VALIDATION OF BENFORD'S LAW THROUGH ECONOMIC STATEMENTS OF COMPANIES CARRYING OUT BUSINESS IN THE PRODUCTION OF BASIC METALS, METALLURGICAL METAL PROCESSING AND FOUNDRY INDUSTRY

PLUNDER Karel, VOZŇÁKOVÁ Iveta, VEVERKOVÁ Jarmila

VSB - Technical University of Ostrava, Faculty of Metallurgy and Materials Engineering, Ostrava, Czech Republic, EU, karel.plunder@vsb.cz, iveta.voznakova@vsb.cz

Abstract

The entire world around us is filled with a certain numerical regularity; this can relate to the height of mountains, winning lottery numbers, numbers in tax declarations, the number of stars - this phenomenon is called Benford's Law. Benford's Law claims that the first significant (valid) place in naturally occurring numbers is more often occupied by low digits than high digits. So far, this topic has been paid little attention in Czech literature. In this paper, we decided to check whether the law also applies in the field of metallurgy. Incidence of digits in the first significant place was assessed using the set of statistical data of selected companies whose business is the production of basic metals, metallurgical metal processing and foundry activities.

Keywords: Benford's Law, metallurgical enterprise, digits in the first significant place

1. INTRODUCTION

This paper aims to confirm Benford's Law, i.e. to provide confirmations in the field of problematic uneven distribution of the first significant digits in naturally occurring numbers.

In many sets of naturally occurring numbers, low digits appear with a certain incidence in the first place more often than high digits. As already mentioned, this phenomenon is called Benford's Law and it has not yet been paid much attention on the part of the domestic professional public, rather on the contrary.

As already published under our conditions, our world carries a sort of mysterious key, a certain enigmatic numerical regularity which includes everything - from lengths of rivers, heights of mountains, the number of stars up to ordinary numbers from tax declarations or numbers winning in the lottery. Just this was affirmed by the scientist Benford who formulated his claims in a specific law.

It is an interesting fact that even major Czech periodicals, such as Hospodářské noviny, dealt with this reality and published an article called "Europe charms with numbers"; the article described how Benford's Law can be used to uncover manipulation with a set of natural numbers. In this case, mainly manipulations in the financial statements of individual countries, namely Greece, were discussed.

For the purpose of this paper, efforts were made to verify the validity of Benford's Law on a statistical data set acquired from publicly available sources where we assessed the incidence of the first digit in selected values of the financial statements of companies whose subject of enterprise is the production of basic metals, metallurgical metal processing and foundry activities.

2. BENFORD'S LAW

Simon Newcomb, American mathematician and astronomer, author of the article "Note on the frequency of use of the different digits in natural numbers", published in 1881, states in this work that the front or first pages of logarithmic tables are noticeably more worn and dirtier than the last pages. This fact led him to conclude that users of these tables, in their activities, significantly more often meet with numbers having the digits 1 or

2 in the first place, whose logarithms are listed on the front pages of tables, compared to those where the first place is occupied with digits 8 or 9. [1], [2]

The naive view would define that it is natural that the first significant digit standardly met by people can be one as well as eight or nine with the same probability. In this case, however, Newcomb claimed something that was in conflict with this intuitive idea.

During the 1930s, Frank Benford noticed the same different wear of the pages in logarithmic tables. In 1938, without knowing of the publication by S. Newcomb, Benford issued an article entitled “The law of anomalous numbers” thanks to which the reality (i.e. the fact that the distribution of incidence of the first significant digit in an array of sets is not even) was named Benford's Law. [3], [4]

2.1. Expression of Benford’s Law

This part specifies the affirmation (i.e. Benford's Law) that the distribution of digits in some naturally occurring sets of numerical data is such that their mantissas (when writing numbers in the decimal system) have Benford's distribution (see **Table 1**).

$$P(m(x) < t) = \log_{10} t, t \in [1,10] \quad (1)$$

Table 1 Benford’s distribution for the first significant digit

First significant digit	Benford’s distribution [%]
1	30.1
2	17.6
3	12.5
4	9.7
5	7.9
6	6.7
7	5.8
8	5.1
9	4.6

To formulate Benford's Law, it is also necessary to specify certain terms that have already occurred. The mantissa (when expressing numbers in the decimal system) is function m: $(0, \infty) \rightarrow [1, 10)$, which for each $x \in (0, \infty)$ can be expressed in the form $x = m(x) \cdot 10^n$ for some $n \in Z$. Such a number $m(x)$ in the interval [1, 10) is just one and the definition is therefore correct (Equation (2)). If it is further necessary to distinguish that the mantissa relates to a number expressed in the decimal system, $m^{(10)}(x)$ designation will be used. Similarly, it is possible to define mantissas for numbers expressed in systems with another base. Random variable X has Benford’s distribution if:

$$P(X < t) = \log_{10} t, t \in [1,10] \quad (2)$$

This affirmation (according the Equation (1)) suffers from one deficiency. It only says that the distribution of mantissas is logarithmic (i.e. Benford's distribution) in some sets of numerical data. It provides no guidance on how to decide in advance which set of data has this property or not. Simon Newcomb as well as Frank Benford reached the expression corresponding to Equation (1), but in different ways. [5]

Benford based his claims on empirical observations. For many years, he collected numerical data across fields and sources. He gathered a set of data containing more than 20,000 figures, such as the catchment areas of

335 rivers, specific latent heat of 1389 chemical compounds, the numbers appearing on the front page of newspapers, and others. For this set of collected data, Benford presented that the first digits really did not appear with the same frequency. In search of a simple formula (see Equation (3)) which could describe the distribution of the first significant digits in his data, he came to the following expression:

$$P(D_1(x) = d) = \log_{10}(1 + \frac{1}{d}), d = 1,2,..,9. \quad (3)$$

3. DATA SELECTION AND PROCESSING

To statistics in general: One of the definitions characterizes statistics as a scientific discipline concerned with the collection, analysis, organization, interpretation and final presentation of data. The task of this discipline consists in the process of describing the properties of the given object or phenomenon using characteristics found from the data. Statistics can be considered the kind of discipline in which it is possible to use the built system for the benefits across all areas in view, both from the sphere of natural, technical and social sciences. [6]

In preparation leading to the data processing, it is necessary to adhere to certain generally accepted facts including the characteristics of the target information, where it is needed to characterize a group (set) on which we want to learn something. This group must be defined before the start of data collection and the data set must characterize the information representatively. Current statistical analysis requires data processing through statistical software. A prerequisite for success is the proper storage of data in the form of a “database” table allowing its processing in any application. Paying attention to data refining prior to analysis is equally important. Any error that occurs or is not found in the phase of data preparation will be reflected in all subsequent steps and may cause invalid results and the need for repeating the analysis.

3.1. CZ-NACE

Classification of economic activities or CZ-NACE is based on the name “Nomenclature générale des activités économiques dans la Communauté Européenne”; it is a statistical classification of economic activities used by the European Union (or the European Community) since 1970. NACE provides a framework for statistical data on activities in many economic areas (e.g. in production, employment, national accounts). Produced statistics using the NACE classification are comparable across the European Union. With a lower level of detail (at higher levels), they can be compared even with world statistics. The use of NACE is mandatory for all Member States of the European Union. At-world-level comparability of data created according to the NACE classification is given by the fact that NACE is part of a system of statistical classifications developed mainly under the auspices of the Statistics Division of the United Nations.

Classifications divide the statistically observed world (reality) into parts that are more or less homogeneous inside, depending on the statistical viewpoint applied. Statistical classifications are characterized by: a) a thorough coverage of the observed reality; b) mutually exclusive categories: each unit can be classified only in one category; c) methodological principles which allow consistent classification of all units in the classification categories. [7]

For the issue investigated, i.e. the determination and selection of a statistical sample / data set, we used CZ-NACE 24 (Production of basic metals, metallurgical metal processing; foundry industry).

3.2. Statistical sample - data set

As previously characterized, the subject of this article is to validate Benford's Law on a data set, specifically on selected values from economic statements of companies that fall within the statistical group CZ-NACE 24 (Production of basic metals, metallurgical metal processing; foundry industry).

Using a publicly available source, the Administrative Registry of Economic Entities, established by the Ministry of Finance of the Czech Republic (on-line database www.info.mfcr.cz/ares), we selected one hundred companies and processed data from their financial statements, such as balance sheets and profit and loss statements, and thus created a data set with five hundred values, intended for further processing. Specifically, we processed the following items of balance sheets: total gross assets for the given year, total net assets for the given year and total net assets for the previous year. Regarding the profit and loss statements, we included the incomes from operations for the given and previous year. After selecting the data, we found that some companies from the hundred selected started their economic activity in the current year and that it is therefore impossible to assess statistical sets concerning the total assets for the previous year in the balance sheets, and the incomes from operations for the previous year in the profit and loss statements. For the actual validation of Benford's Law, we therefore used the remaining data set containing three hundred numbers. Statistical processing was carried out with regard to the three statistical sets divided according to their meaning, i.e. total gross assets, total net assets and incomes from operations, and also with regard to the whole set of three hundred numbers. I would like to add that the selected companies were only legal persons, specifically commercial companies, joint-stock companies (29 %) and limited liability companies (71 %).

4. RESULTS OF THE ANALYSIS PERFORMED

In total, we investigated four data sets (data selected values from economic statements of companies with business proceses - production of basic metals, metallurgical metal processing, foundry industry) of naturally occurring numbers; the first three sets contained one hundred samples (numbers) and the fourth set was a set of the first three sets containing three hundred samples (numbers).

The first performed analysis showed that the statistical set of values occurring as total gross assets from one hundred balance sheets of selected companies with identical economic activity as mentioned above contained numbers with the following distribution of first significant digits (see **Table 2**): 1 (incidence 26 %), 2 (incidence 21 %) … 9 (incidence only 4 %). Comparison between Benford's distribution and the results of the analysis for the first set is graphically shown in **Figure 1**.

Table 2 Results of the first nalysis for the incidence of the first significant digit

First significant digit	Distribution according to the analysis [%]
1	26
2	21
3	18
4	7
5	8
6	7
7	4
8	5
9	4

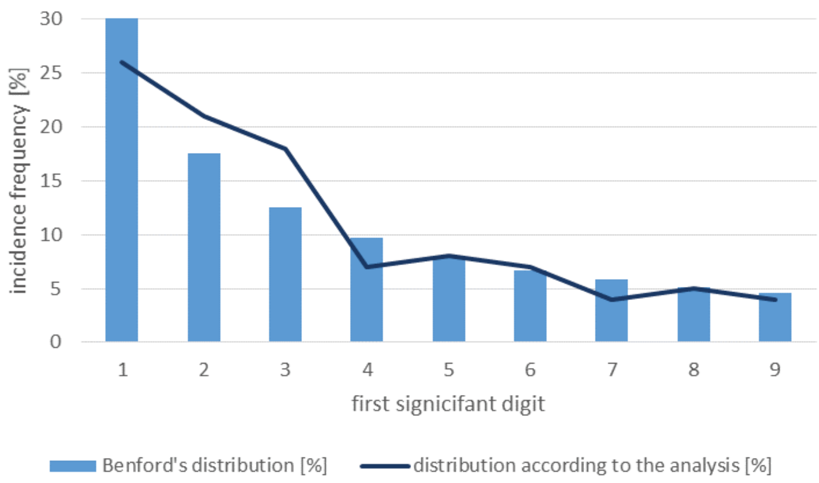

Figure 1 Graphical representation of the incidence of the first digit for the first

The following analysis of the incidence of the first significant digit for the next set of natural numbers (total net assets) showed the following distribution: 1 (incidence 21 %), 2 (incidence 23 %) 3 (incidence 16 %), 4 (incidence 14 %), 5 (incidence 5 %), 6 (incidence 8 %), 7 (incidence 3 %), 8 (incidence 5 %) and 9 (incidence 5 %) - see **Table 3** and the graphical representation in **Figure 2**.

Table 3 Results of the second analysis for the incidence of the first significant digit

First significant digit	Distribution according to the analysis [%]
1	21
2	23
3	16
4	14
5	5
6	8
7	3
8	5
9	5

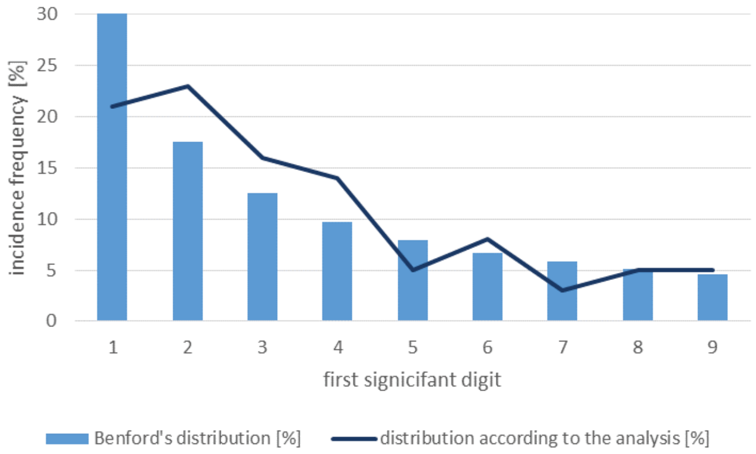

Figure 2 Graphical representation of the incidence of the first digit for the second analysis

The third analysis again assessed he sample with one hundred values - incomes from operations for the current accounting period from the profit and loss statements of one hundred companies. Distribution of the first significant digit was as follows: 1 (incidence 27 %), 2 (incidence 16 %), 3 (incidence 9 %), 4 (incidence 11 %), 5 (incidence 10 %), 6 (incidence 5 %), 7 (incidence 6 %), 8 (incidence 6 %) and 9 (incidence 8 %) - see **Table 4** and the graphical representation in **Figure 3**.

Table 4 Results of the third analysis for the incidence of the first significant digit

First significant digit	Distribution according to the analysis [%]
1	27
2	16
3	9
4	11
5	10
6	5
7	6
8	6
9	8

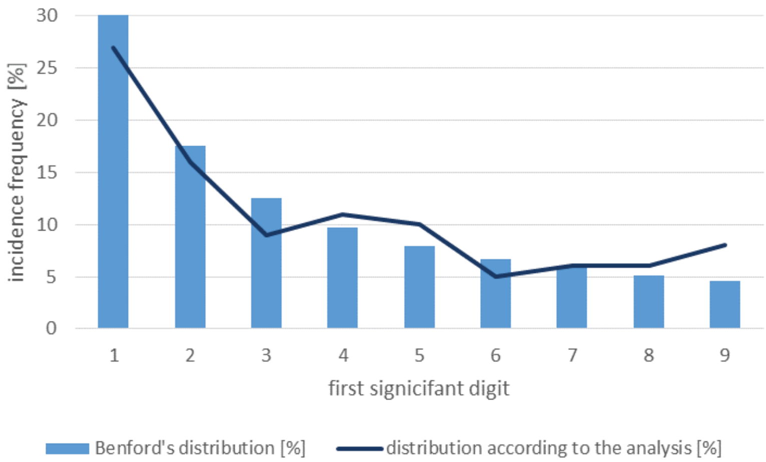

Figure 3 Graphical representation of the incidence of the first digit for the third analysis

The last analysis is conducted for the set of all of the selected data; it therefore comprises three hundred values. The respective results of the incidence of the first significant digit are shown in **Table 5**; the graphical representation of the distribution found by the analysis and Benford's distribution are presented in **Figure 4.**

Table 5 Results of the fourth analysis for the incidence of the first significant digit

First significant digit	Distribution according to the analysis [%]
1	24.7
2	20.0
3	14.3
4	10.7
5	7.7
6	6.7
7	4.3
8	5.3
9	5.7

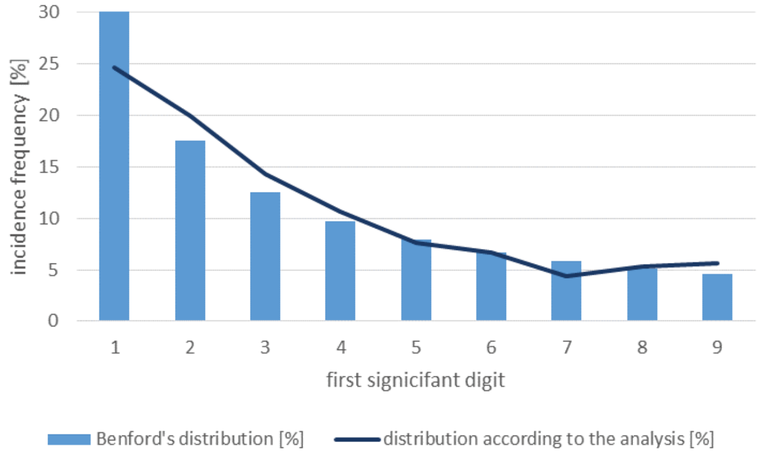

Figure 4 Graphical representation of the incidence of the first digit for the fourth analysis

5. CONCLUSION

Motivation for research in the area of the presented issue has two roots. One of them is to popularize the topic of Benford's Law or Benford's distribution, also due to the fact that relatively little attention is paid to this issue in scientific circles whereas the topic is also largely neglected in the context of domestic publications. Noticed already by Newcomb in 1881 and then by Benford in 1938, the phenomenon is a striking fact that appears in the sets of naturally occurring numbers in our world. Another motivation for research in this field was the possibility of verifying the mentioned law or distribution on sets of numbers and values submitted by companies operating in the field of metal processing. In our opinion, the results of the four analyses performed on the set of three hundred values confirmed the validity of Benford's Law on the incidence of the first significant digits in sets of naturally occurring numbers. There is a possibility than to outputting the results on data sets of the metallurgical companies. It could concern the economical as well as the production data. In this case Benford´s Law could serve as a validity verification tool of the metallurgical concern data.

ACKNOWLEDGEMENTS

The work was supported by the specific university research of Ministry of Education, Youth and Sports of the Czech Republic No. SP2016/107.

REFERENCES

[1] NEWCOMB, S. *Note on the frequency of use of the different digits in natural numbers*, Amer. J. Math. Vol. 4, 1881, pp. 39-40.

[2] KANTOREK, P. Benfordův zákon, *Vesmír*, 1998, Vol. 77, pp. 583

[3] BENFORD, F.: *The law of anomalous numbers*, Proc. Amer. Philos. Soc. Vol. 78, 1938, pp. 551-572.

[4] KOSSOVSKY, A. E. *Benford's Law: Theory, the General Law of Relative Quantities, and Forensic Fraud Detection Applications.* Singapore: World Scientifi c Publishing, 2015.

[5] PLAČEK, M. Application of Benford's Law to government macroeconomic data: the case of trade balance. *Trendy ekonomiky a managementu.* 2014, vol. VIII, no. 19, p. 45-51. ISSN 1802-8527.

[6] BUDÍKOVÁ, M., M. KRÁLOVÁ and B. MAROŠ. *Guide to basic statistical methods.* Prague: Grada Publishing, a.s., 2010. ISBN 978-80-247-3243-5.

[7] Methodological guide to NACE (CZ-NACE). Czech Statistical Office. [online] [cit 2016-04-01]. Available from: https://www.czso.cz/documents/10180/23174387/metodicka_prirucka_cz_nace_rev_2.pdf/e26ebee3-a5b2-48a1-a036-75e14cdb8944?version=1.0