**2,52**	**17,56**	**8,13**	**1,56**	**5,44**	**0,47**	**17,56**	62	**46**
F2-20	**0,4**	**5,83**	**2,62**	**0,51**	**1,38**	**4,95**	**3,24**	**3,07**	**4,92**	**3,77**	**5,83**	Sum:	
F2-30	**3,88**	**5,19**	**0,71**	**1,34**	**2,06**	**5,35**	**4,41**	**1,93**	**2,04**	**3,5**	**5,35**	94	**76**
F2-40	**13,04**	**2,23**	**3,55**	**3,97**	**1,1**	**10,43**	**3,28**	**3,51**	**7,36**	**2,74**	**13,04**		
F2-50	**1,72**	**3,74**	**4,4**	**2,09**	**2,22**	**0,65**	**1,05**	**2,54**	**3,11**	**0,83**	**4,4**		
F2-60	**2,73**	**1,98**	**2,72**	**2,77**	**3,96**	**1,73**	**3,99**	**2,57**	**5,42**	**12,2**	**12,2**		
F2-70	**1,43**	**1,19**	**2,16**	**3,75**	**0,28**	**39,86**	**1,86**	**1,63**	**2,82**	**2,12**	**39,86**		
F2-80	**1,9**	**40,15**	**3,85**	**2,99**	**0,48**	**2,23**	**2,99**	**2,77**	**2,05**	**2,18**	**40,15**		
F2-90	**1,93**	**2,34**	**2,34**	**0,92**	**3,44**	**0,29**	**2,84**	**2,97**	**1,69**	**0,12**	**3,44**	1,96	**2,58**
max	**13,6**	**40,1**	**9,5**	**6,1**	**4,9**	**39,9**	**8,1**	**7,4**	**10,1**	**14,7**	**40,1**		

FV3 dataset analysis result (3)

benford.pl



FV5 dataset analysis result (1)

benford.pl

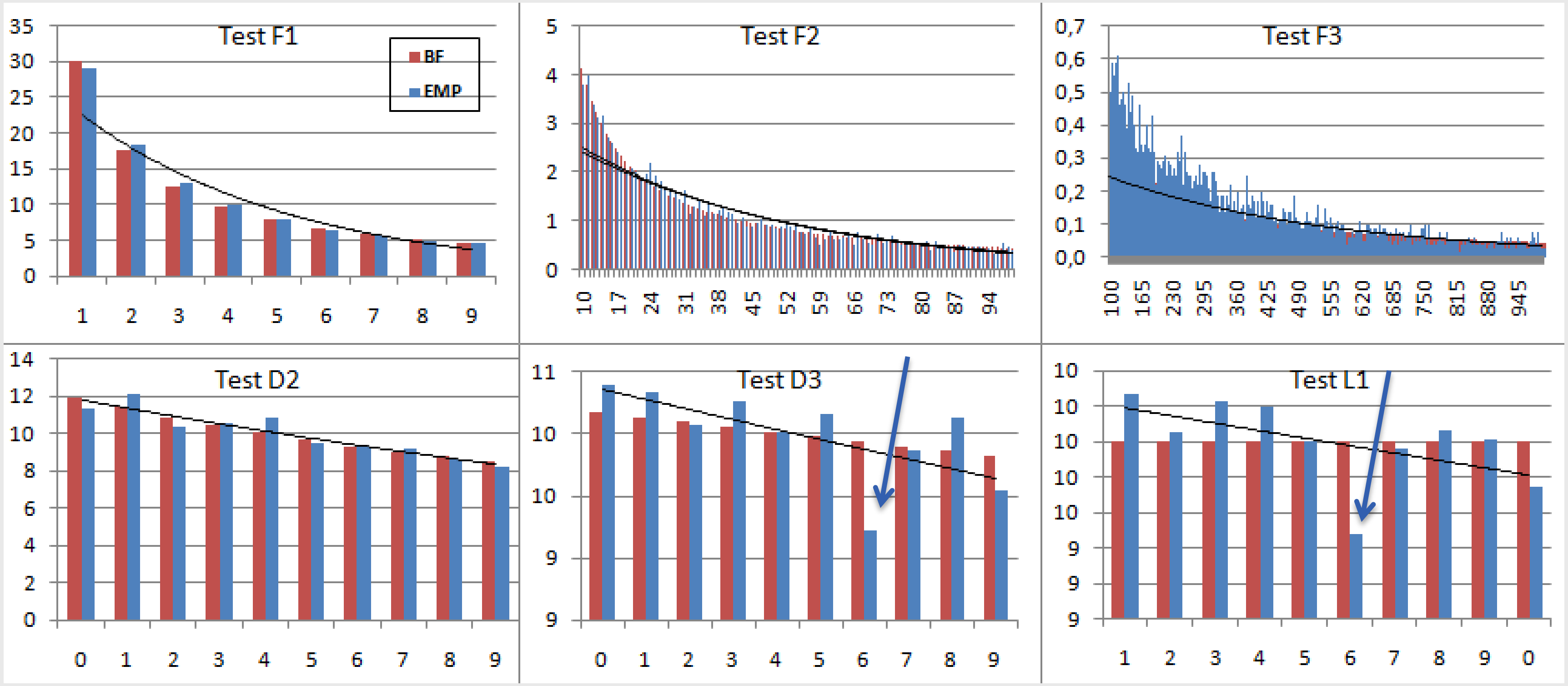
			α=0,05	1,36	1,36	1,75
			α=0,01	1,63	1,63	2,00
	chi	chi-0,05	p(chi)	KS1	KS2	KS3
F1	12,6	15,5	0,126	0,77	1,09	1,15
F2	107,8	112	0,085	0,85	1,20	1,58
F3	1208,6	969,9	0,000	6,32	8,93	6,32
D2	20,3	16,9	0,016	0,44	0,63	0,78
D3	6,8	16,9	0,659	0,49	0,69	0,49
L1	5,5	16,9	0,786	0,55	0,77	0,55

FV5 dataset analysis result (2)

Name	Sales invoices - polygraphic company												
FV5	Source:	Confident											
Test z	0	1	2	3	4	5	6	7	8	9	max	α=0,05	α=0,01
F1	-	2,38	2,29	1,38	0,98	0,07	0,71	0,16	1,14	0,2	2,38	2	0
D2	1,94	2,4	1,89	0,36	2,72	0,77	0,04	0,62	0,56	0,92	2,72	2	1
D3	0,66	0,6	0,09	0,62	0,03	0,53	2,16	0,09	0,81	0,86	2,16	1	0
L1	0,92	0,18	0,79	0,69	0,01	1,83	0,14	0,21	0,05	0,87	1,83	0	0
F2-10	1,8	1,04	0,61	0,54	0,97	0,64	0,33	0,45	1,76	1,42	1,8	5	2
F2-20	0,4	0,23	0,32	0,87	3,24	1,7	1,35	0,26	0,73	0,87	3,24	Sum:	
F2-30	0,09	2,06	1,71	0,08	1,2	1,16	1,9	0,19	0,92	1,58	2,06	10	3
F2-40	1,22	1,03	0,62	0,78	1,22	0,91	0,21	1,13	0,33	0,43	1,22		
F2-50	0,41	0,41	1,02	0,56	1,72	0,17	0,34	0,82	1,26	2,59	2,59		
F2-60	1,09	0,65	0,94	1,04	0,21	1,34	1,74	1,62	0,18	1,16	1,74		
F2-70	1,67	1,41	0,84	0,11	1,52	0,15	0,94	0,46	0,49	0,13	1,67		
F2-80	0,83	0,85	1,71	1,04	0,34	1,07	0,58	2	0,41	0,34	2		
F2-90	0,4	0,18	0,04	0,45	0,24	2,02	1,1	1,57	0,34	0,46	2,02	1,96	2,58
max	1,9	2,4	2,3	1,4	3,2	2	2,2	2	1,8	2,6	3,2		

FV5 dataset analysis result (3)

benford.pl



Chi-square test – quotient of empirical values to critical values

Chi-square	F1	F2	F3	D2	D3
FV 1	2,23	3,05	2,84	3,37	6,96
FV 2	0,83	0,97	1,59	0,90	14,42
FV 3	21,26	46,04	14,57	34,69	13,59
FV 4	2,79	2,35	2,50	0,72	0,96
FV 5	0,93	0,97	1,20	0,65	0,24
FV 6	2,85	3,04	3,04	6,81	23,25
FV 7	1,06	0,93	1,08	0,55	0,43
FV 8	5,74	3,44	1,78	5,66	6,92
FV 9	1,84	1,04	2,44	0,61	33,36
FV 10	3,01	2,26	1,40	2,23	0,38
FV 11	8,57	12,23	12,68	11,18	68,18
FV 12	6,80	10,20	11,84	30,00	162,57

Chi-square test – the rank of quotients of empirical values to critical values

Chi-square	D2	D3	F2	F1	F3	Sum
FV 3	4	4	4	4	4	20
FV 11	4	4	4	4	4	20
FV 12	4	4	4	4	4	20
FV 6	4	4	3	2	3	16
FV 8	4	4	3	4	1	16
FV 1	3	4	3	2	2	14
FV 9	0	4	1	1	2	8
FV 10	2	0	2	3	1	8
FV 4	0	0	2	2	2	6
FV 2	0	4	0	0	1	5
FV 7	0	0	0	1	1	2
FV 5	0	0	0	0	1	1

	rank	n	%	%
good	0	14	23,3	36,7
not bad	1	8	13,3	
rather bad	2	9	15,0	
bad	3	6	10,0	48,3
v.bad	4	23	38,3	
		60	100	

Chi-square test – Correlation matrix

Correlation matrix	Average emp./critic.	Coeff. of variation	Kurtosis	Skewness	No. of observations
Average emp./critic.	1,000	-0,149	0,101	0,122	0,459
Coeff. of variation	-0,149	1,000	0,907	0,908	0,000
Kurtosis	0,101	0,907	1,000	0,993	0,259
Skewness	0,122	0,908	0,993	1,000	0,258
No. of observations	0,459	0,000	0,259	0,258	1,000

Benford Law usage (1)

- Detection of fraud data or unintentional accounting error;
- Credit card fraud;
- Money laundering;
- Telecommunications fraud;
- Computer intrusion;
- Medical and scientific fraud;
- Analysis of data exchange;
- Election fraud;
- Designing a storage architecture in computers.

Benford Law usage (2)

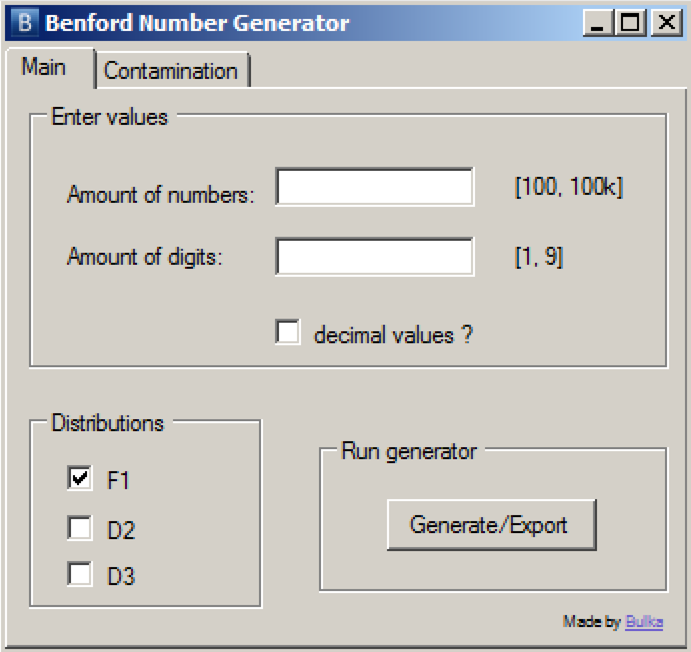
Conclusion

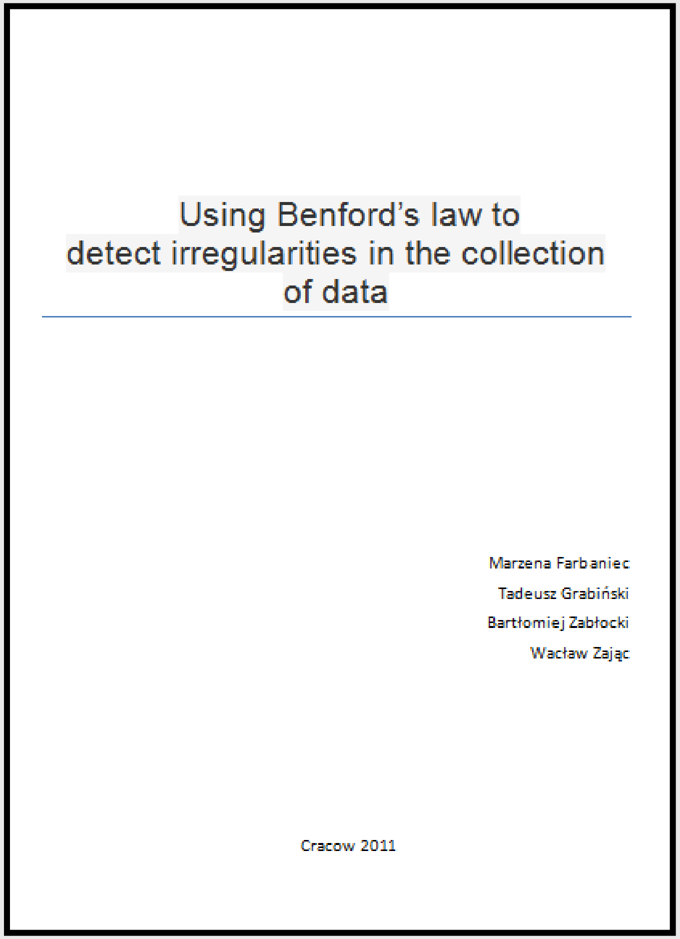
- Analysis technique can detect errors and conscious manipulation of data;

- All the calculated parameters, statistics and measures should be accordingly examinate in order to better interpretation of data and results obtained;

- In the case of anomalies in the distribution, closer examination of suspect parts of the data set is recommended, as they may indicate a conscious attempt to interfere with the data;

Currently in progress




Using Benford's law to detect irregularities in the collection of data

Marzena Farbaniec
Tadeusz Grabiński
Bartłomiej Zabłocki
Wacław Zając

Cracow 2011

Thank you !

info@benford.pl

Marzena Farbaniec
Tadeusz Grabiński
Bartłomiej Zabłocki
Wacław Zając

Andrzej Frycz Modrzewski Cracow University

References

- Durtschi C., Hillson W., Pacini C. The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data. Journal of Forensic Accounting 1524-5584 Vol. V(2004) , 17-34.
- Nigrini M. J., Miller S. J. Data Diagnostics Using Second Order Tests of Benford's Law. A Journal of Practice & Theory Vol. 28, No. 2 (Nov. 2009)
- Watrin Ch., Struffert R., Ullmann R. Benford's Law: an instrument for selecting tax audit targets? Rev Manage Sci (2008) 2:219–237
- Benford F. The Law of Anomalous Numbers. Proceedings of the American Philosophical Society, Vol. 78, No. 4 (Mar. 31, 1938), pp. 551-572
- Hill T. P. A Statistical Derivation of the Signicant-Digit Law. Stastical Science, Vol.10, No. 4 (1995), pp. 354-363
- Hill T. P. Base-Invariance Implies Benford's Law, Proceedings of the American Mathematical Society, Vol. 123, No. 3 (Mar 1995)