**Testing Benford’s Laws (Non)Conformity within Disclosed Companies’ Financial Statements among**

**Hospitality Industry in Romania**

**Claudiu HERTELIUa,b,\*, Ionel JIANUc, Irina Maria DRAGANa,**

**Simona APOSTUa, Iuliana LUCHIANd**

aDepartment of Statistics and Econometrics, Bucharest University of Economic Studies, Romania bSchool of Business, London South Bank University, UK cDepartment of Accounting and Audit, Bucharest University of Economic Studies, Romania dPhD Student, Business Administration Doctoral School, Bucharest University of Economic Studies, Romania

\**Corresponding author:* Email: hertz@csie.ase.ro , hertelc2@lsbu.ac.uk; Address: Department of Statistics and Econometrics, Virgil Madgearu Building, University of Economic Studies, Calea Dorobantilor 15-17, Bucharest, 010552 Sector 1, Romania; Phone: +4 0722 455 586

**Abstract**

Nowadays, the Benford’s Laws is tested in plenty of academic domains as the first check of data reliability. Benford’s Laws non-conformity does not necessarily mean that there is a case of data manipulation. However, such a test can raise the first signal for further prudence in conclusions based on that specific dataset. Romania is considering to have a higher rate for fiscal evasion, and the cash propensity within the hospitality industry is high. In this context, we expect that hospitality industry companies’ fiscal statements to be Benford non-conformant. The dataset consists of more than 574 thousand Romanian companies’ disclosed financial statements for the year 2013. The control variable is the company’s size (from the number of employees point of view). Against our expectations, we find very small discrepancies in the dataset’s empirical distributions compared to Benford’s Laws. This is the case for the first digit, second digit and the first two combined digits.

**Key-words:** Benford’s Laws; Newcomb Law; tourism data; hospitality industry; financial statements; profit before tax

**Highlights:**

* Companies’ disclosed financial statements conformity to Benford’s Laws are tested
* The large dataset consist of public data obtained from the Romanian Ministry of Finance
* Against our expectations, we find minimal discrepancies in the empirical distributions in the dataset when compares to Benford’s Laws

**1. Context**

The high level of tax evasion in Romania is almost an axiomatic statement (Toader, 2007; Constantin *et al.*, 2011; Brezeanu *et al.*, 2011; Canagarajah *et al.*, 2012; Lazar, 2013; Todor, 2018; Androniceanu *et al.*, 2019; Bucur *et al.*, 2019). In addition, the propensity of cash payments regarding the hospitality industry suggests a high risk for tax evasion (Hasseldine & Bebbington, 1991; Brown, 1998; Ott, 2002; Crnogorac & Lago-Peñas, 2018; Jaliu & Ravar, 2019 ). Other unethical, immoral or illegal behaviour categories are specific to the hospitality industry. These categories are grounded from a company’s perspective, deceived by its own employees and the customers deceived regarding the services requested or paid for them (Murphy, 2015). This field has recently faced another challenge: the collaborative economy (Williams & Horodnic, 2017).

This article aims to test how the companies’ financial statements in the hospitality industry in Romania comply (or not) with distribution laws, currently known as Benford’s Laws.

Regarding the beginnings of this law of distribution, we can appreciate that its paternity belongs to an American[[1]](#endnote-1) mathematician/astronomer: Simon Newcomb (1881). Studying the books used by the students at the library in the multiplication process (at that time, in the absence of computers, the students used the logarithmic tables), he observed that the beginning part (corresponding to numbers beginning with 1, 2 or 3) of these books is much more used than the ending part (corresponding to the numbers beginning with 7, 8 or 9).

After more than half a century, an American engineer/ physicist Frank Benford (1938) rediscovered this distribution law by publishing much larger material (twenty-two pages compared to Newcomb’s two-pages text) and well documented empirical study (all the work was done manually!). Another half-century went by until another American[[2]](#endnote-2) came, Mark John Nigrini (1992), bringing back the benefits of using Benford’s Laws in the field of accounting and/ or auditing. He succeeded in his doctoral thesis (Nigrini, 1992) developed at the University of Cincinnati to revitalise and popularise in the area of socio-economic sciences the laws currently known as Benford’s Laws.

It is worth mentioning that regarding the chronology of the works dedicated to Benford’s Laws between the mentioned landmarks (1881-1938-1992), there were many entries from different academic areas: mathematics, physics, psychology, statistics etc. Prominent members of the academic community have made such contributions: the famous French mathematician Henri Poincaré (1912), Nobel laureate in economics George Joseph Stigler[[3]](#endnote-3) (1945) or reputed Italian statistician Corrado Gini (1957). A comprehensive list of academic papers focused on Benford’s Laws is available on the website (WWW1, 2020). The enthusiasm with which Benford’s Laws is currently applied/tested/analysed is also illustrated by the multiple domains and territories from which the data sets are used. Thus, as a symbolic example, we distinguish the economic and financial areas (Clippe & Ausloos, 2012; Mir *et al.*, 2014; Ausloos *et al.*, 2016; Riccioni & Cerqueti, 2018; Druica *et al.*, 2018; Shi *et al.*, 2018; Kaiser, 2019; Miranda-Zanetti *et al.*, 2019; Jianu & Jianu, 2021; Ausloos *et al.*, 2021), scientometrics area (Hürlimann, 2015; Alves *et al.*, 2016; Mir & Ausloos, 2018), demography area (Ausloos *et al.*, 2015), Google searches area (Ileanu *et al.*, 2018), hydrology area (Ausloos *et. al.*, 2017; Huang *et al.*, 2019), astronomy area (Alexopoulos & Leontsinis, 2014), religion (Mir, 2014) or, very recent, epidemiology of Covid-19 (Lee *et al.*, 2020; Kennedy & Yam, 2020). On the institutional level, very recent, European Commission organised a dedicated conference on Stresa (Italy) on Benford’s Laws use, methods and further advance (WWW2, 2020).

In the current paper, we analyse how the financial statements of the Romanian companies from the hospitality industry follow or not Benford’s Laws. The hypotheses to be tested is that, in the context of a high level of tax evasion and the propensity of cash payments in the field, the information disclosed in the financial statements of these companies does not follow Benford’s Laws.

**2. Methodology and data sources**

**2.1. Benford’s Laws**

Contrary to other distribution laws studying numbers in their integrality, Beford’s laws are looking to numbers’ digits. Although there are generalisations, the most frequent approach is checking the distribution of the first digit from a specific dataset (BL1). Hence, concordant to studies which are dealing with this approach (Newcomb, 1881; Benford, 1938; Durtschi, 2004; Nigrini, 2015; Ausloos *et al.*, 2015), the expected (theoretically) Benford probability of occurrence of a specific digit (d1 from one to nine) on the first position (D1) within a number is:

d1=(1,2,3,…,9) (1)

With basic computations from equation (1) we will have the expected Benford probabilities as those shown in table 2.1.

**Table 2.1.** First digit (d1) Benford’s Law expected frequencies on the first position (D1) within a number

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| d1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| P(D1=d1) | 0.301 | 0.176 | 0.125 | 0.097 | 0.079 | 0.067 | 0.058 | 0.051 | 0.046 |

On the next step when testing Benford’s Laws, one should test the second digit distribution (BL2) in a specific dataset. Using the specific computation method suggested by Durtschi (2004), the expected (theoretically) Benford probability for a specific digit (d2 from zero to nine) on the second position (D2) within a number is:

d2=(0,1,2,3,…,9) (2)

Hence, the expected Benford probability for zero to occur on the second digit within a number can be computed as follow:

Similar, after computations, using equation (2) for all other values (from one to nine), the expected Benford probabilities for the second digit in a number are presented in the table 2.2.:

**Table 2.2.** The second digit (d2) expected Benford’s Law frequencies on the first position (D2) within a number

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| d2 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| P(D2=d2) | 0.1197 | 0.1139 | 0.1088 | 0.1043 | 0.1003 | 0.0967 | 0.0934 | 0.0904 | 0.0876 | 0.0850 |

Although its occurrence in the researches on the empirical datasets is much lower than BL1 and/ or BL2, the probabilities of the combinations of the first two digits (having form) within a specific number is still a useful tool. Following the approach suggested in the specific literature (Durtschi, 2004; Nigrini, 2015) the expected (theoretically) Benford probability for the occurrence of a particular combination having form on the first two positions within a number is (BL12):

=(10,11,12,33,…,98,99) (3)

Using equation (3) for all first two digits combinations =(10,11,12,33,…,98,99) after basic computations, the expected Benford probabilities are enlisted in Table 2.3.

**Table 2.3.** The first two digits combined () expected Benford’s Law frequencies on the first two positions within a number

|  | **P(** |  | **P(** |  | **P(** |
| --- | --- | --- | --- | --- | --- |
| 10 | 0.04139 | 40 | 0.01072 | 70 | 0.00616 |
| 11 | 0.03779 | 41 | 0.01047 | 71 | 0.00607 |
| 12 | 0.03476 | 42 | 0.01022 | 72 | 0.00599 |
| 13 | 0.03218 | 43 | 0.00998 | 73 | 0.00591 |
| 14 | 0.02996 | 44 | 0.00976 | 74 | 0.00583 |
| 15 | 0.02803 | 45 | 0.00955 | 75 | 0.00575 |
| 16 | 0.02633 | 46 | 0.00934 | 76 | 0.00568 |
| 17 | 0.02482 | 47 | 0.00914 | 77 | 0.00560 |
| 18 | 0.02348 | 48 | 0.00895 | 78 | 0.00553 |
| 19 | 0.02228 | 49 | 0.00877 | 79 | 0.00546 |
| 20 | 0.02119 | 50 | 0.00860 | 80 | 0.00540 |
| 21 | 0.02020 | 51 | 0.00843 | 81 | 0.00533 |
| 22 | 0.01931 | 52 | 0.00827 | 82 | 0.00526 |
| 23 | 0.01848 | 53 | 0.00812 | 83 | 0.00520 |
| 24 | 0.01773 | 54 | 0.00797 | 84 | 0.00514 |
| 25 | 0.01703 | 55 | 0.00783 | 85 | 0.00508 |
| 26 | 0.01639 | 56 | 0.00769 | 86 | 0.00502 |
| 27 | 0.01579 | 57 | 0.00755 | 87 | 0.00496 |
| 28 | 0.01524 | 58 | 0.00742 | 88 | 0.00491 |
| 29 | 0.01472 | 59 | 0.00730 | 89 | 0.00485 |
| 30 | 0.01424 | 60 | 0.00718 | 90 | 0.00480 |
| 31 | 0.01379 | 61 | 0.00706 | 91 | 0.00475 |
| 32 | 0.01336 | 62 | 0.00695 | 92 | 0.00470 |
| 33 | 0.01296 | 63 | 0.00684 | 93 | 0.00464 |
| 34 | 0.01259 | 64 | 0.00673 | 94 | 0.00460 |
| 35 | 0.01223 | 65 | 0.00663 | 95 | 0.00455 |
| 36 | 0.01190 | 66 | 0.00653 | 96 | 0.00450 |
| 37 | 0.01158 | 67 | 0.00643 | 97 | 0.00445 |
| 38 | 0.01128 | 68 | 0.00634 | 98 | 0.00441 |
| 39 | 0.01100 | 69 | 0.00625 | 99 | 0.00436 |

**2.2. The dataset**

The analysed dataset information at the individual (company) level is a public one. According to the Romanian legislation, the following variables from the financial statements disclosed by any company are publicly available (WWW3, 2020): (i) turnover (in lei[[4]](#endnote-4)); (ii) profit before tax (in lei); (iii) number of employees and (iv) the main activity domain (four digits NACE). The fiscal reference year is 2013. The dataset contains information about 574 221 companies.

The very well-known nomenclature: NACE (statistical classification of economic activities in the European Community) is provided, maintained and updated by Eurostat (2008). According to NACE, the economic activities are split in 21 sections (coded with capital letters from A to U) and 88 divisions (numbered from 01 to 99). We include only those companies which are belonging to the hospitality industry to reach our aim. In NACE these companies are recalled in Section I: Accommodation and food service activities consisting of two divisions: Division 55, namely Accommodation and Division 56, namely Food and Beverage service activities.

From a different perspective, testing (the first) Benford’s Law conformity requires numbers that are not equal to zero. However, some registered companies are dormant. This category of companies is identified using the following flags: (i) turnover is zero; (ii) the number of employees is zero; and (iii) the profit before tax is zero. A company is considered to be dormant if all these flags are jointly raised. The synthetic information with the companies’ distribution after assigning the “dormant” label is shown in Table 2.4.

**Table 2.4.** Companies’ distribution by status

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Dormant status?** | **All companies** | **out of which** | | |
| **Hospitality industry** | **out of which** | |
| **Accommodation** | **Food and Beverages** |
| No | 488 842 | 24 087 | 5 381 | 18 706 |
| Yes | 85 379 | 4 263 | 982 | 3 281 |
| **Total** | **574 221** | **28 350** | **6 363** | **21 987** |

**Data source:** Authors’own calculations using raw dataset (WWW3, 2020)

**Note:** Light green background highlights the sample(s) used in the current manuscript.

At a glance, data inspection reveals that the share of dormant companies is relatively constant at a level of about 15% (for both the whole dataset and when it is split into sections/ divisions taken into consideration). Out of more than 574 thousand companies, the share of those operating in the hospitality industry is 4.94% (28 350 companies). Within this section, almost one in five (22.4%) companies are active in accommodation while the rest (77.6%) is offering food and beverage services.

Another control variable used in the current analysis is the company size (from the number of employees point of view). There is used the Eurostat (WWW4, 2020) definition for the company size: (i) micro-companies (those having less than ten employees); (ii) small companies (those having between 10 and 49 employees); (iii) medium companies (those having between 50 and 249 employees) and (iv) large companies (those having more than 250 employees). To calibrate the way in which the used dataset fits official statistics information (Romanian Statistical Yearbook, ASR, 2014), in table 2.5. the dual sources of data distribution of companies by size is shown.

**Table 2.5.** Companies’ distribution active in the hospitality industry by size in 2013

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Size classes** | **Dataset** | **ASR, 2014** |
| 1. | Micro (1-9 employees) | 20 852 | 20 682 |
| 2. | Small (10-49 employees) | 2 963 | 3 324 |
| 3. | Medium (50-249 employees) | 250 | 266 |
| 4. | Large (at least 250 employees) | 21 | 25 |
|  | **Total** | **24 086** | **24 297** |

**Data source:** Authors’own calculations using the raw dataset(WWW3, 2020)and ASR (2014)

**Note:** Light orange background highlights that the sample is not big enough to test Benford’s Law conformity for the first two combined digits. Light blue background highlights that the sample is not big enough to test any Benford’s Laws conformity.

When comparing the dual sources of data companies’ distribution in 2013 in the hospitality industry, a good concordance with official statistics data is found for the whole size classes except one. Indeed, for small companies, there are in the dataset about 10% fewer companies than has been reported in the statistical yearbook (2 963 small companies in dataset versus 3 324 in the statistical yearbook).

The specific data analysis for testing Benford’s Laws (non)conformity focuses on the **profit before tax**.The sample (hospitality industry companies)size consists of over 24 thousand companies not being dormant. In Table 2.4., the companies under consideration for this current research are highlighted by a light green background. For clarity, the sample size is provided for each Benford’s Law (non)conformity test. We choose to do so since in some cases: (i) the first digit of the profit before tax is zero and/ or (ii) the second digit of profit before tax does not exist.

Data analysis is performed via SPSS 16.0 while the pictures are designed in MS PowerPoint 2013.

**2.3. Benford’s Laws (non)conformity statistical testing**

The most popular test to check if an actual distribution (in absolute figures) is obeying Benford’s Laws is Chi-square () test.

The two hypotheses to be tested are:

H0: The empirical distribution does obey Benford’s Law

H1: The empirical distribution does not obey Benford’s Law

The computed figures for this test for each version (BL1, BL2 and BL12) are the following:

(4)

where are expected (theoretical) Benford frequencies for the first digit (Table 2.1.) while represents the actual (empirical) first digit frequencies;

(5)

where are expected (theoretical) Benford frequencies for second digit (Table 2.2.) while is used for actual (empirical) second digit frequencies;

(6)

where are the expected Benford frequencies for the first two combined digits (Table 2.3.) while denotes the actual (empirical) first two combined digits frequencies.

The used probability level to test if the actual distribution fits (or not) the Benford’s one is set to 0.95 (significance level: α=0.05). Critical values for this significance level are 15.51 for the first digit (eight degrees of freedom), 16.92 for the second digit (nine degrees of freedom) and 112.02 for the first two combined digits (89 degrees of freedom).

Classical in testing statistical hypotheses, if the computed value for the Chi-square test () is greater than its corresponding critical value, the null hypothesis (H0) is rejected.

It is commonly accepted that to use Chi-square () test the minimum level for absolute figures for any expected frequency is five. Under this restriction, when first or second digit distribution is tested, the minimum sample sizes should be 50, while for the case of the first two combined digits the threshold is 450. In the context of companies’ distribution by above-mentioned criteria (NACE division – table 2.4. and company size – table 2.5.), since there are: (i) 21 large companies (highlighted with light blue background in table 2.5) hence no testing of Benford’s Laws (neither for the first digit nor for second or first two combined digits); (ii) 250 medium companies (highlighted with a light orange background in table 2.5) hence no testing of Benford’s distribution for first two combined digits.

**3. Data analysis**

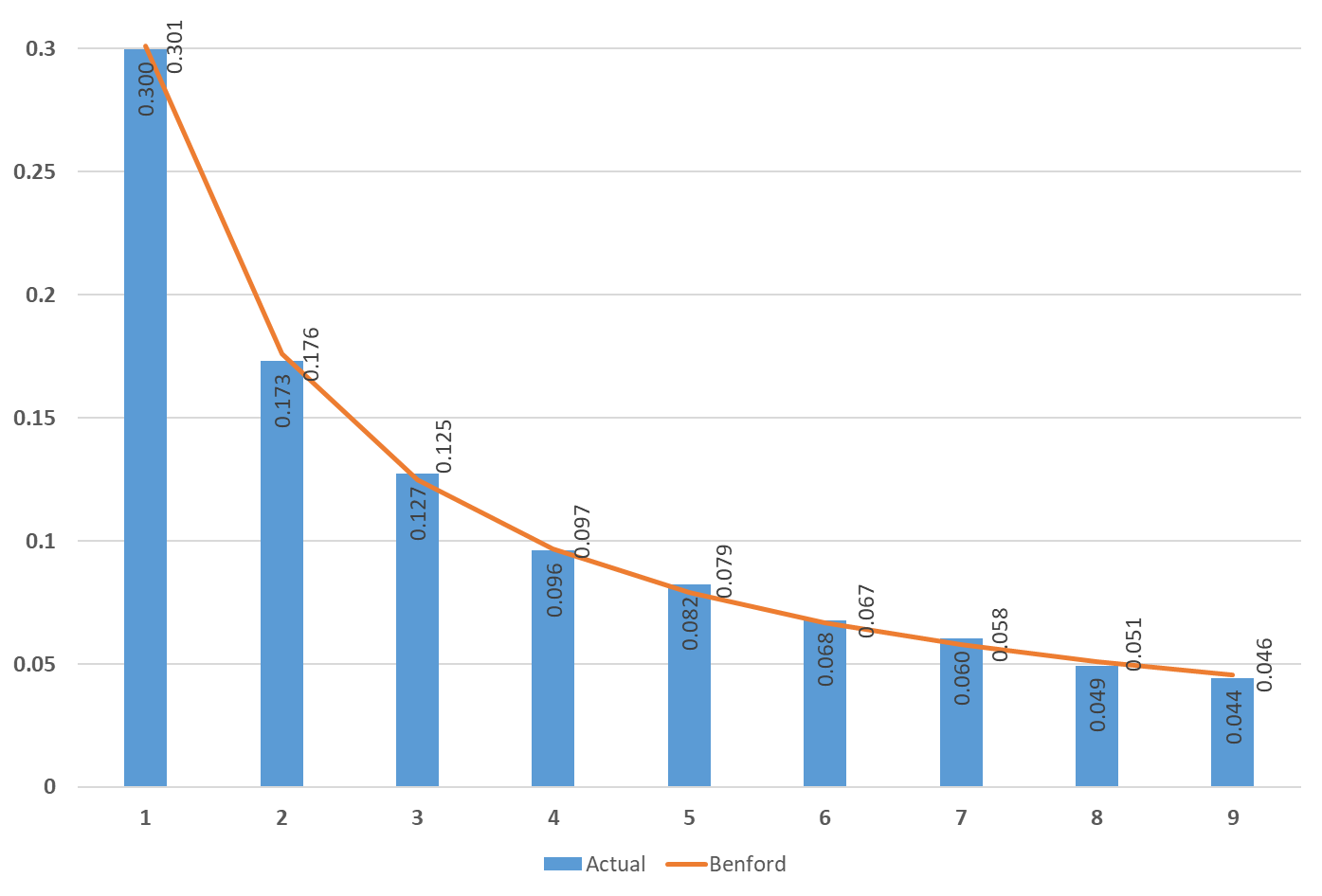
The conformity test of the empirical distributions for various cases (first digit, second digit or the combination of the first two digits) and samples (the whole section of the hospitality industry, the accommodation division, the food and beverage division, micro type companies, small companies, medium-sized companies) do not imply any special computational skills.

The first step is the visual inspection of the respective distributions (figures 3.1-3.17). How the empirical distributions obey Benford’s Laws is surprising. The description of the results identified visually is divided into two sections. The first one regards the main sample (hospitality industry) divided into two divisions (accommodation and food and beverage). The second section focuses on the samples obtained according to the company size criterion.

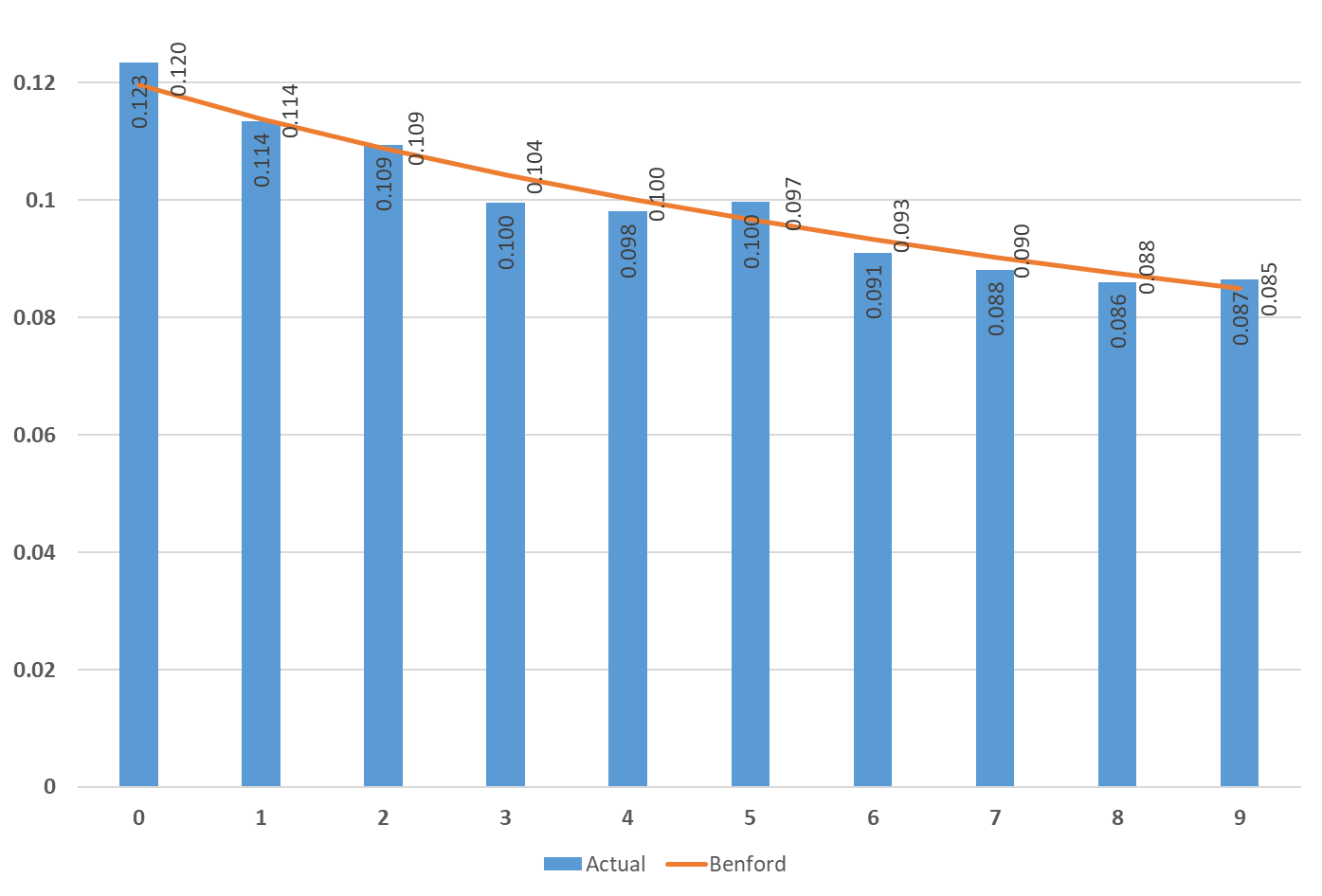
**3.1. The visual approach for the entire section of the hospitality industry and its separation on the accommodation and food and beverage divisions**

In the first three pictures, we have tested conformity against Benford’s Laws for the first digit, the second digit and the combination of the first two digits for the entire section of the hospitality industry. The following six figures consider the same issue for the accommodation (three figures) and food and beverage (three figures) divisions. In these nine figures, the empirical distribution follows Benford’s Laws almost perfectly. There are minimal dissimilarities that can be identified.

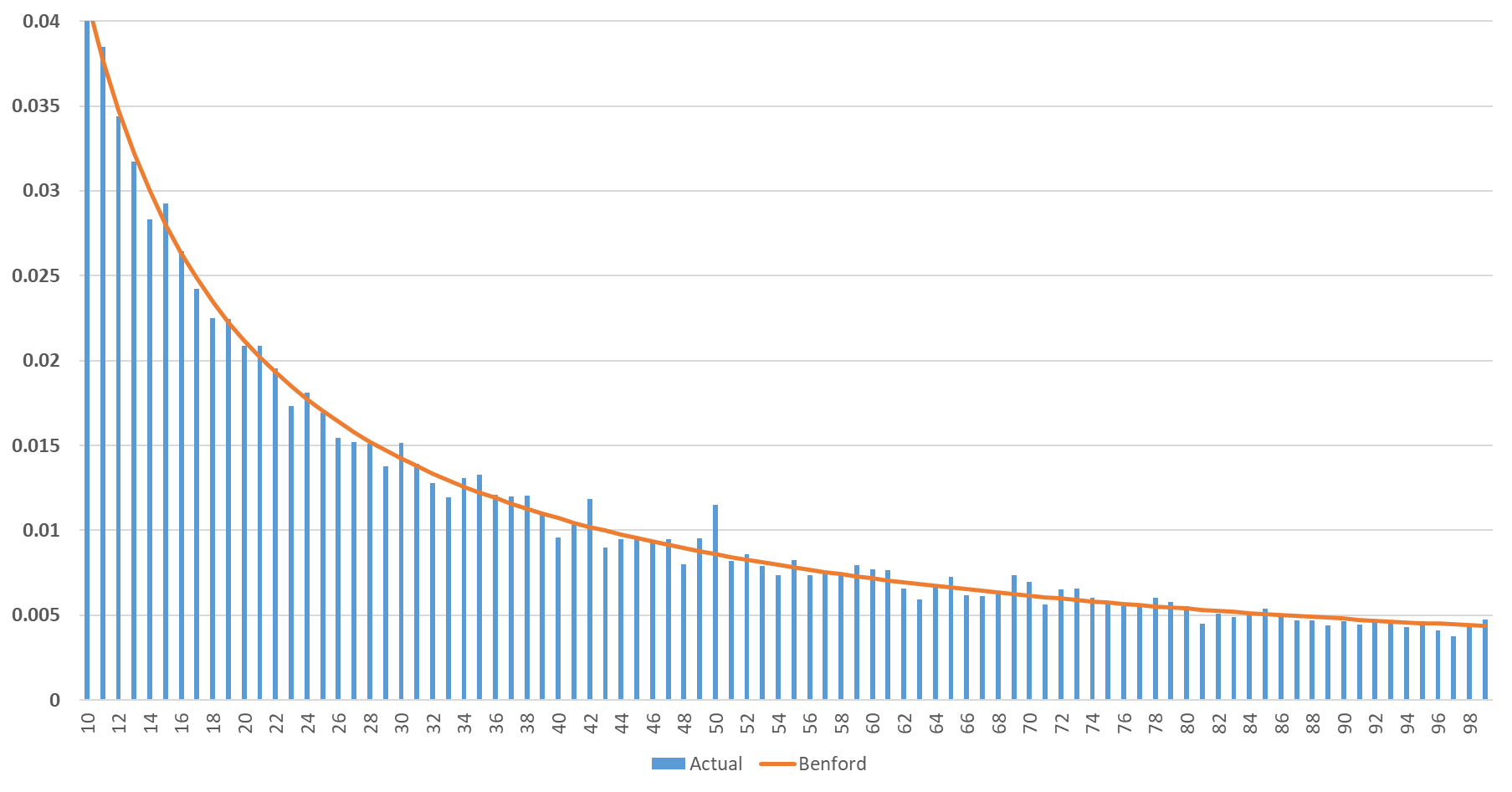
Thus, in the case of the first digit, the empirical frequency for digit 1 is slightly lower than the theoretical distribution (30.1%): 30.0% for the entire hospitality industry (figure 3.1) and the food and beverage division (figure 3.7); 29.7% for the accommodation division (figure 3.4). Likewise, digit five is known in accounting as registering levels higher than its expected (theoretical) value of 7.9% (Nigrini, 1999; Nigrini, 2015). In our case, the empirical values are 8.2% for the entire section of the hospitality industry (figure 3.1) and the food and beverage division (figure 3.7), respectively 8.4% for the accommodation division.



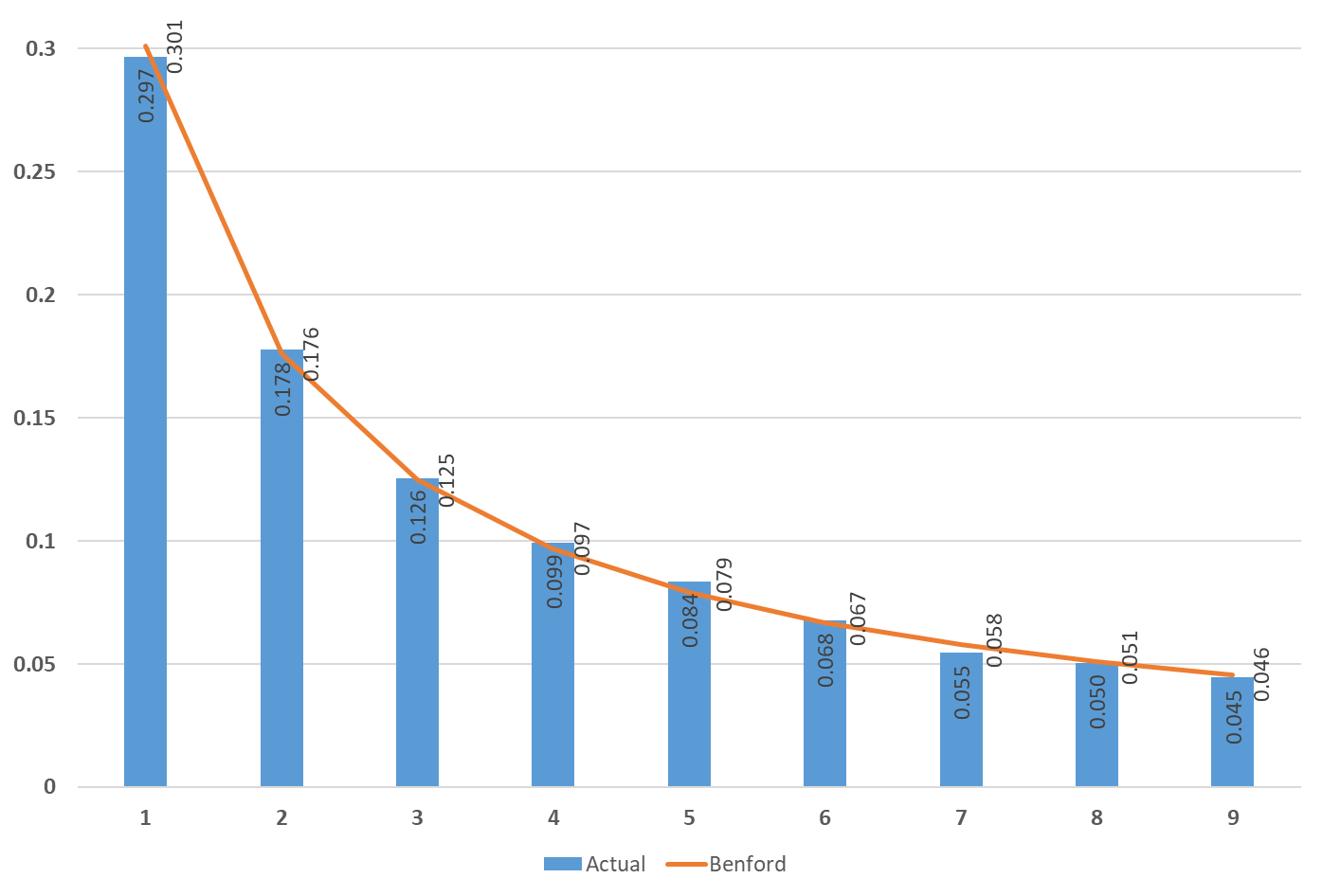
**Figure 3.1.** The distribution of the first digit (actual and expected) for the entire section of the hospitality industry



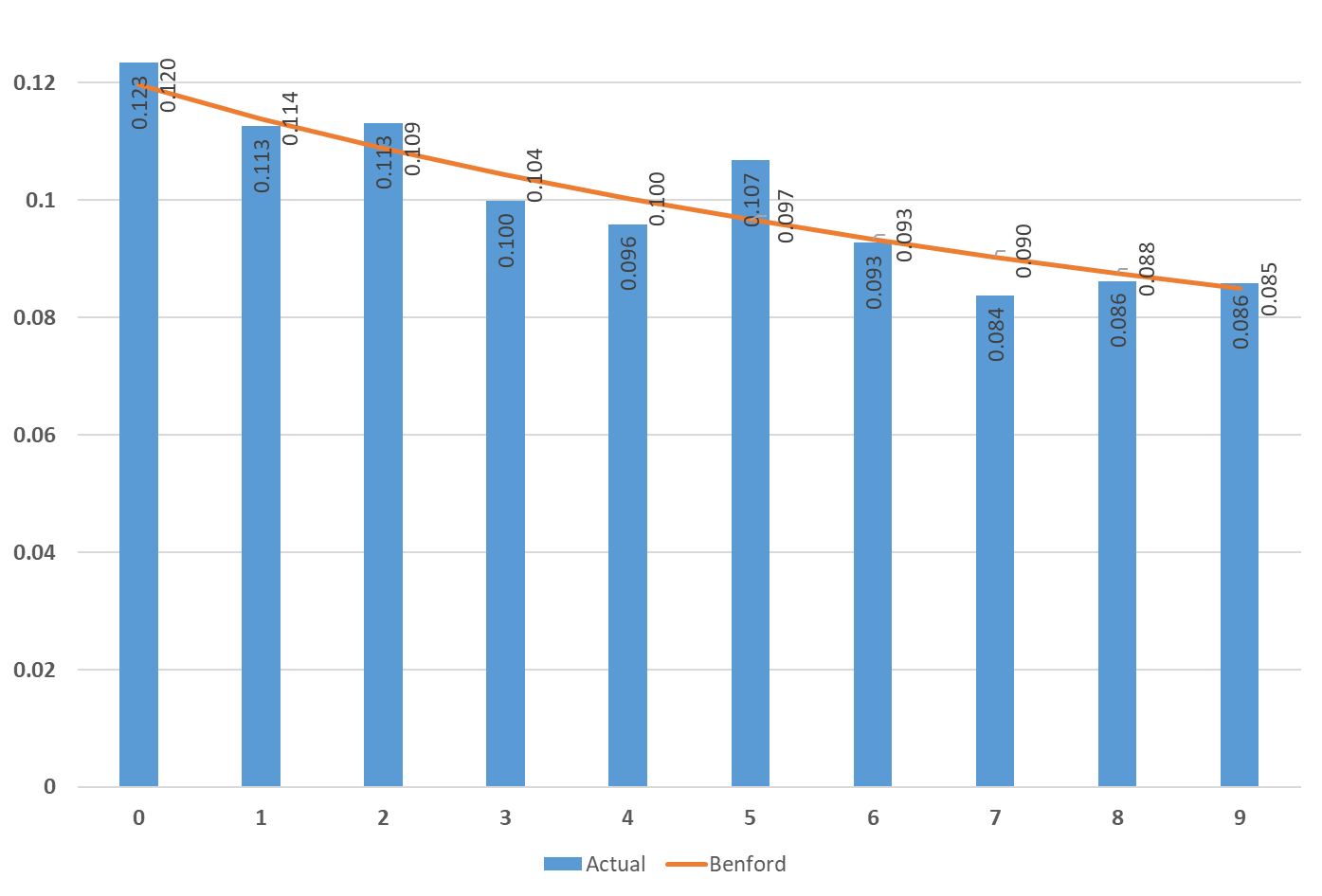
**Figure 3.2.** The distribution of the second digit (actual and expected) for the entire section of the hospitality industry



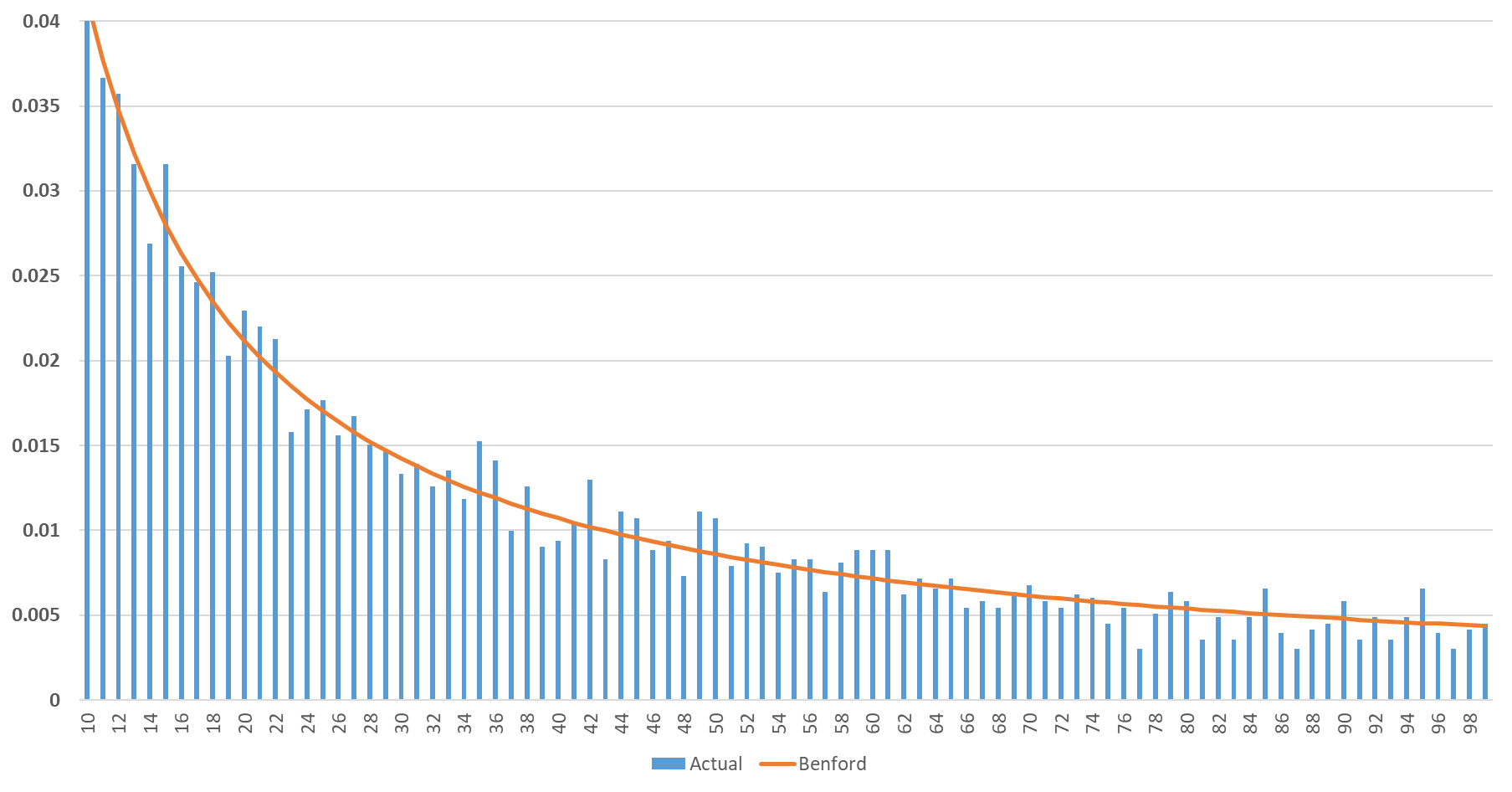
**Figure 3.3.** The distribution of the first two-digit combinations (actual and expected) for the entire section of the hospitality industry



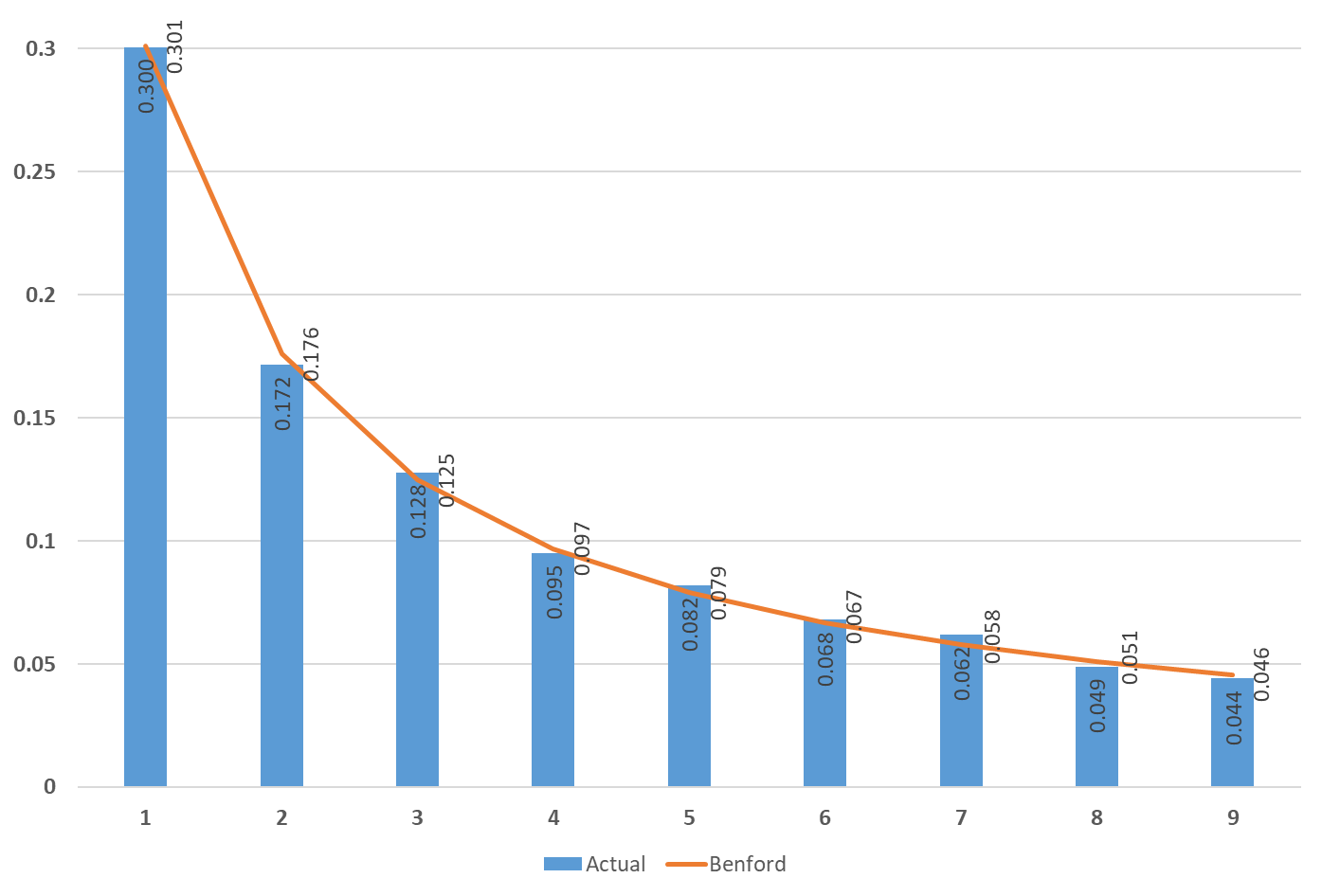
**Figure 3.4.** The distribution of the first digit (actual and expected) for the accommodation division



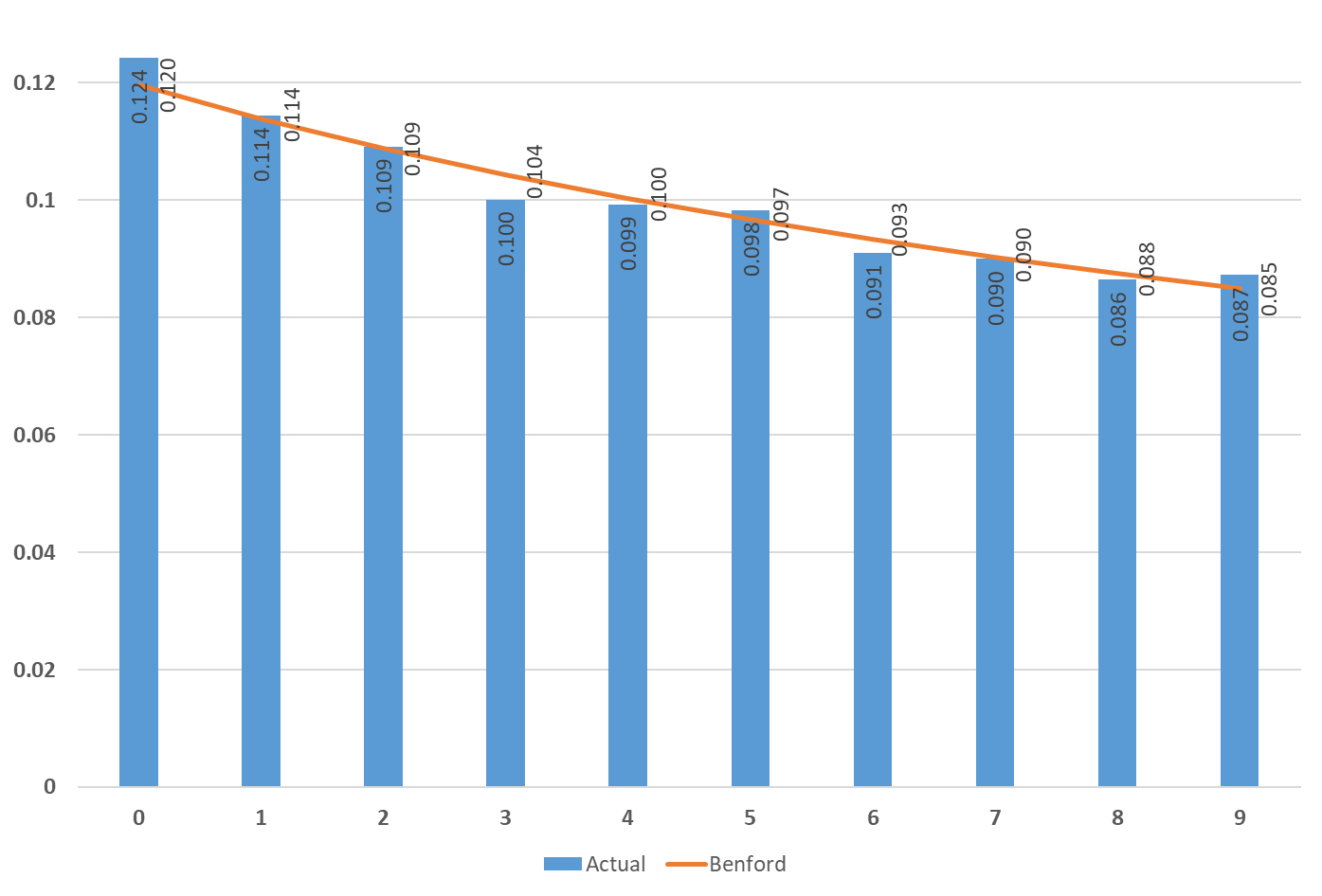
**Figure 3.5.** The distribution of the second digit (actual and expected) for the accommodation division



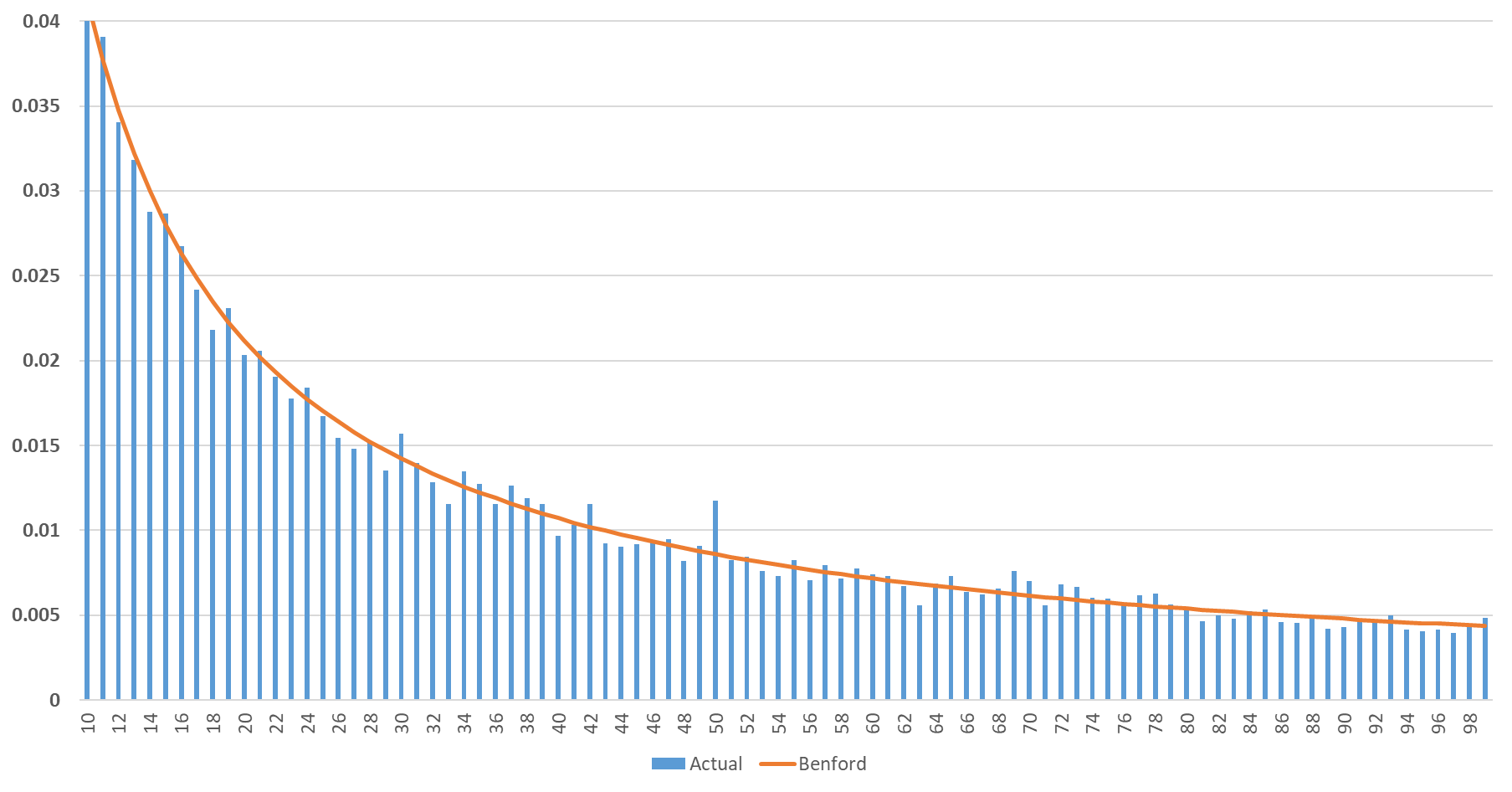
**Figure 3.6.** The distribution of the first two-digit combinations (actual and expected) for the entire section of the accommodation division



**Figure 3.7.** The distribution of the first digit (actual and expected) for the food and beverage division



**Figure 3.8.** The distribution of the second digit (actual and expected) for the food and beverage division



**Figure 3.9.** The distribution of the first two-digit combinations (actual and expected) for the entire section of the food and beverage division

Regarding the second digit, it reveals that there was a higher frequency than expected (12.0%) in the case of the digit 0, respectively 9.7% in the case of the digit 5. For the entire section of the hospitality industry (figure 3.2.), like for the accommodation division (figure 3.5.) the value calculated for the digit 0 was 12.3%, while the food and beverage division reached 12.4% (figure 3.8). For digit five, there were established levels of 10.0% (hospitality industry section - figure 3.2.), 10.7% (accommodation division - figure 3.5.) and 9.8% (food and beverage division - figure 3.8.).

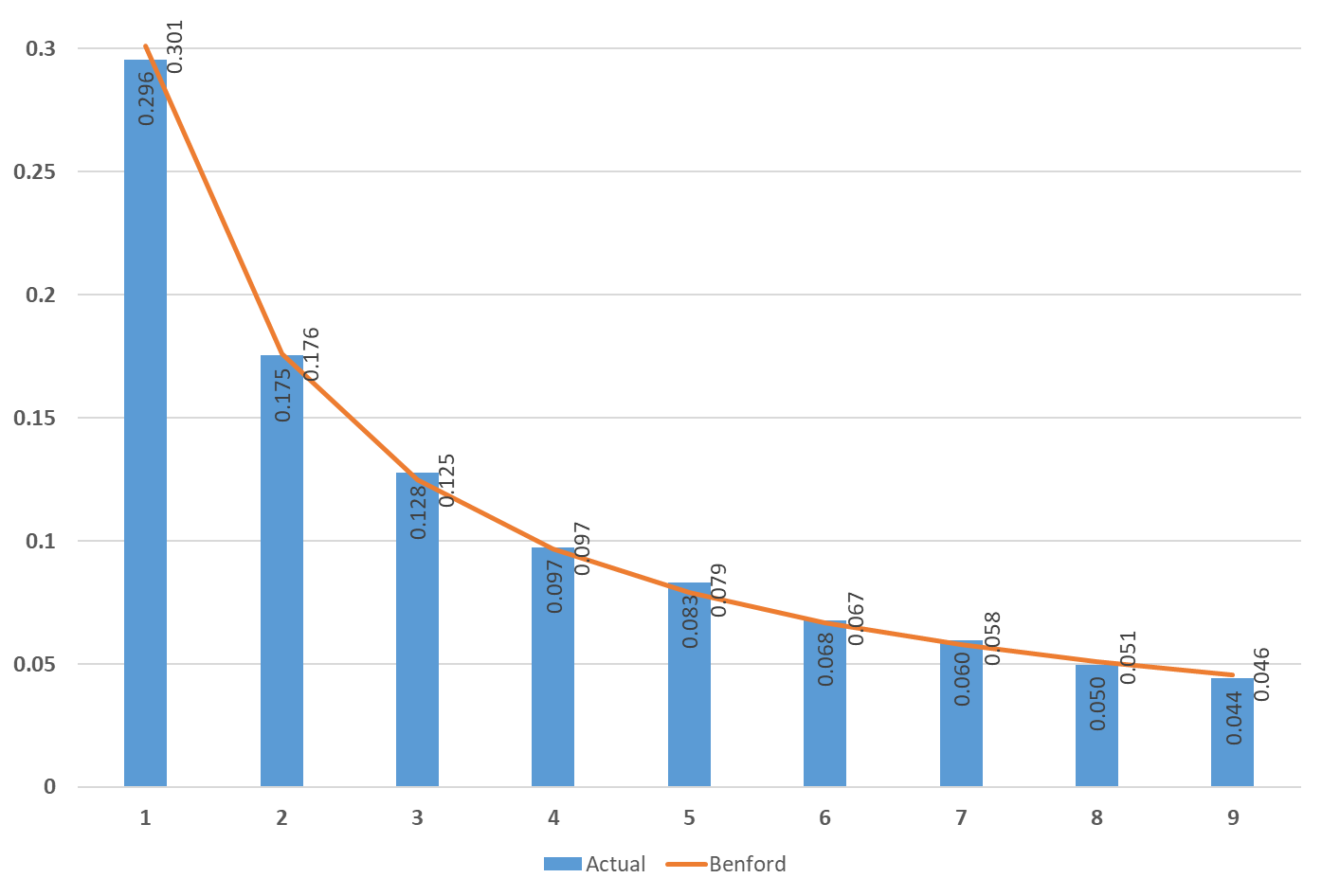
If a data set conforms to the distribution of Benford’s Laws for the first and second digit processed separately, it is a good evidence that the combination of the two digits follows the same distribution.

Furthermore, any dissimilarities against Benford’s Laws in the case of the entire hospitality industry section (Figure 3.3) or the food and beverage division (Figure 3.9) are almost absent. Similar to other studies’ results (Nigrini, 1999; Nigrini 2015), there are only two very small spikes for leading digits 50, whose theoretical frequency (Benford) is 0.86%, but the empirical level is 1.15% for the whole section of the hospitality industry (figure 3.3.) and respectively 1.18% for the food and beverage division (figure 3.9.). The dissimilarities are somewhat more obvious (albeit at a small level) regarding the accommodation division (figure 3.6.). According to other studies (Nigrini, 2015), the spikes appear mainly in the case of combinations ending in 0 or 5 (15, 35, 50, 60, 85 or 95).

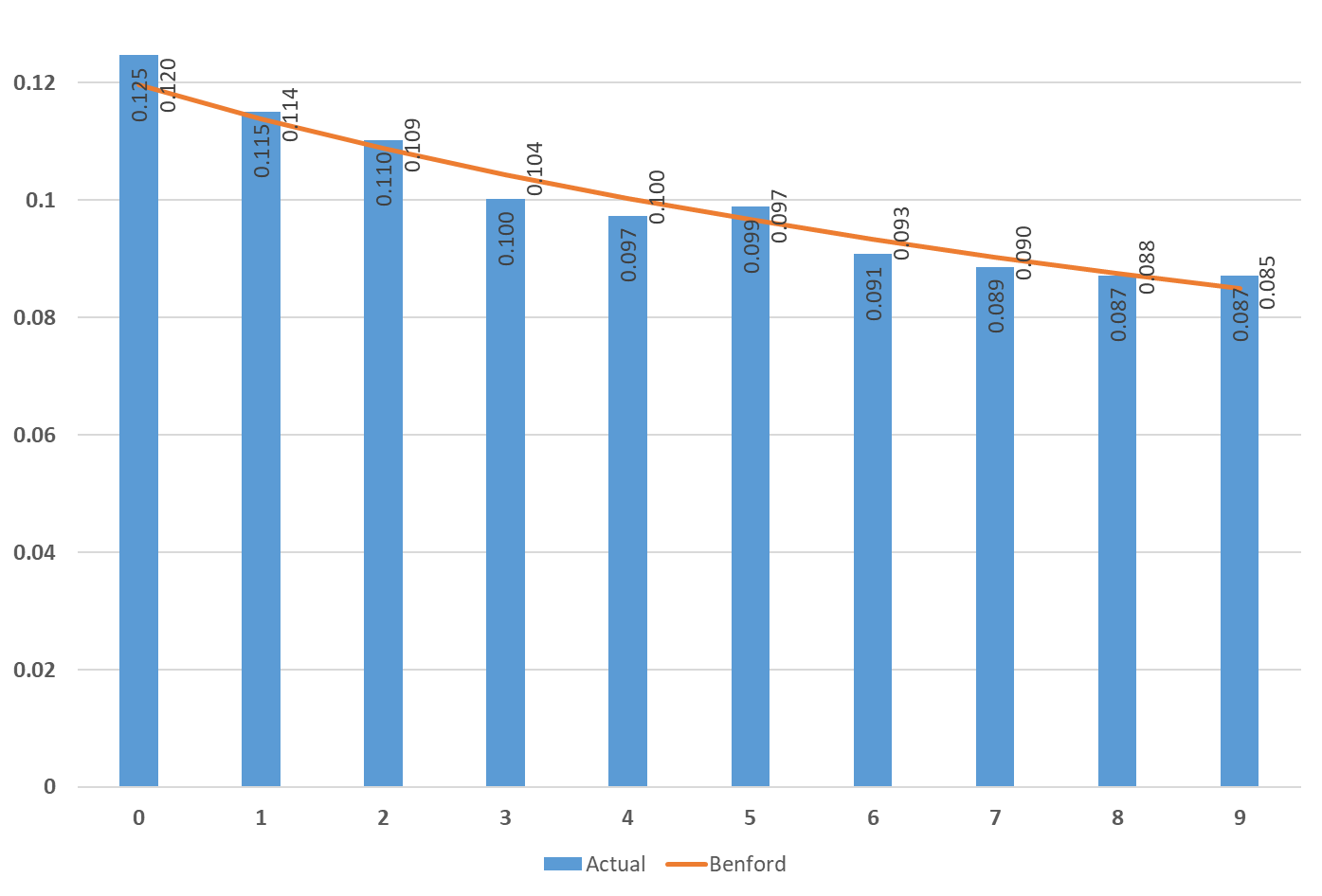
**3.2. The visual approach for the samples obtained according to company size**

In micro companies, the visual approach of the empirical distribution compared with Benford’s Laws shows very few variations. For the first digit (figure 3.10) and also for the second one (figure 3.11.) the dissimilarities are hardly perceptible (most often at the third decimal place). As in the previous section (3.1.), in the case of the second digit, the empirical level for 0 (12.5%) and 5 (9.9%) is a few tenths higher than the expected one. Also, if one looks at Figure 3.2. for the first two digits combined, the only obvious deviation is in the case of numbers starting with 50 (where the empirical level is 1.15% compared to the expected value of 0.86%).

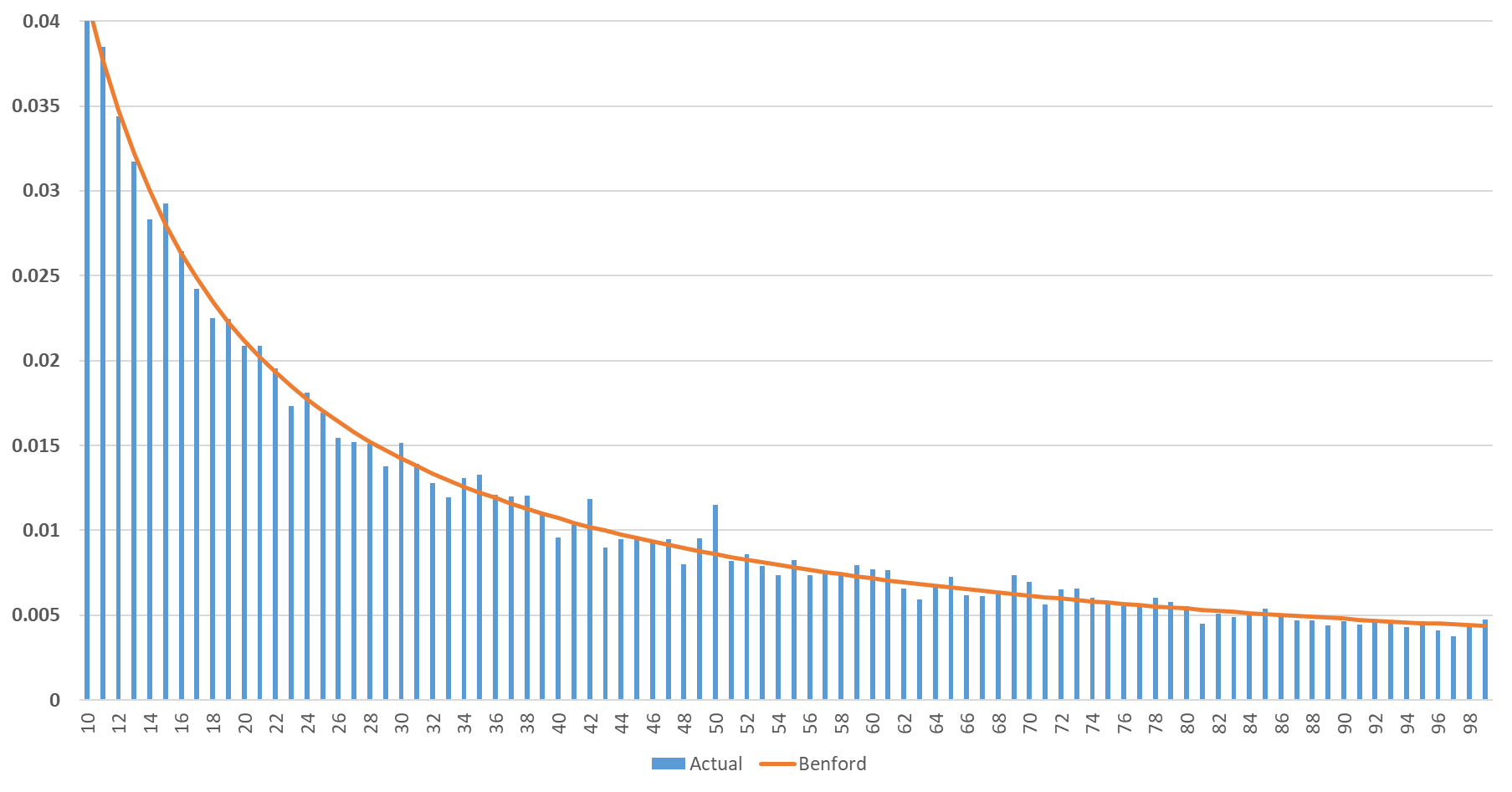
Moreover, for smaller volume samples (slightly below 3,000 small companies and even 250 medium-sized companies), increased visually identified volatility is found. Regarding the first digit, for small companies (figure 3.13.), there are dissimilarities greater than or equal to one percentage point for the first time. Thus, for digit one, the empirical level is 32.9% compared to the expected value (30.1%), for digit two, the empirical level (15.9%) is lower than the expected one (17.6%), and in the case of digit four, we find the empirical level of 8.7% (compared to the expected level of 9.7%). The dissimilarities are quite important, and by aggregation, as a consequence, the null hypothesis is eventually rejected (the distribution is not Benford’s Law conformant).

****

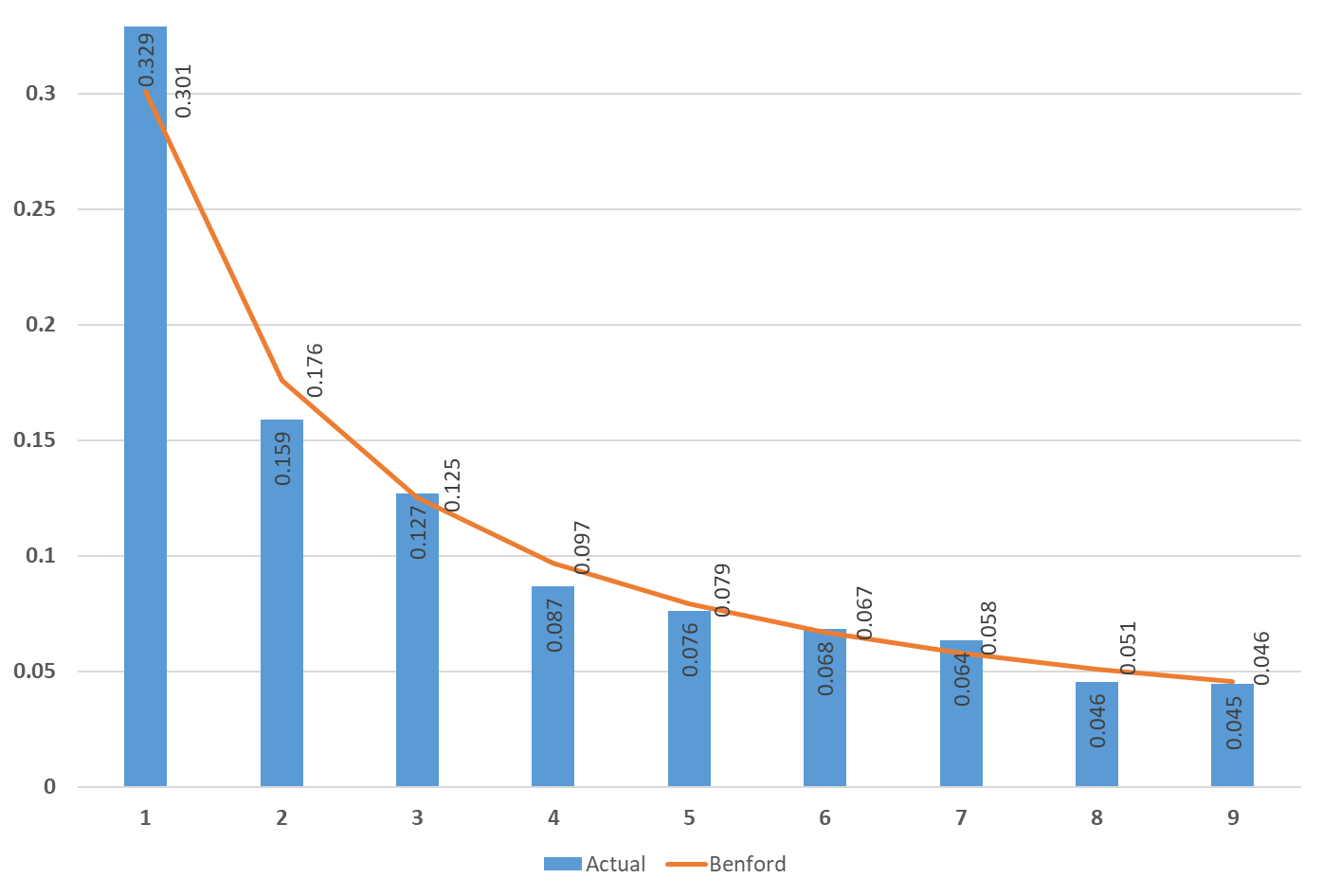
**Figure 3.10.** The distribution of the first digit (actual and expected) for the micro-companies from the entire section of the hospitality industry



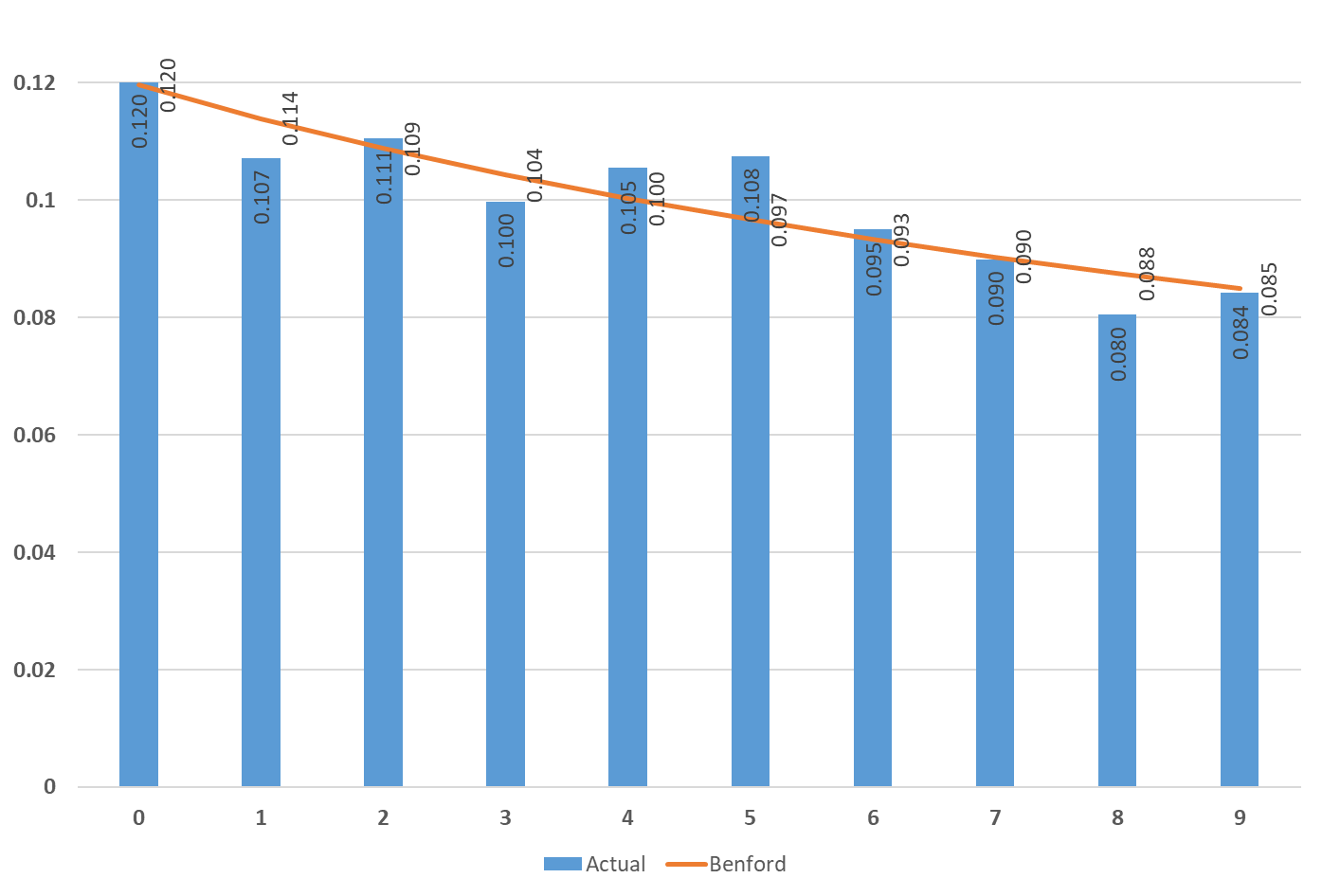
**Figure 3.11.** The distribution of the second digit (actual and expected) for the micro-companies from the entire section of the hospitality industry



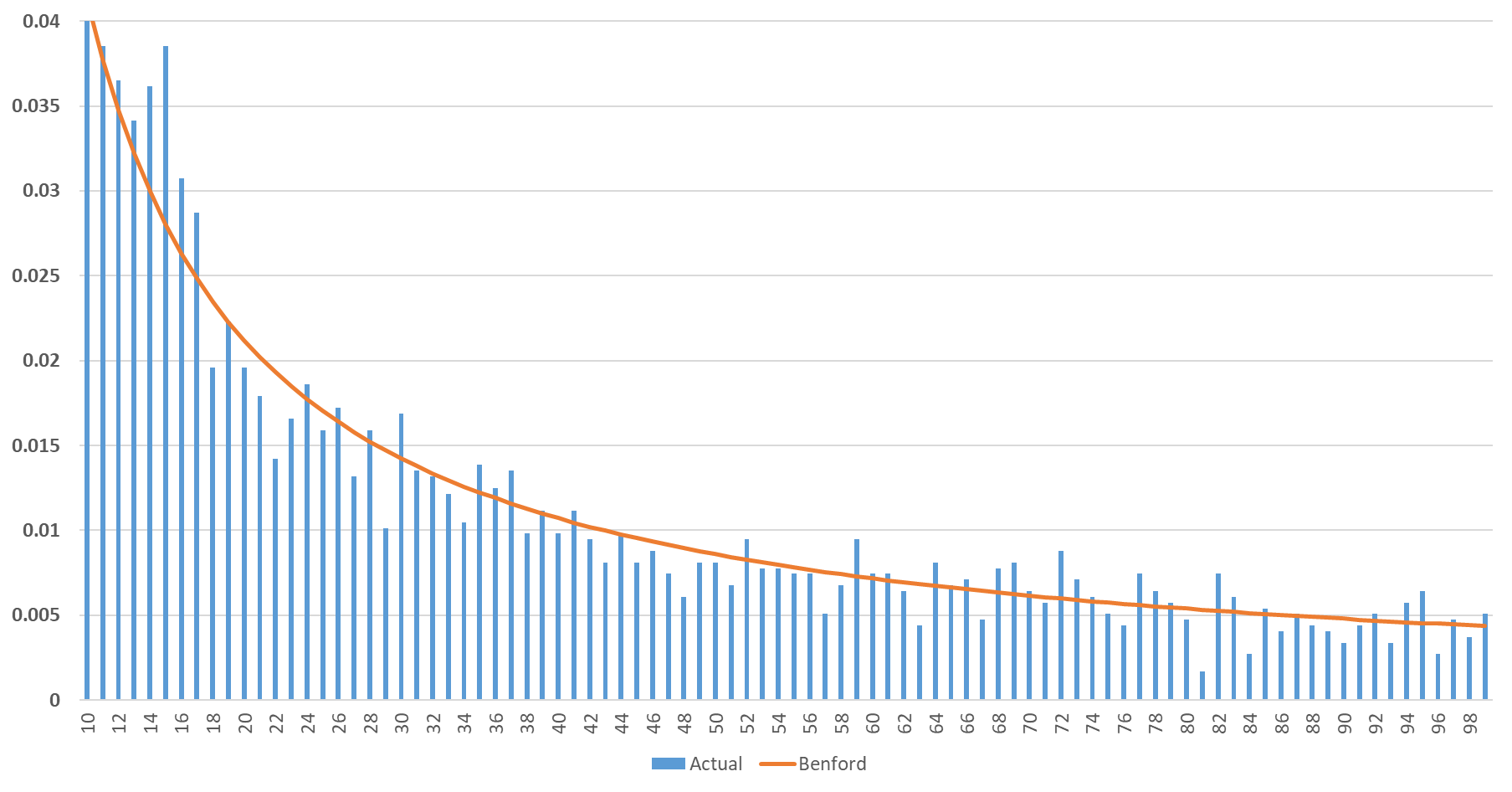
**Figure 3.12.** The distribution of the first two-digit combinations (actual and expected) for the micro-companies from the entire section of the hospitality industry



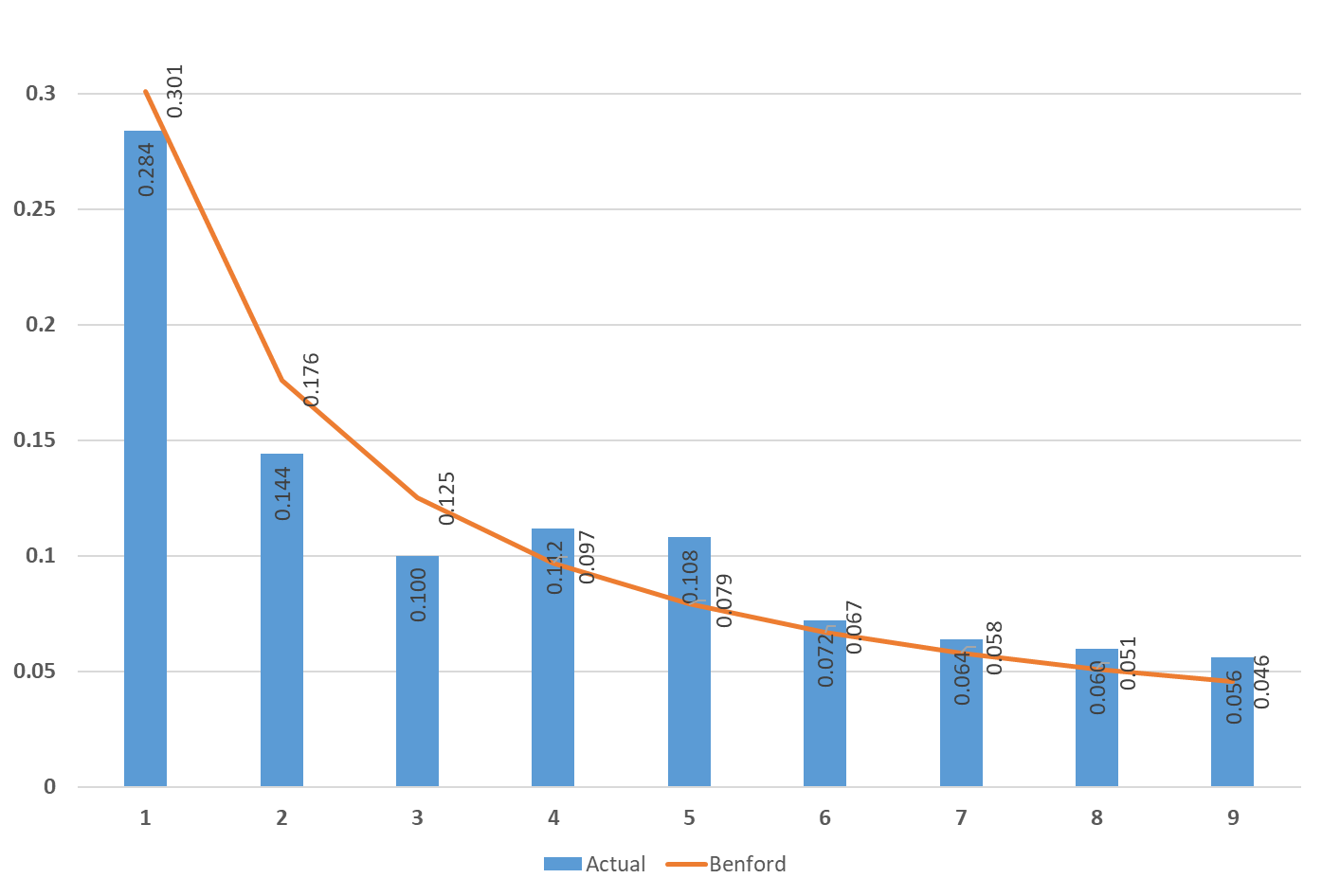
**Figure 3.13.** The distribution of the first digit (actual and expected) for the small companies from the entire section of the hospitality industry



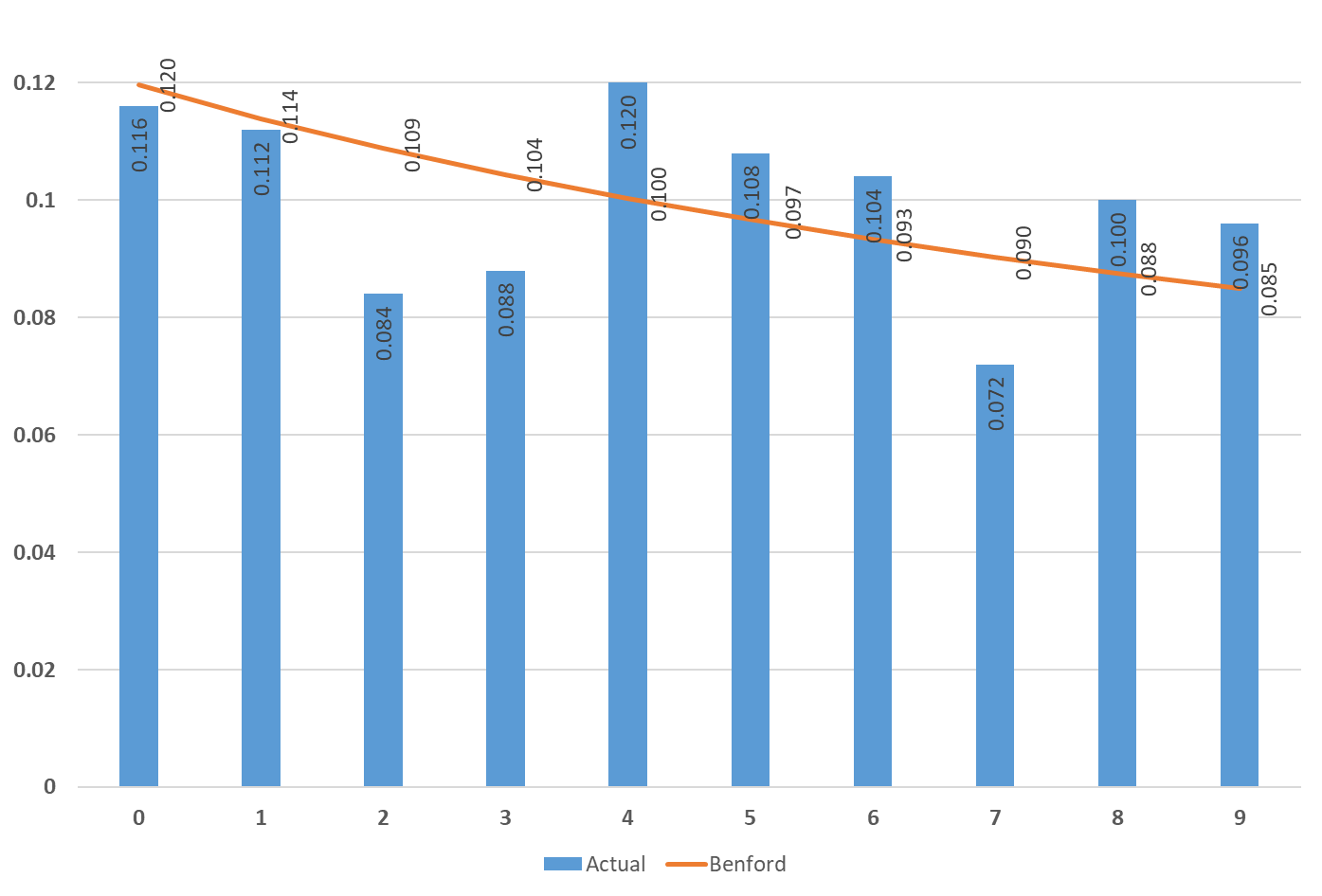
**Figure 3.14.** The distribution of the second digit (actual and expected) for the small companies from the entire section of the hospitality industry



**Figure 3.15.** The distribution of the first two-digit combinations (actual and expected) for the small companies from the entire section of the hospitality industry



**Figure 3.16.** The distribution of the first digit (actual and expected) for the medium-sized companies from the entire section of the hospitality industry



**Figure 3.17.** The distribution of the second digit (actual and expected) for the medium-sized companies from the entire section of the hospitality industry

There are techniques (D’Ambra *et al.*, 2018) in which the decomposition of the elements of the cumulative Chi-Square statistics can be performed or analysis dedicated to small companies (Ceptureanu *et al.*, 2017). However, due to the fact that the computed Chi-Square statistics (19.38) is quite close to the critical value (15.51), the potential incorporation of special teqniques will not add so much of an additional explanatory power. Hence, the decomposition is not performed. Concerning the second digit (figure 3.14.) there are disparities from the expected levels, but the deviations are not so important: for digit five, the empirical level is 10.8%, compared to the expected level of 9.7%; for digit eight, the empirical level is 0.8%, compared to the expected level of 0.88%; for digit one the empirical level is 10.7%, below 11.4% the expected level. Therefore, it is somewhat surprising that if one looks at the combination of the first two digits (figure 3.15.), we notice only a few spikes, with slightly more significant dissimilarities for the leading digits 14, 15, 16 and 17, while the other combinations have empirical levels close to expected.

The visually identified volatility is increasing when one studies the case of medium-sized companies (figures 3.16. and 3.17.). Both, for the first and the second digit, there are frequent differences (plus or minus) of two percentage points. The maximum difference is recorded in the first digit two (figure 3.16.) where the empirical level is 14.4% compared to 17.6% as expected. It seems slightly surprising that the null hypothesis (conformity with Benford’s Law) is not rejected in this situation.

**3.3. Analytical methods for testing (non)conformity against Benford’s Laws**

The visual interpretation of the distributions led, theoretically, towards establishing the analytical methods for testing (non)conformity against the theoretical distributions. Hence, this is the context in which we dedicate the present section to implementing the methodology set out in section 2.3. As a manner of presentation, it will proceed as follows: (i) each sample will be examined in separate tables (Tables 3.1.-3.6.); (ii) the results for the first digit, second digit and first two-digit combinations are presented separately; (iii) the items regarding the graphical representation to which the Chi-square test refers, the calculated value of the Chi-square test; the critical value; the degrees of freedom and the sample size, are pointed out to provide an overview on the statistical validity[[5]](#endnote-5) of the outcomes; (iv) in addition, the colour code is used to signal two situations: the case of rejecting the null hypothesis (conforms to Benford’s Laws) have been highlighted in grey, while areas with a light golden background indicate some persistent visual identified volatility (even if this volatility is not big enough to lead to Benford’s Laws non-conformity).

**Table 3.1.** Summary results for the Chi-square testing of (non)conformity against Benford’s Laws – entire section of the hospitality industry

|  |  |  |  |
| --- | --- | --- | --- |
| **Indicator** | **First digit** | **Second digit** | **First two combined digits** |
| Figure | Figure 3.1. | Figure 3.2. | Figure 3.3. |
| Chi-square test | 12.32 | 15.43 | 94.71 |
| Critical value (significance level 0.05) | 15.51 | 16.92 | 112.02 |
| Degrees of freedom | 8 | 9 | 89 |
| Sample size | 24,037 | 23,975 | 23,975 |

**Data source:** Authors’ own calculations

**Table 3.2.** Summary results for the Chi-square testing of (non)conformity against Benford’s Laws – accommodation division

|  |  |  |  |
| --- | --- | --- | --- |
| **Indicator** | **First digit** | **Second digit** | **First two combined digits** |
| Figure | Figure 3.4. | Figure 3.5. | Figure 3.6. |
| Chi-square test | 3.47 | 12.05 | 86.66 |
| Critical value (significance level 0.05) | 15.51 | 16.92 | 112.02 |
| Degrees of freedom | 8 | 9 | 89 |
| Sample size | 5,370 | 5,349 | 5,318 |

**Data source:** Authors’ own calculations

**Note:** Light golden background highlights that there is some persistent visual identified volatility.

**Table 3.3.** Summary results for the Chi-square testing of (non)conformity against Benford’s Laws – food and beverage division

|  |  |  |  |
| --- | --- | --- | --- |
| **Indicator** | **First digit** | **Second digit** | **First two combined digits** |
| Figure | Figure 3.7. | Figure 3.8. | Figure 3.9. |
| Chi-square test | 14.11 | 9.76 | 91.12 |
| Critical value (significance level 0.05) | 15.51 | 16.92 | 112.02 |
| Degrees of freedom | 8 | 9 | 89 |
| Sample size | 18,667 | 18,626 | 18,626 |

**Data source:** Authors’ own calculations

**Table 3.4.** Summary results for the Chi-square testing of (non)conformity against Benford’s Laws, entire section of the hospitality industry – micro companies

|  |  |  |  |
| --- | --- | --- | --- |
| **Indicator** | **First digit** | **Second digit** | **First two combined digits** |
| Figure | Figure 3.10. | Figure 3.11. | Figure 3.12. |
| Chi-square test | 10.75 | 14.67 | 98.38 |
| Critical value (significance level 0.05) | 15.51 | 16.92 | 112.02 |
| Degrees of freedom | 8 | 9 | 89 |
| Sample size | 20,805 | 20,745 | 20,745 |

**Data source:** Authors’ own calculations

**Table 3.5.** Summary results for the Chi-square testing of (non)conformity against Benford’s Laws, an entire section of the hospitality industry – small companies

|  |  |  |  |
| --- | --- | --- | --- |
| **Indicator** | **First digit** | **Second digit** | **First two combined digits** |
| Figure | Figure 3.13. | Figure 3.14. | Figure 3.15. |
| Chi-square test | 19.38 | 8.05 | 90.83 |
| Critical value (significance level 0.05) | 15.51 | 16.92 | 112.02 |
| Degrees of freedom | 8 | 9 | 89 |
| Sample size | 2,960 | 2,958 | 2,958 |

**Data source:** Authors’ own calculations

**Note:** Light golden background highlights that there is some persistent visual identified volatility. Light grey background denotes Benford’s Law non-conformity.

**Table 3.6.** Summary results for the Chi-square testing of (non)conformity against Benford’s Laws, an entire section of the hospitality industry – medium-sized companies

|  |  |  |
| --- | --- | --- |
| **Indicator** | **First digit** | **Second digit** |
| Figure | Figure 3.16. | Figure 3.17. |
| Chi-square test | 7.36 | 5.42 |
| Critical value (significance level 0.05) | 15.51 | 16.92 |
| Degrees of freedom | 8 | 9 |
| Sample size | 250 | 250 |

**Data source:** Authors’ own calculations

**Note:** Light golden background highlights that there is some persistent visual identified volatility.

Unexpectedly, testing with analytical methods leads to the conclusion that the disclosed financial statements of the Romanian companies, which are operating in the hospitality industry, conforms to Benford’s Laws. There is only one exception: the first digit for small companies from the entire section of the hospitality industry. In this case, the empirical value of the Chi-square test (19.38) slightly exceeds the critical value (15.51).

Other persistent volatilities, identified and discussed in the previous section, do not affect conformity with Benford’s Laws. In all cases (Figures 3.6., 3.14., 3.15., 3.16. and 3.17.), the null hypothesis (actual (empirical) distributions conform to Benford’s Laws) is accepted. The financial statements of the companies operating in the hospitality industry obey Benford’s Laws to a greater extent.

**4. Conclusions**

To the best of our knowledge, this research paper is the first dedicated to analysing compliance with Benford’s Laws of financial statements of Romanian companies. Although Benford’s Laws has been tested on data specific to Romania, either demographic (Ausloos *et al.*, 2015) or related to the distribution of bank deposits (Druica *et al.*, 2018), the financial statements data have not been analysed so far. An important point regarding the validity of this approach is the use of an almost exhaustive dataset[[6]](#endnote-6). Another relatively unique feature is that the methodology of this paper aimed to test both the first and second and the combination of the first two digits for the variable of interest (the result of the economic activity). Such approaches are relatively rare in the literature.

The context in which it was intended to test the hypothesis of (non) compliance of financial statements with Benford’s Laws included two factors that could have largely had negative effects in this direction. This is about the high level of tax evasion doubled by the propensity to use cash in the hospitality industry (especially regarding the food and beverage division).

Whether the first or second digits are analysed separately or, on the contrary, the combination of the first two digits is chosen, with one exception, all samples follow (conform to) Benford’s Laws. Even when the null (compliance) hypothesis is rejected, the value of the square Chi () statistic is relatively close to the critical value. From the point of view of the samples, produced according to the NACE code (division 55: accommodation; division 56: food and beverage and respectively the two divisions consolidated in the hospitality industry) there is no difference regarding the compliance with Benford’s Laws. The sample that led to the rejection of the null hypothesis was created according to the size criterion of the companies.

The tax evasion combined with a propensity for cash payments from the hospitality industry seems not to adversely affect compliance with Benford’s Laws. These particularities of Romania that, in fact, do not affect the results are located at the same level as the records provided by other specialised papers (Mir *et al.*, 2014; Shi *et al.*, 2018). The results in detail (including the specific figures that deviated somewhat more significantly from the expected levels) are in line with other thematic studies (Nigrini, 1999; Nigrini, 2015).

From the perspective of future research directions, one can consider that a longitudinal analysis is desirable (in the sense of validating or not this static view on the year 2013). Also, the analysis of other sections of NACE, or the Romanian economy as a whole, is a wish to prove useful and interesting information for the academic and/ or decision-making environment. One can conclude that non-compliance with Benford’s Laws does not, by default, mean data manipulation. It may be a clue but not necessarily proof. The fact that, in the present paper, contrary to the initial expectations, we find that the financial statements of the companies in the hospitality industry are very much in compliance with Benford’s Laws is of interest for the academia as well as for the general audience.

**Acknowledgements**

This work is partially supported by a grant from the Romanian National Authority for Scientific Research and Innovation, CNDS-UEFISCDI, project number PN-III-P4-ID-PCCF-2016-0084. Many thanks to (alphabetical order): Tudorel Andrei, Marcel Ausloos, Gabriel Badescu, Jozsef Benedek, Roy Cerqueti, Sebastian Ceptureanu, Adriana Dardala, Daniel David, Bogdan Vasile Ileanu, Iulia Jianu, Cristian Litan, Zoltan Neda, Adrian Pana, Pasquale Sarnacchiaro, Biagio Simonetti and Denisa Vasilescu for their suggestions and comments on an initial version of the manuscript. Thanks to Gurjeet Dhesi and Valerio Ficcadenti (London South Bank University, UK) as well to Elena-Diana Jianu (student at the National College “Şcoala Centrală” of Bucharest, Romania) who checked the English on the later version of the manuscript.

**References**

1. Alexopoulos, T., & Leontsinis, S. (2014). Benford’s Laws in astronomy. *Journal of Astrophysics and Astronomy*, *35*(4), 639-648.
2. Alves, A. D., Yanasse, H. H., & Soma, N. Y. (2016). An analysis of bibliometric indicators to JCR according to Benford’s Laws. *Scientometrics*, *107*(3), 1489-1499.
3. Androniceanu, A., Gherghina, R., & Ciobanasu, M. (2019). The interdependence between fiscal public policies and tax evasion. *Administratie si Management Public*, (32), 32-41.
4. AS, (2019). Statistical Inference in the 21st Century: A World Beyond p < 0.05, *The American Statistician, available online:* https://www.tandfonline.com/toc/utas20/73/sup1?nav=tocList&fbclid=IwAR35wRQV9VRYY5dGjciO2wr55ghuZ0poiVlJfEAn2yhYOFayCFtEbVUMuWw
5. ASR (2014). Anuarul Statistic al României. *Institutul Național de Statistică,* Bucharest.
6. Ausloos, M., Castellano, R., & Cerqueti, R. (2016). Regularities and discrepancies of credit default swaps: a data science approach through Benford’s Laws. *Chaos, Solitons & Fractals*, *90*, 8-17.
7. Ausloos, M., Cerqueti, R., & Lupi, C. (2017). Long-range properties and data validity for hydrogeological time series: The case of the Paglia river. *Physica A: Statistical Mechanics and its Applications*, *470*, 39-50.
8. Ausloos, M., Herteliu, C., & Ileanu, B. (2015). Breakdown of Benford’s Laws for birth data. *Physica A: Statistical Mechanics and its Applications*, *419*, 736-745.
9. Ausloos, M., Ficcadenti, V., Dhesi, G., & Shakeel, M. (2021). Benford’s laws tests on S&P500 daily closing values and the corresponding daily log-returns both point to huge non-conformity. *Physica A: Statistical Mechanics and its Applications*, *574*, 125969.
10. Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American philosophical society*, 551-572.
11. Brezeanu, P., Celea, S., & Stanciu, A. P. (2011). Forms of tax evasion in Romania. Analytical perspective. *Annals of the University of Petroşani. Economics*, *11*, 33-42.
12. Brown, D. O. (1998). In search of an appropriate form of tourism for Africa: lessons from the past and suggestions for the future. *Tourism Management*, *19*(3), 237-245.
13. Bucur, A., Dobrotă, G., & Dumitraşcu, O. (2019). Implications of Fiscal Pressure on the Sustainability of the Equilibrium and Performance of Companies. Evidences in the Rubber and Plastic Industry from Romania. *Sustainability*, *11*(7), 2082.
14. Canagarajah, S., Brownbridge, M., Paliu, A., & Dumitru, I. (2012). *The challenges to long run fiscal sustainability in Romania*. The World Bank.
15. Ceptureanu, S. I., Ceptureanu, E. G., & Visileanu, E. (2017). Comparative analysis of small and medium enterprises organizational performance in clothing industry. *Industria Textila*, *68*(2), 156.
16. Clippe, P., & Ausloos, M. (2012). Benford’s Laws and Theil transform of financial data. *Physica A: Statistical Mechanics and its Applications*, *391*(24), 6556-6567.
17. Constantin, D. L., Goschin, Z., & Danciu, A. R. (2011). The Romanian economy from transition to crisis. Retrospects and prospects. *World Journal of Social Sciences*, *1*(3), 155-171.
18. Crnogorac, M., & Lago-Peñas, S. (2018). *Tax evasion in Former Yugoslavian countries* (No. 1811). Universidade de Vigo, GEN-Governance and Economics research Network.
19. D’Ambra, L., Amenta, P., & D’Ambra, A. (2018). Decomposition of cumulative chi-squared statistics, with some new tools for their interpretation. *Statistical Methods & Applications*, *27*(2), 297-318.
20. Druica, E., Oancea, B., & Vâlsan, C. (2018). Benford’s Laws and the limits of digit analysis. *International Journal of Accounting Information Systems*, *31*, 75-82.
21. Durtschi, C., Hillison, W., & Pacini, C. (2004). The effective use of Benford’s Laws to assist in detecting fraud in accounting data. *Journal of forensic accounting*, *5*(1), 17-34.
22. Eurostat (2008). N.A.C.E. Rev. 2–statistical classification of economic activities in the European community. *Office for Official Publications of the European Communities,* Luxemburg.
23. Gini, C. (1957) Sulla frequenza delle cifre iniziali dei numeri osservati. Bull. Inst. Internat. Stat., 29th session, 35(2), pp. 57-76.
24. Hasseldine, D. J., & Bebbington, K. J. (1991). Blending economic deterrence and fiscal psychology models in the design of responses to tax evasion: The New Zealand experience. *Journal of Economic Psychology*, *12*(2), 299-324.
25. Huang, X., Maçaira, P. M., Hassani, H., Oliveira, F. L. C., & Dhesi, G. (2019). Hydrological natural inflow and climate variables: Time and frequency causality analysis. *Physica A: Statistical Mechanics and its Applications*, *516*, 480-495.
26. Hürlimann, W. (2015). On the uniform random upper bound family of first significant digit distributions. *Journal of Informetrics*, *9*(2), 349-358.
27. Ileanu, B. V., Ausloos, M., Herteliu, C., & Cristescu, M. P. (2018). Intriguing behavior when testing the impact of quotation marks usage in Google search results. *Quality & Quantity*, 1-13.
28. Jaliu, D. D., & Ravar, A. S. (2019). Informal Tourism Economy and EU Funding: The Case of Romania. In *Yellow Tourism* (pp. 193-207). Springer, Cham.
29. Jianu, I., & Jianu, I. (2021). Reliability of Financial Information from the Perspective of Benford’s Law. *Entropy*, *23*(5), 557.
30. Kaiser, M. (2019). Benford’s Laws As An Indicator Of Survey Reliability—Can We Trust Our Data?. *Journal of Economic Surveys*, *33*(5), 1602-1618.
31. Kennedy, A. P., & Yam, S. C. P. (2020). On the authenticity of COVID-19 case figures. *PLOS ONE*, *15*(12), e0243123.
32. Lazar, R. E. (2013). Tax evasion between legality and criminal offense. *Procedia-Social and Behavioral Sciences*, *92*, 462-466.
33. Lee, K. B., Han, S., & Jeong, Y. (2020). COVID-19, flattening the curve, and Benford’s law. *Physica A: Statistical Mechanics and its Applications*, *559*, 125090.
34. Mir, T. A. (2014). The Benford law behavior of the religious activity data. *Physica A: Statistical Mechanics and its Applications*, *408*, 1-9.
35. Mir, T. A., & Ausloos, M. (2018). Benford’s Laws: A “sleeping beauty” sleeping in the dirty pages of logarithmic tables. *Journal of the Association for Information Science and Technology*, *69*(3), 349-358.
36. Mir, T. A., Ausloos, M., & Cerqueti, R. (2014). Benford’s Laws predicted digit distribution of aggregated income taxes: the surprising conformity of Italian cities and regions. *The European Physical Journal B*, *87*(11), 261.
37. Miranda-Zanetti, M., Delbianco, F., & Tohmé, F. (2019). Tampering with inflation data: A Benford law-based analysis of national statistics in Argentina. *Physica A: Statistical Mechanics and its Applications*, *525*, 761-770.
38. Murphy, J. (2015) Responsible sale and service of alcohol for the tourism, hospitality and retail industries, Goodfellow Publishing Ltd, Oxford: England.
39. Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. *American Journal of mathematics*, *4*(1), 39-40.
40. Nigrini, M. J. (1992). *The detection of income tax evasion through an analysis of digital distributions* (Doctoral dissertation), University of Cincinnati, OH, USA.
41. Nigrini, M. J. (1999). I’ve got your number. *Journal of accountancy*, *187*(5), 79-83.
42. Nigrini, M. J. (2015). Persistent patterns in stock returns, stock volumes, and accounting data in the US capital markets. *Journal of Accounting, Auditing & Finance*, *30*(4), 541-557.
43. Ott, K. (2002). The underground economy in Croatia. *Occasional paper series-Institute of Public Finance*, *7*(12), 1-29.
44. Poincaré, H. (1912) *Répartition des décimales dans une table numérique*. pp 313-320 in: Calcul des Probabilités, Gauthier-Villars, Paris.
45. Riccioni, J., & Cerqueti, R. (2018). Regular paths in financial markets: Investigating the Benford’s Laws. *Chaos, Solitons & Fractals*, *107*, 186-194.
46. Shi, J., Ausloos, M., & Zhu, T. (2018). Benford’s Laws first significant digit and distribution distances for testing the reliability of financial reports in developing countries. *Physica A: Statistical Mechanics and its Applications*, *492*, 878-888.
47. Stigler, G.J. (1945) The distribution of leading digits in statistical tables. University of Chicago, Regenstein Library, Special Collections, George J. Stigler Archives.
48. Toader, S. A. (2007). Overall analysis of the tax evasion phenomenon and its dynamics in Romania after 1989. *Romanian Economic and Business Review*, *2*(3), 98.
49. Todor, A. (2018). Willing to Pay? The Politics of Engendering Faith in the Post-Communist Romanian Tax System. *The Leap of Faith. The Fiscal Foundations of Successful Government in Europe and America, Oxford University Press, Oxford*.
50. Williams, C. C., & Horodnic, I. A. (2017). Regulating the sharing economy to prevent the growth of the informal sector in the hospitality industry. *International Journal of Contemporary Hospitality Management*, *29*(9), 2261-2278.
51. WWW1 (2020). http://www.benfordonline.net/list/chronological retrieved in many occasions in 2020.
52. WWW2 (2020). https://ec.europa.eu/jrc/en/event/workshop/benfords-law-conference retrieved in December 2020.
53. WWW3 (2020). https://data.gov.ro/dataset/situatii\_financiare\_2013 retrieved in many occasions in 2019-2020.
54. WWW4 (2020). https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Small\_and\_medium-sized\_enterprises\_(SMEs) retrieved in December 2020.

1. Born in Canada. [↑](#endnote-ref-1)
2. Born in South Africa. [↑](#endnote-ref-2)
3. Granted in 1982. [↑](#endnote-ref-3)
4. Romanian currency. [↑](#endnote-ref-4)
5. Following the recent debate in the academic environment, on the relevance of the concept of p value and / or "statistical significance", the preference for the use of the term "statistical validity" is given. On this topic see, for example: AS, (2019). Statistical Inference in the 21st Century: A World Beyond p < 0.05, *The American Statistician, available online:* https://www.tandfonline.com/toc/utas20/73/sup1?nav=tocList&fbclid=IwAR35wRQV9VRYY5dGjciO2wr55ghuZ0poiVlJfEAn2yhYOFayCFtEbVUMuWw [↑](#endnote-ref-5)
6. Even if the authors of the current manuscript consider that the dataset (WWW3, 2020) is exhaustive, since there are some small differences when compare to official statistics’ reports (ASR, 2014) (as is shown in Table 2.5.) the prudence in claims is the best approach. [↑](#endnote-ref-6)