

TOURISM DATA IN CROATIA ASSESSED BY BENFORD'S LAW

<https://doi.org/10.47743/jopafl-2021-19-13>

Hrvoje MATAKOVIĆ
Institute for Tourism
Zagreb, Croatia
hrvoje.matakovic@iztzg.hr

Abstract: *Tourism data is crucial for effective tourism management since it enables national and local authorities to shape public policies in tourism and also enables the tourism industry to make appropriate business decisions. In 2016 new tourism data information system, called eVisitor, was introduced in Croatia, and this new system significantly eased collection and data processing. The aim of this paper is to assess the quality of statistical data on tourist traffic and to determine whether the technical improvement of the data collection system, which eased reporting on tourist traffic to information providers, contributed to the quality of collected data. This is done by applying Benford's distribution of first digits, i.e. Benford's law, to the collected data. Benford's law is based on the thesis that the first digits in numbers are not uniformly distributed and gives an expected pattern of numbers in the tabular data. Data that is not manipulated, accidentally or intentionally, should follow Benford's distribution of first digits, and deviations from Benford's distribution indicate that the data is compromised in some way. The conducted analysis has shown that the introduction of a new user-friendly data system did not affect the quality of collected data, but that the origin of the tourists was more important: data for domestic tourists have shown a statistically significant deviation from the expected Benford's distribution, so it can be concluded that their quality is lower than the data for foreign tourists.*

Keywords: *tourism, tourism data, tourism statistics, Benford's Law, Croatia*

Introduction

Tourism is an “important economic, social and cultural factor in modern societies” (Stipetić, 1998: 121) - tourism in many countries foster the economy, and beside the fact it significantly contributes to GDP and employment, it also contributes to general social development (Volo, 2004). Tourism is especially important for countries like Croatia, which base a significant part of their incomes on tourism - according to data published by the World Travel and Tourism Council in 2019, travel and tourism directly contributed to Croatia's GDP by 25% and its share in total employment in 2019 was 25.1% (WTTC, 2020). These data indicate that tourism is one of the key economic branches in Croatia and to adequately manage tourism, it is necessary to have accurate and complete data to identify new development trends, new market opportunities, demand characteristics, and similar issues (Baldiraga & Mamula, 2012). All those and wider data and indicators should comprise statistics, i.e. tourism statistics.

Generally speaking, statistics can be defined as “a branch of applied mathematics concerned with the collection and interpretation of quantitative data and the use of systematization methods” (Baldiraga & Mamula, 2012: 56). Tourism statistics is a specific subtype of statistics and its basic aim is to “accurately quantify tourism flows” (De Cantis & Ferrante, 2013). Tourism statistics can have different meanings, depending on who uses it as a source of information: 1) it enables national and local authorities to monitor and

quantify the economic, socio-cultural, and environmental effects of tourism, 2) it enables planners to plan and make decisions related to tourism development, such as the adoption of tourism development strategies and policies, monitoring of tourism development and monitoring of tourism sustainability; 3) it provides decision-makers from the tourism industry with data necessary to implement market analysis, strategic planning, investments, and promotion design, 4) it provides international statistical agencies and researchers with internationally comparable data, thus enabling the creation of new knowledge (Wöber, 2000; Volo & Giambalvo, 2008; Aroca, Brida, & Volo, 2013; De Cantis, Parroco, Ferrante, & Vaccina, 2015; Krajnović, Jeličić, & Šćiran-Rizner, 2020).

The system of tourism statistics encompasses basic tourism statistics, i.e. statistics of tourist consumption and tourist offer, but also production factors used in tourism and other related activities (Ivandić & Marušić, 2009). The study of statistical measurement in tourism began in the 1930s (Wöber, 2000) and the systematic collection of data on tourism in most European countries began after World War II. The first systematic collection of travel data was published in 1949 by the British Travel Association (Volo, 2020). In Croatia, data on tourism have been systematically collected since 1954, and from those early beginnings, the system of data on tourism and travel has been trying to harmonize with modern needs and trends.

The main source of official statistical data in Croatia, including data on tourism or data related to tourism is the Croatian Bureau of Statistics. The Croatian Bureau of Statistics publishes data about tourism on a monthly and annual basis and is the most comprehensive and reliable source of statistical data in Croatia. Until 2016, the basic way to collect data on tourist traffic in commercial accommodation (hotels, apartments, summer houses, camping sites, etc.) was to fill in the Monthly Report on Tourists and Tourist Nights (so called TU-11 form). Reporting was carried out at the level of the accommodation units, and in this form, each accommodation unit provided data on the number of tourist arrivals and overnight stays, the mode of tourist arrival, and for foreign tourists it was also stated the country of their origin. Reports for private rooms, suites and houses directly rented by citizens/households were filled out by tourist boards. These reports were usually made on the basis of reception records in guest books and check-ins and checkouts in tourist boards (Kelebeh & Javor, 1998). This form was filled out on a monthly basis, in hard copy or in the excel spreadsheet form, and was submitted to the Croatian Bureau of Statistics. Several authors (Vlahović, 1998; Ivandić & Marušić, 2009; Baldiraga & Mamula, 2012) pointed out that this system should be informatically modernized, and that users should be able to enter data more easily, and in this way, more accurate data could be collected. The issue of a user-friendly ICT system for the collection of tourist data is not a specifically Croatian problem - similar problems were pointed out in Italy (Volo & Giambalvo, 2008) and in Chile (Aroca et al., 2013).

This outdated and inappropriate way of data collection was replaced in 2016 by a new information system called eVisitor. The eVisitor system is based on an Internet interface into which accommodation providers enter data about tourists. eVisitor has several purposes and only one of them is purely statistical: this system allows check-in and check-out of guests by accommodation providers via the Internet, registration of foreign tourists to the Ministry of Interior, and it is also used to calculate and control the collection of sojourn tax. The system can also be used for data processing and analysis and reporting for statistical purposes - via eVisitor it is possible to monitor tourists' movements almost

in real-time, according to one or more criteria such as type of accommodation, location, country or place of residence, gender, age of tourists, duration of stay and similar (Krajnović et al., 2020). The data entered into this system are also used by the Croatian Bureau of Statistics for the preparation of reports and analyses. This method of data collection is much more user friendly than the previously used data collection via the Monthly Report on Tourists and Tourist Nights (TU-11 form), and data collection has been significantly simplified and facilitated, so it can be expected that data collected via eVisitor should be more accurate and complete than previously collected data.

The aim of this paper is to assess the quality of statistical data on tourist traffic and to determine whether the technical improvement of the data collection system, which eased reporting on tourist traffic to information providers, contributed to the quality of collected data. In this paper, by applying Benford's distribution of first digits, i.e. Benford's law, which will be explained in detail in section 1, to the collected data, it will be determined whether there are differences in the quality of data collected through eVisitor and data previously collected through Monthly Report on Tourists and Tourist Nights. Data that is not manipulated, accidentally or intentionally, should follow Benford's distribution of first digits, and deviations from Benford's distribution indicate that the data is compromised in some way.

Benford's law

In this paper, tourism traffic data will be screened using Benford's law to identify potential data manipulations, which may be intentional or accidental. Benford's law is based on the thesis that the first digits in numbers are not uniformly distributed as might be expected, and gives expected pattern of numbers in the tabular data (Hill, 1995). Existence of a certain regularity in the appearance of the first digits in numbers was for the first time noticed by the mathematician and astronomer Simon Newcomb: Newcomb noticed that the first few pages of logarithmic tables were more worn than later pages, and he assumed that people viewed numbers starting with the first digit 1, 2, or 3 more than numbers starting with digits 7, 8, or 9. Based on that observation, he concluded that in the data sets there exist more lower numbers than higher and that, consequently, the first digits in the numbers will be more often lower numbers than higher. Newcomb published his observations in 1881 in the article "Note of the frequency of the use of different digits in natural numbers" (Newcomb, 1881), however, at that time his discovery went unnoticed (Geyer & Pepple Williamson, 2004).

Unaware of Newcomb and his discovery, nearly 60 years later American physicist Frank Benford also noticed that the first few pages of logarithmic tables were more worn than later, and like Newcomb, he assumed that people more viewed the logarithms of lower numbers than higher. In his famous article "The Law of Anomalous Numbers" (Benford, 1938), Benford tested the assumption that lower digits are more likely to occur in numbers than higher, and he analysed digits in 20 very different groups of numerical data, such as death rates, atomic weights, American League baseball results in 1936 championship, etc. His analysis of the appearance of the digits in numbers showed that in each of these completely different lists of numbers, smaller digits appear more often as the first digit than larger ones, and that there is a certain pattern which exists in that appearance. Based on this analysis, Benford concluded that numbers in nature follow logarithmic distribution

rather than uniform, as might be assumed, and that the appearance of the first digit in a number depends on the distance between that number and its follower divided by the entire length of the scale, which is characteristic of the logarithmic base 10. From that, it results that the number 1 will appear as the first digit on average in 30.6% of the numbers, or 0.306 as the proportion, which roughly corresponds to the logarithm of the number 2, and that the number 2 will appear in the proportion of 0.185 which is approximately logarithm of 3/2. This pattern goes all the way to number 9, which will appear in a ratio of 0.047, which roughly corresponds to a logarithm of 10/9 (Hickman & Rice, 2010).

Based on these observations, Benford developed a formula for the expected frequencies of the digits in the number lists, and these frequencies became known as the Benford’s distribution, or Benford’s law, as follows (Nigrini, 2011: 87):

$$P(D_1 = d_1) = \log(1 + 1/d_1) \quad d_1 \in \{1, 2, \dots, 9\} \tag{1}$$

where D_1 representing the first digit, P indicates the probability of occurrence in data sets and \log refers to the log to the base 10. The expected frequencies of occurrence of the first digits are shown in Table 1.

Table 1: Expected frequencies of first digits

First digit	1	2	3	4	5	6	7	8	9
Frequency	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

Source: Benford, 1938: 556.

The fact that we can reliably predict, and explain which digits will first appear in numbers has led to the application of Benford’s law in various fields. Benford’s law is widely applied in the field of accounting and auditing, where it was used to analyse the occurrence of digits in financial statements (Amiram, Bozanic, & Rouen, 2015; Asslani & Naco, 2014; Clippe & Ausloos, 2012; Nigrini & Miller, 2009; Omerzu & Kolar, 2019; Shi, Ausloos, & Zhu, 2017; Shrestha, 2016) and for the analysis of stock exchange trading (Corazza, Ellero, & Zorzi, 2010; Jayasree, Pavana Jyothi, & Ramya, 2018). On the basis of Benford’s law analyses of various statistical data were also performed: crime statistics (Hickman & Rice, 2010), GDP statistics (Holz, 2014), air quality data (Stoerk, 2015), aerobiological data (Docampo, del Mar Trigo, Aira, Cabezudo, & Flores-Moya, 2009), occupational hygiene data (de Vocht & Kromhout, 2013), research results published in the American Journal of Sociology (Diekmann, 2007) and to compare deficit data and social security data (Rauch, Götttsche, Brähler, & Kronfeld, 2014). In Croatia, so far Benford’s law has been applied only in the field of accounting and auditing for the analysis of financial statements (Cunjak Mataković, 2019; Dumičić & Cunjak Mataković, 2019; Kopal, Nemeth, & Leinweber, 2019; Papić, Vudrić, & Jerin, 2017; Slijepčević & Blašković, 2014).

To use Benford’s law for data analysis, data must satisfy several conditions. The records should represent the sizes of facts or events, datasets should have at least 1000

records, and there should be no default range of data, i.e. built-in maximum and minimum values for the data (Nigrini, 2011). Likewise, the mean should be greater than the median, which means that smaller values predominate in the data set, and skewness should be positive. Additionally, numbers assigned in some way (like zip codes, invoice numbers, etc.) or those created under the influence of human thoughts, such as prices set at a psychological threshold of 0.99 or 1.99, cannot be used in the analysis (Durtschi, Hillison, & Pacini, 2004).

Tourism data – features and challenges

As mentioned earlier, statistics plays a significant role in tourism management, since it makes easier for stakeholders to plan and make decisions essential for tourism management. Significant steps were taken since the 1970s to increase countries' capacities to collect tourism activities data and to store, process, and report data, collected at the national level, in a coherent way (Frechtling & Hara, 2016) since different countries developed different data collection systems, based on different methodological settings, which made mutual comparisons difficult or even impossible. Many documents have been developed within various international organizations, aimed at overcoming differences in national tourism data collection systems and enabling international comparisons: in 1973 the European Travel Commission published a “European Program for Tourism Statistics”, in 1989 the World Tourism Organization published “Guidelines for the Collection and Presentation of Domestic Tourism Statistics” and “Recommendations on Tourism Statistics” in 1994, and OECD published “Manual on Tourism Economic Accounts” in 1996 (Lickorish, 1997). At the level of the European Union, in 1995 the “Directive on the collection of statistical information in the field of tourism” was adopted. A particularly important document is Regulation 692/2011, since it requires European Union member states to collect and submit to the European central authorities information on: a) the capacity and occupancy of tourist accommodation establishments; b) tourism nights spent in non-rented accommodation; c) tourism demand (domestic and outbound tourism); d) characteristics of same-day visits. Although at the level of international bodies various definitions and methodologies are established, data sources and the collection process are still issues that are implemented at the level of national authorities (Lickorish, 1997).

Data for tourism statistics are mainly collected in three ways: 1) counting passengers or conducting interviews at state borders; 2) surveys of households and destinations about tourism activities; 3) recording of tourist arrivals and overnight stays in establishments that provide accommodation to tourists (Volo, 2004; Volo & Giambalvo, 2008). The first two methods are used for demand side analysis and the third is used for the supply side. In this paper we will make a more detailed analysis of the data collected for the supply side through accommodation statistics. Supply side information is collected by national statistical authorities: they collect data on the number of accommodation establishments, and for each accommodation establishment is recorded the number of rooms, beds, tourist arrivals and overnight stays, and similar.

Data collection for the supply side has certain limitations: first, it is impossible to determine which guests are tourists and which come for some other purposes, such as business, which is often the case in urban destinations, or as seasonal workers, students, and similar. Another problem is that the visitor can stay in several different establishments

during their trip, and each stay will be recorded separately, which will lead to an unrealistic increase in the number of visitors (so-called “double counting effect”) (De Cantis et al., 2015: 3).

Also, for the supply side information, a special problem is tourist activities that are not recorded or are not accurately recorded in official statistics, so there emerges the issue of hidden tourism activities, which is in the literature called “underground tourism” (De Cantis et al., 2015) or “concealed tourism” (Parroco & Vaccina, 2004). The term “underground tourism” refers to official establishments registered for tourism activities, which present only partial information about their activities, i.e. not completely accurate information on the number of tourist arrivals and overnight stays, and thus reduce the accuracy and reliability of official statistics (De Cantis et al., 2015). This “misreported consumption of tourism products” (Volo, 2004) can lead to overestimated demand forecasts for certain products, and at the end supply will exceed demand. Likewise, tourism that is deliberately hidden from public authorities reduces tax revenues, thus preventing local governments to properly manage the local community but also to care about maintenance of environmental resources (Parroco & Vaccina, 2004).

The size and significance of tourism hidden in this way depend to a significant extent on “socio-economic behaviour and on the territorial order of interested areas” (Parroco & Vaccina, 2004). By a literature review it can be concluded that there exist two main reasons for inaccurate or incomplete reporting of tourist traffic: one of them is negligence, and the other is tax evasion (Vlahović, 1998; Javor & Kalčić, 2003; Volo, 2004; Volo & Giambalvo, 2008; De Cantis et al., 2015; Krajnović et al., 2020), or the possibility of obtaining tax reliefs (Guizzardi & Bernini, 2012).

The estimated differences between the actual situation and the one recorded in the official statistics are significant. For Italy, underreporting in accommodation statistics was estimated to at least 16% in the period between 2007 and 2009 (Guizzardi & Bernini, 2012). It is estimated that in some tourist districts in Sicily, according to 2001 data, hidden tourism accounted for up to 1/5 (Parroco & Vaccina, 2004). Fontana and Pistone (2010) on data for the period from 2000 to 2007 found that in the province of Piedmont in Italy the actual number of overnight stays was 30% higher than the officially recorded number. De Cantis et al. (2015) using data for 2010 estimated that in Sicily “the ratio between unobserved and observed nights is almost equal to 4” (De Cantis et al., 2015: 11), however, it should be taken into account that under the rules of the statistical system in Italy unobserved overnight stays also included overnight stays in non-commercial accommodation. Milićević and Galić (2018) estimated that in Bosnia and Herzegovina the level of the informal economy in tourism in 2016 was between 40 and 60% of GDP. Çakmak and Çenesiz (2020) analysing data for the period from 1975 to 2017 found that in Thailand, informal GDP in tourism amounts to about 50% of formal GDP in tourism. Research on unreported tourism in Croatia is quite rare. Stučka (2002) estimated that unregistered overnight stays in 1998 amounted to 33 to 39% of the total registered overnight stays, between 39 and 53% in 1999 and between 12 and 22% in 2000. Javor and Kalčić (2003) state that “20 – 50 per cent of private rental accommodation is statistically unrecorded” (Javor & Kalčić, 2003: 101).

Considering the large importance of tourism for overall economic activities, but also rather large volume of hidden tourism in the above-mentioned countries, various

methods have been developed to assess unreported tourism activists. There are three basic approaches for assessment (Parroco & Vaccina, 2004; De Cantis et al., 2015):

- 1) comparison of different statistical sources on tourism - data collected for the demand side is compared with supply side data, and it is also possible to compare different statistical sources (e.g. data on tourism and data on traffic);
- 2) use of indirect measures, where indicators about the presence of tourists are analysed - indicators used in such research are electricity and water consumption, amount of waste produced, newspapers sold and similar;
- 3) ad hoc surveys on different samples - it is possible to conduct surveys on residents, where they are questioned about their travel plans, and the costs they have made or plan to make in tourism activities. Tourists can also be surveyed, and surveys are conducted at state borders or in tourist destinations. The most accurate data will be obtained for costs incurred in the previous 24 hours from the time of the survey (Frechtling, 2006).

Except for its positive sides, each of these three methodological approaches for assessment of unreported tourism also has negative aspects. The negative sides of the comparison of different statistical sources are the lack of uniformity of different statistical sources, low details in specific territorial units, and the lack of information for specific variables. When using indirect indicators, it is difficult to separate tourist components from the entire population which is being observed, and it is also difficult to estimate the translation coefficient of an individual indicator in the number of tourists. Ad hoc surveys are complex, expensive, and difficult to replicate, thus making historical analysis impossible (De Cantis et al., 2015). In order to avoid the negative sides of these three approaches, this paper will use Benford's law to assess the quality of the submitted statistical data on tourist arrivals and overnight stays in Croatia. Benford's law allows a simple and effective assessment of large data sets since "it has been shown that manipulated, unrelated, or created numbers usually do not follow Benford's law, which can be ascribed to most people's misconceptions of randomness and distributions of real data" (de Vocht & Kromhout, 2013: 297). It has also been found that in cases where data has been accidentally altered without the fraud intention, for example when errors have occurred during data processing or copying, because of these unintentional errors the data will deviate from Benford's distribution (ibid.).

According to our knowledge, so far Benford's law was not applied to tourism statistics. Benford's law has been very rarely applied to tourism-related research, with the exception of Jawabreh, Bader, Saleh and Alrabei (2018) and Chemin and Mbiekop (2015). Jawabreh et al. (2018) did the research of fraud in companies from the hospitality sector in Jordan and tested their financial statements by Benford's law. Chemin and Mbiekop (2015) did research on child sex tourism in India and used Benford's law to test the quality of criminal data.

Data and methodology

Tourism data

In this analysis there was used data on tourist arrivals and overnight stays in 30 destinations in Croatia, collected on monthly basis, and since the first year included in this analysis is 2013, in the analysis there were included destinations that had the most tourist

visits and overnight stays in 2013¹. Considering that tourism in Croatia is very seasonal, with most tourist activities in the period from June to September, only those destinations that had tourist arrivals and overnight stays in all 12 months of the year were included in the analysis, in order to avoid lack of data for winter months. For the period from 2013 to 2015, it was used data collected by the Croatian Bureau of Statistics, and published in the First Releases, which are issued on a monthly basis. Data for the preparation of these monthly releases was collected through the Monthly Report on Tourists and Tourist Nights (TU-11 form), which is explained in more detail in the introduction. The analysis for the period from 2016 to 2018 includes the same 30 destinations, and the data were collected from the eVisitor system.

A separate analysis was done for domestic and foreign tourists, in order to determine whether the origin of tourists affects the recording of data.

Methodology

In order to analyse the quality of the submitted statistical data on the tourist arrivals and overnight stays, the data will be analysed by using Benford's law. Three tests were used to analyse the conformity of tourism activity data with Benford's first-digit distribution: The Chi-square test, the Kolmogorov-Smirnov test, and the Mean Absolute Deviation.

The chi-square test is based exclusively on the analysis of absolute frequencies, and it will be used to determine how much the actual frequency conforms to the expected frequency, i.e. to the Benford's distribution. The chi-square test for analysing the first digits was calculated as follows, as shown in equation 2 (Nigrini, 2011: 112):

$$\text{Chi - square} = \sum_{i=1}^K \frac{(AC-EC)^2}{EC} \quad (2)$$

Where AC and EC represent the Actual Count and Expected Count, and K represents the number of bins. The number of degrees of freedom equals $K - 1$, which means that for the first digit the test is evaluated by using 8 degrees of freedom. At the significance level of 0.05, with 8 degrees of freedom, the critical value is 15.51.

The Kolmogorov-Smirnov test was also used to test compliance with the expected Benford's distribution. This test is based on the cumulative density function, i.e. the density functions of the two distributions are compared. To determine whether a result is statistically significant can be done by the following formula:

$$\text{Kolmogorov - Smirnov} = \frac{1.36}{\sqrt{N}} \quad (3)$$

where 1.36 is constant for a significance level of 0.05 and N is the number of records (Nigrini, 2011: 113). The Kolmogorov-Smirnov test takes the largest deviation from Benford's distribution, which is determined by the cumulative difference that exists between the empirical distribution of digits 1 to 9 and the theoretical distribution, i.e. the

¹ The destinations included in the analysis are Crikvenica, Krk, Mali Lošinj, Opatija, Rijeka, Lovran, Puntar, Senj, Plitvička jezera, Biograd na Moru, Zadar, Starigrad, Šibenik, Vodice, Hvar, Makarska, Split, Supetar, Podstrana, Okrug, Novigrad, Poreč, Pula, Rovinj, Umag, Fažana, Dubrovnik, Korčula, Orebić and Zagreb.

Benford’s distribution (Amiram et al., 2015). This test was used to determine compliance with the Benford’s distribution, compared to a critical value, which equals 0.041 for a significance level of 0.05.

The third used test is Mean Absolute Deviation test. This test does not take into account the sample size, and is calculated by using the formula:

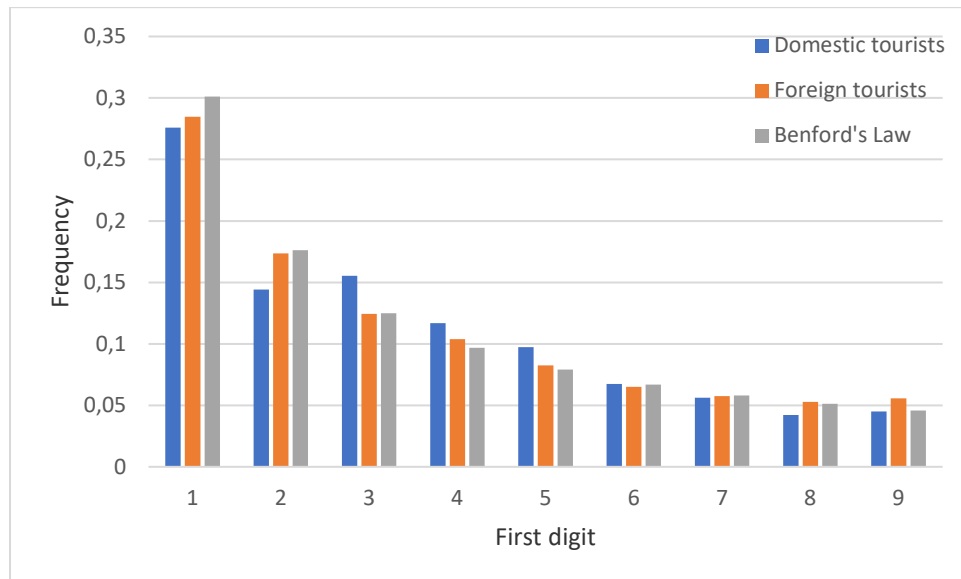
$$\text{Mean Absolute Deviation} = \frac{\sum_{i=1}^K |AP-EP|}{K} \tag{4}$$

where EP denotes the expected proportion, AP the actual proportion and K represents the number of bins, which equals 9 for the first digit (Nigrini, 2011: 114). The Mean Absolute Deviation test is calculated as the sum of the absolute difference between the empirical frequency and each digit, from 1 to 9, and the theoretical frequency which is based on Benford’s law, divided by the number of leading digits used (Amiram et al., 2015). The Mean Absolute Deviation test has no analytically derived critical value, but Nigrini (2012) defined empirically based criteria for analysing deviations from the expected Benford’s distribution. He proposed four ranks: “close conformity” (0.000-0.006), “acceptable conformity” (0.006-0.012), “marginally acceptable conformity” (0.012-0.015) and “nonconformity” (above 0.015).

Results

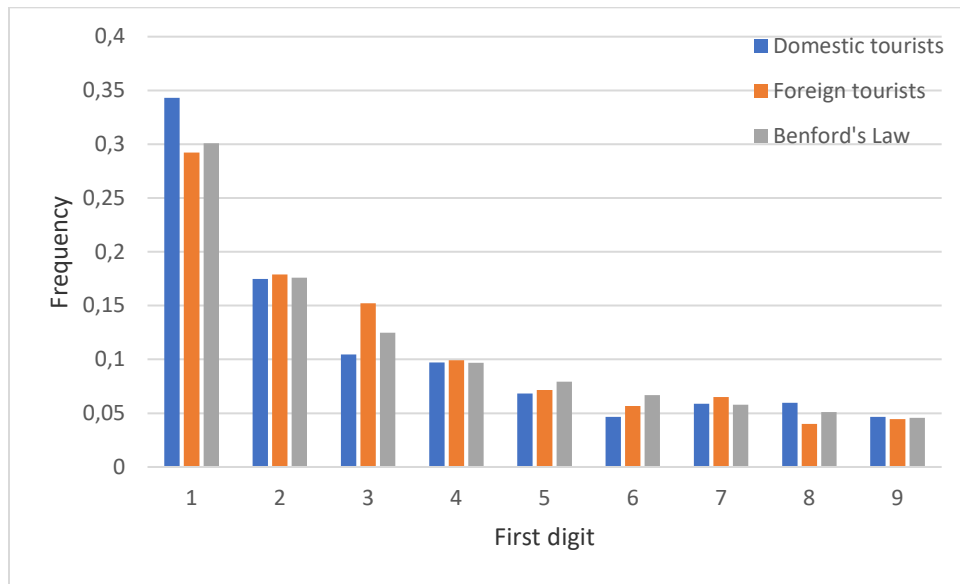
As stated above, for data that is not accidentally or intentionally manipulated, it is expected that the first digits in the numbers will follow Benford’s distribution. The distribution of the first digits in the data for the period from 2013 to 2015 is shown in Figure 1.

Figure 1. Expected and actual distribution of first digits, 2013-2015 tourist arrivals data



Source: Croatian Bureau of Statistics

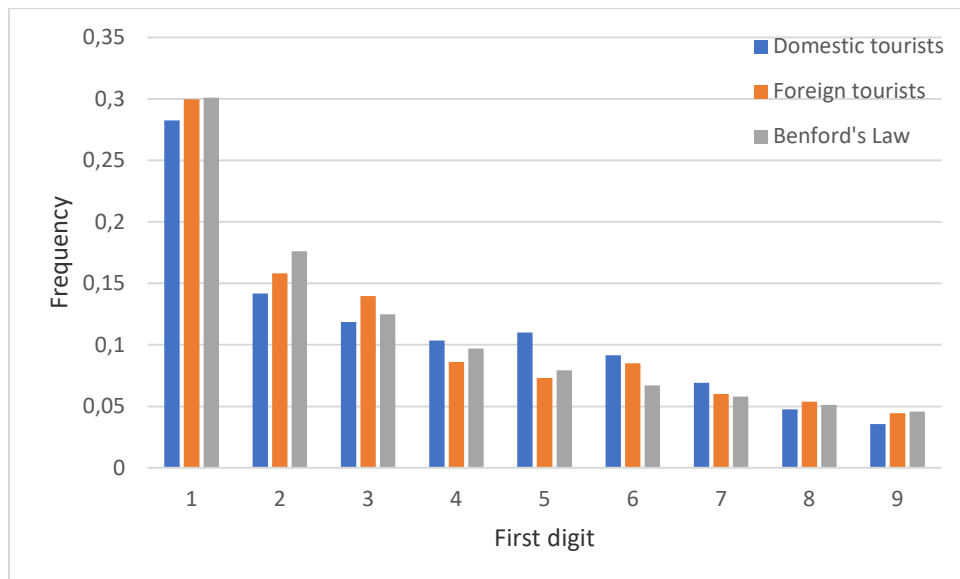
Figure 2. Expected and actual distribution of first digits, 2013-2015 tourist overnight stays data



Source: Croatian Bureau of Statistics

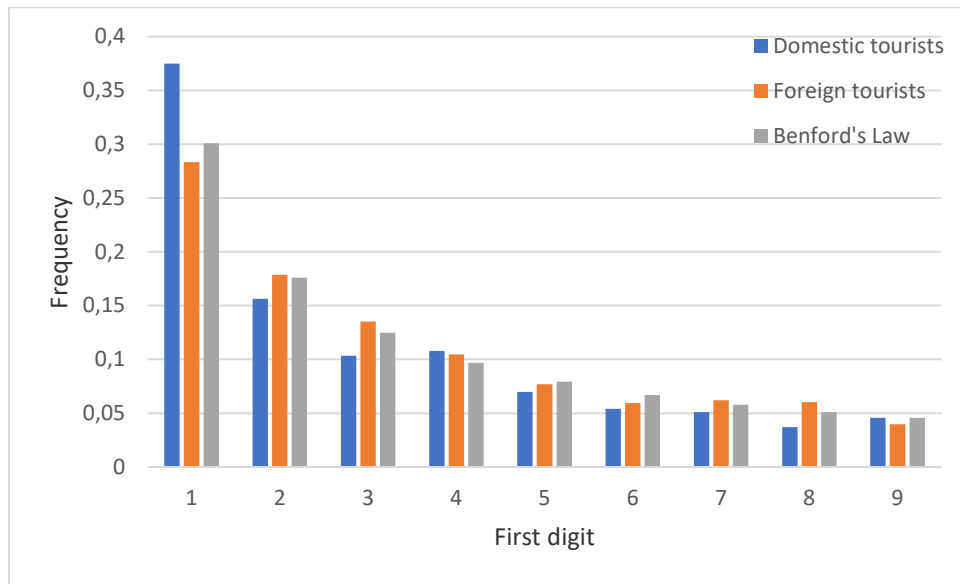
As can be seen from Figure 1, for tourist arrivals data the largest deviations from the expected Benford's distribution were found for the digit 1, both for domestic and foreign tourists, and for domestic tourists' larger deviations for digits 2 and 3 was also found. In the data for tourist overnight stays deviation from the expected distribution is found for digit 1 for domestic tourists, and for digit 3 deviations both for domestic and foreign tourists were found, except that the number of domestic tourists is lower than the expected Benford's distribution and the number of foreign tourists is higher.

Figure 3. Expected and actual distribution of first digits, 2016-2018 tourist arrivals data



Source: Croatian Bureau of Statistics

Figure 4. Expected and actual distribution of first digits, 2016-2018 tourist overnight stays data



Source: Croatian Bureau of Statistics

For the period from 2016 to 2018, the distribution of the first digits is shown in Figures 3 and 4. Regarding tourist arrivals, for domestic tourists digits 2, 5 and 6 deviate the most from the Benford’s distribution, and for foreign tourists the most deviate digits are 2, 3 and 6. For tourist overnight stays the largest deviation is evident for the digit 1 in the data for domestic tourists, and also the digits 2 and 3. In the data for foreign tourists, the digits 1 and 3 deviate the most.

Looking at these data, it can be concluded that in both observed periods a larger deviation of digits was found in the data for domestic tourists than for the foreign, and the largest deviation was found in digit 1 in the data for tourist overnight stays in both observed periods.

Three tests were used to analyse deviations from the Benford’s distribution: The Chi-square test, the Kolmogorov – Smirnov test, and the Mean Absolute Deviation.

The results of the Chi-square test for the period from 2013 to 2015 are shown in Table 2. The obtained results are statistically significant for a significance level of 0.05, and with 8 degrees of freedom the critical value equals 15.51. In tourist arrivals results, for domestic tourists the chi-square value is 26.99, and for foreign tourists is 4.13, so it can be concluded that the data for domestic tourists have shown statistically significant deviation from the Benford’s distribution. The results are similar for overnight stays – for domestic tourists the chi-square value is also higher than the critical value (equals 19.55) and for foreign tourists are lower (equals 12.93), so here it leads to conclusion that the data for domestic tourists have shown statistically significant deviation from the expected distribution.

Table 2. Chi-square test for 2013 – 2015 data

Digit	Tourist arrivals		Tourist overnight stays	
	Domestic tourists	Foreign tourists	Domestic tourists	Foreign tourists
1	26.99	4.13	19.55	12.93
2	15.51	15.51	15.51	15.51
3	15.51	15.51	15.51	15.51
4	15.51	15.51	15.51	15.51
5	15.51	15.51	15.51	15.51
6	15.51	15.51	15.51	15.51
7	15.51	15.51	15.51	15.51
8	15.51	15.51	15.51	15.51
9	15.51	15.51	15.51	15.51

	Chi-square	Chi-square	Chi-square	Chi-square
1	2.23	0.94	6.26	0.28
2	6.23	0.04	0.01	0.05
3	7.88	0.00	3.52	6.38
4	4.42	0.54	0.00	0.06
5	4.43	0.16	1.62	0.82
6	0.00	0.07	6.53	1.73
7	0.06	0.00	0.01	0.90
8	1.71	0.06	1.57	2.67
9	0.02	2.31	0.02	0.04
Total	26.99	4.13	19.55	12.93

Table 3 shows results of the chi-square test for the period from 2016 to 2018. The obtained results are similar to those for the period from 2013 to 2015: the chi-square value for domestic tourists in both observed categories (tourist arrivals; overnight stays) is higher than the critical value (15.51) and therefore it could be concluded that the difference between them and the expected Benford's distribution is statistically significant. Data for foreign tourists conforms to the expected distribution since the chi-square value is lower than the critical value.

Table 3. Chi-square test for 2016 – 2018 data

Digit	Tourist arrivals		Tourist overnight stays	
	Domestic tourists	Foreign tourists	Domestic tourists	Foreign tourists
	Chi-square	Chi-square	Chi-square	Chi-square
1	1.20	0.01	19.48	1.12
2	7.16	1.97	2.40	0.04
3	0.36	1.88	4.05	0.91
4	0.49	1.32	1.34	0.66
5	12.92	0.51	1.20	0.07
6	9.59	5.32	2.71	0.95
7	2.25	0.09	0.86	0.30
8	0.27	0.13	4.09	1.72
9	2.49	0.04	0.00	0.83
Total	36.73	11.27	36.13	6.63

The Kolmogorov-Smirnov test was also used to test the conformity with the expected Benford's distribution. As can be seen from Table 4, empirical values for domestic tourists, both for tourist arrivals and for overnight stays, are higher than the critical value which equals 0.041. In the data for the period from 2013 to 2015 for tourist arrivals the empirical value is 0.057 and for overnight stays is 0.042. In the period from

2016 to 2018, the empirical values are also higher than the critical value and are 0.059 (tourist arrivals) and 0.074 (overnight stays). All results of the Kolmogorov-Smirnov test for foreign tourists in both observed periods are lower than the critical value, so it can be concluded that only the data for domestic tourists have shown statistically significant deviation from the expected Benford’s distribution.

Table 4. Kolmogorov–Smirnov Z test

		Empirical value	
		2013 - 2015	2016 - 2018
Tourist arrivals	Domestic tourists	0.057	0.059
	Foreign tourists	0.019	0.020
Tourist overnight stays	Domestic tourists	0.042	0.074
	Foreign tourists	0.023	0.017

The third conducted test was the Mean Absolute Deviation test. The values obtained by the MAD test are shown in Table 5. According to empirically based criteria for compliance to expected Benford’s distribution, defined by Nigrini (2012), the obtained results can be classified into the following ranks: as only data for foreign tourist arrivals in the period from 2013 to 2015 (MAD = 0.00489) can be classified as “close conformity”, while in the category “acceptable conformity” all other data for foreign tourists can be classified as well as for tourist overnight stays for domestic tourists in the period from 2013 to 2015. All other data for domestic tourists should be classified as “nonconformity” since their MAD test result is higher than 0.015.

Table 5. Mean Absolute Deviation test

		MAD	
		2013 - 2015	2016 - 2018
Tourist arrivals	Domestic tourists	0.01531	0.01628
	Foreign tourists	0.00489	0.00821
Tourist overnight stays	Domestic tourists	0.01164	0.01896
	Foreign tourists	0.00876	0.00747

The results of the MAD test have shown, as the previous two tests, that the data for foreign tourists conforms more with the expected Benford’s distribution than the data for domestic tourists, and similar results were obtained in both observed periods.

Discussion

Tourism is a complex social phenomenon, and its management is a demanding task, among others, because of its dynamic nature. Accurate and complete data on tourism activities is one of the necessary prerequisites for successful tourism management. Except for the accuracy and completeness of the information, the time dimension is also extremely important in tourism: data should be available in a timely manner, to enable that strategic and business decisions are based on them. To meet all these requirements, in 2016 a new information system in tourism called eVisitor was introduced in Croatia. This system has significantly reduced the administrative workload in the tourist’s registration process and

also allowed all interested users to monitor tourist traffic in almost real-time by various parameters. The quality of the system can be seen from the fact that it received an award at the International Tourism Trade Fair in Madrid in 2018, and it was also assessed by experts, who analysed its characteristics and functionality, with a 4+ grade, out of 5 possible points (Krajnović et al., 2020).

User-friendliness is very important in such data collection systems - for example, in Italy introduction of a new web-based service for tourism data has led to significant improvements in the collection process, with a large reduction of missing values (Fontana & Pistone, 2010). In the already mentioned experts' evaluation of eVisitor, the functions of the system "Collection and processing of data on accommodation providers and their accommodation facilities", "Registration and deregistration of tourists" and "Calculation and control of sojourn tax" received the highest possible grades, i.e. 5 out of 5 points. Having all that in mind, it can be concluded that there are no obstacles for the usage of this system by accommodation providers since the system is easy to use, free of charge and can be installed on a mobile phone, and therefore does not require the users to own a computer.

With the launch of this system in Croatia, for the first time a comprehensive database was available on tourists and their stay, and the collection and processing of data were significantly improved. In this way, in addition to the already mentioned timeliness, the comprehensiveness of the data was achieved. However, for the system to be fully functional and bearing in mind the information it contains possibly shapes public policies and makes business decisions, the data must be accurate and complete. This was also one of the aims of this paper, to determine whether the data collected by the new system, which is technically more advanced and easier to use, is more reliable than the previously collected data. Several authors pointed out that the quality of previously collected data was questionable: "our already traditional negligence/disrepair in data collection and processing is a kind of "destruction" factor for the system" (Vlahović, 1998: 195), therefore the data were "mostly incomplete" and "random, due to the existence of the grey economy" (Krajnović et al., 2020: 128). This aim was achieved by analysing data on the basis of Benford's Law, a method that "enables easy screening of large data sets to identify potential problems with the data, manipulated or other, for further, more in-depth investigation" (de Vocht & Kromhout, 2013: 297).

One of the basic thesis of Benford's law is that digits in numbers do not appear randomly, but that there is a regularity in the appearance of these digits and it was the basis of our analysis to determine whether tourist data recorded by accommodation providers is in line with Benford's distribution of the first digits. The conformity of digits occurrence with Benford's distribution would indicate the accuracy of the data, and a deviation from Benford's distribution may indicate that the data is inaccurate, i.e. that data was accidentally or intentionally manipulated. In order to achieve this, the data for the period from 2013 to 2015, which were collected by the Croatian Bureau of Statistics through the Monthly Report on Tourists and Tourist Nights (TU-11 form), was analysed separately, and due to the relative complexity of data collection, their accuracy can be doubted. The second data group contains data for the period from 2016 to 2018, which was collected through the eVisitor system, which allowed users to easily and quickly enter data, so it can be assumed that because of simplicity and user-friendliness this data is more reliable than those collected in the previous period. Both analysed periods included larger tourist destinations in the analysis which are visited by a higher number of tourists and therefore

have a developed infrastructure but also experience and tradition of doing business with tourists. Therefore, it can be presumed their business maturity and that the collected data is mainly accurate and complete. Data for domestic and foreign tourists were analysed separately, to determine whether the origin of tourists had an impact on the quality of the collected data.

Statistical analysis made by using three tests (Chi-square test, Kolmogorov-Smirnov test, and Mean Absolute Deviation test) showed very similar results in both observed periods, and therefore it can be concluded that the characteristics of the data collection system did not affect the quality of collected data, but that the origin of the tourists was more important. Data on foreign tourists in both observed periods conform to Benford's distribution and therefore it can be assumed that their accuracy is not questionable. Data for domestic tourists in all three statistical tests in both observed periods showed a deviation from the expected Benford's distribution, so it can be concluded that their quality is lower than the data for foreign tourists.

When analysing this data it should be taken into account that Benford's law is "not a lie detector" (Omerzu & Kolar, 2019: 56), but only an indicator of increased risk of fraud or error, since deviations from Benford's law indicate that digits in a certain set of numbers are manipulated in some way, but by Benford's law it cannot be determined whether data manipulations are accidental or intentional. Unintentional errors may occur for technical reasons during the physical collection and data processing - for example during data transfer (if not all data is transferred or if numbers are cutoff on two decimals) or errors such as manual typos during copying of data or accidental misplacement of digits when copying data from one medium to another (de Vocht & Kromhout, 2013). Unintentional manipulation can also occur due to negligence in data collection and delivery – reporting on tourist traffic represents a waste of time for accommodation providers, especially for smaller ones, so most of them will not be motivated to report tourist traffic (if by doing they will not get some benefits or if it is certain that nonreporting of tourists will not be sanctioned). The most common reason for intentional data manipulation, i.e. for avoiding reporting on the actual number of tourists is to avoid paying sojourn taxes and even more often to avoid paying taxes, which is a domain of the informal economy. Although the informal economy is quite widespread in Croatia, tax evasion on tourist overnights stays is not a phenomenon specific only for Croatia but is evident in other Mediterranean countries such as Italy (Volo, 2004; Volo & Giambalvo, 2008; Guizzardi & Bernini, 2012; De Cantis et al., 2015) and Bosnia and Herzegovina (Milićević & Galić, 2018).

There are two potential explanations why in both observed periods the data for domestic tourists have shown a statistically significant deviation from the expected Benford's distribution, unlike the data for foreign tourists. Domestic tourists, especially the middle-aged and elderly, still prefer to pay the accommodation cost in cash. When the accommodation cost is paid in cash, it is easier to "hide the tourist": since there is no trace of payment, there is also no trace of his stay in the accommodation facility, so his stay will not be recorded anywhere and the accommodation provider will avoid paying tax and sojourn tax. Paying invoice in cash, which results in tax evasion, is quite common in Croatia: according to research conducted by Škrinjarić, Recher and Budak (2017) it was estimated that in 2014, the unofficial economy from household consumption was 0.69 percent of GDP. Another possible explanation is the legal obligation for all accommodation service providers to report the stay of foreign tourists to the Ministry of the Interior: foreign

tourists must be reported to the Ministry of Interior within one day of their arrival and check-out also within one day after their departure. To avoid possible problems with the police, accommodation providers more regularly record the arrivals of foreign tourists, and for domestic tourists, who are not subject to this legal obligation, take a different approach and their stay is only partially recorded. Knowledge of the language and customs of the country probably also contributed to avoiding the registration of domestic more often than foreign tourists.

In this paper was not conducted analysis according to the type of accommodation provider, since such data is not publicly available for the period from 2013 to 2015. Therefore, it is not possible to determine how accurately the tourist's data is kept in a particular accommodation type. Since larger hotels are systems with prescribed procedures and business policies, it can be assumed that they have accurately reported the number of tourists. In smaller hotels, especially family-owned ones, it is possible that the number of tourists was decreased to avoid paying taxes and the same could be in campsites, especially smaller ones. Private accommodation establishments pay a flat income tax, so the amount of their tax burden does not depend on the number of tourists who stayed in their establishment (Vranar, 2015). Administering tourists data for the accommodation provider is still a cost of time, which is especially important in the summer at the peak of the season when there are a lot of tourists and a lot of work around them, so to save time they do not record domestic tourists on an up-to-date basis.

Conclusions

The professional and general public in Croatia welcomed the introduction of the eVisitor information system because it was assumed that it would modernize, simplify, and improve business in tourism. It was also assumed that this system would bring order to the accommodation renting system, since it has been obvious for a longer period of time that a number of accommodation providers do not operate in accordance with the law and business ethics, by accommodating tourists in facilities that are not registered to perform accommodation services or that the actual number of tourist overnights stays was not recorded. Regarding the introduction of order in the system, eVisitor showed results already in the first year of its implementation: one of the indirect effects of the eVisitor introduction was an increase in the number of registered accommodation providers. Namely, in 2016, the first year in which the eVisitor system was implemented, the number of registered accommodation providers increased by 7000, i.e. from 67000 to 74000 establishments. The introduction of eVisitor also facilitated the supervision of accommodation providers, and in the first 6 months of 2016, the State Inspectorate filed minor charges against 40 illegal accommodation providers. Illegal accommodation providers were discovered by the State Inspectorate by comparing the records of registered accommodation providers from eVisitor with the ads on booking.com and Airbnb web pages (Dobrota, 2016), so it can be assumed that a certain number of unregistered accommodation providers decided to legalize their activities due to the transparency provided by eVisitor.

Unfortunately, neither the introduction of the new information system nor the existing legal framework influenced the accommodation service providers to register domestic tourists more accurately, which was also shown in this paper. The tourism

authorities in Croatia, primarily the Ministry of Tourism and the Tourist Boards, should make additional efforts to achieve more accurate recording of tourist arrivals and overnight stays, and these efforts should go in two directions: one is educational, to point out to accommodation service providers the consequences of incorrect tourism traffic recording. The second is repressive: increased controls and the expansion of the powers of supervisory bodies could force accommodation providers to report real data and to perform their tax obligations accordingly. The effects of inaccurate reporting are not exclusively fiscal but are much broader: overuse of resources leads to a decline in the tourist experience and lower prices in nonregistered accommodation establishments lead to unfair competition (De Cantis et al., 2015: 16).

This paper has two limitations. It was not analysed non-commercial accommodation, since detailed data for non-commercial accommodation for the period before 2016 is not publicly available, but only total data at the county level are available. It could be expected that the situation in non-commercial accommodation is even worse than in commercial accommodation since due to its very nature it is more difficult and complex to keep proper records of the number of tourists, as it is difficult to distinguish “real” tourists from relatives and friends of accommodation owners. Another limitation is that it was not possible to make analysis according to the type of accommodation provider (hotels, camping sites, private accommodation establishments, etc.), since for the period from 2013 to 2015 publicly available is only total data, without distinguishing the particular type of accommodation provider.

References

1. Amiram, D., Bozanic, Z., & Roeun, E. (2015). Financial statement errors: evidence from the distributional properties of financial statement numbers. *Review of Accounting Studies*, 20 (4), 1540-1593. <https://doi.org/10.1007/s11142-015-9333-z>
2. Aroca, P., Brida, J. G., & Volo, S. (2017). Tourism statistics: Correcting data inadequacy. *Tourism Economics*, 23 (1), 99-112. <https://doi.org/10.5367/2Fte.2015.0500>
3. Asslani, A., & Naco, M. (2014). Using Benford's Law for Fraud Detection in Accounting Practices. *Journal of Social Science Studies*, 1 (2), 129-143.
4. Baldigara, T., & Mamula, M. (2012). Tourism statistics in Croatia: Present status and future challenges. *Procedia - Social and Behavioral Sciences*, 44, 53-61, <https://doi.org/10.1016/j.sbspro.2012.05.004>
5. Benford, F. (1938). The Law of Anomalous Numbers. *Proceedings of the American Philosophical Society*, 78 (4), 551-572.
6. Çakmak, E., & Çenesiz, M. A. (2020). Measuring the size of the informal tourism economy in Thailand. *International Journal of Tourism Research*, 22, 637-652, <https://doi.org/10.1002/jtr.2362>
7. Chemin, M., & Mbiokop; F. (2015). Addressing child sex tourism: The Indian case. *European Journal of Political Economy*, 38, 169-180. <https://doi.org/10.1016/j.ejpoleco.2015.02.005>
8. Clippe, P., & Ausloos, M. (2012). Benford's law and Theil transform of financial data. *Physica A: Statistical Mechanics and its Applications*, 391 (24), 6556-6567. <https://doi.org/10.1016/j.physa.2012.07.063>
9. Corazza, M., Ellero, A., & Zorzi, A. (2010). Checking financial markets via Benford's law: The S&P 500 case. In M. Corazza & C. Pizzi (Eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance* (pp. 93-102). Milano: Springer.
10. Cunjak Mataković, I. (2019). The empirical analysis of financial reports of companies in Croatia: Benford distribution curve as a benchmark for first digits. *Croatian Review of Economic, Business and Social Statistics*, 5 (2), 90-100. <https://doi.org/10.2478/crebss-2019-0014>
11. De Cantis, S., Parroco, A. M., Ferrante, M., & Vaccina, F. (2015). Unobserved tourism. *Annals of Tourism Research*, 50, 1-18. <https://doi.org/10.1016/j.annals.2014.10.002>

12. De Cantis, S., & Ferrante, M. (2013). Tourism statistics and unobserved tourism: empirical evidences in Sicily. *DIEM: Dubrovnik International Economic Meeting*, 1 (1).
13. De Vocht, F., & Kromhout, H. (2013). The Use of Benford's Law for Evaluation of Quality of Occupational Hygiene Data. *Annals of Occupational Hygiene*, 57 (3), 296-304. <https://doi.org/10.1093/annhyg/mes067>
14. Diekmann, A., & Jann, B. (2010). Benford's Law and Fraud Detection: Facts and Legends. *German Economic Review*, 11 (3), 397-401. <https://doi.org/10.1111/j.1468-0475.2010.00510.x>
15. Dobrota, A. (2016). e-Visitor preplasio iznajmljivace smjestaja na crno? Retrieved from <https://www.cimerfraj.hr/aktualno/e-visitor-preplasio-iznajmljivace-smjestaja-na-crno>.
16. Docampo, S., del Mar Trigo, M., Cabezudo, B., & Flores-Moya, A. (2009). "Benford's law applied to aerobiological data and its potential as a quality control tool". *Aerobiologia*, 25, 275-283.
17. Dumičić, K., Cunjak Mataković, I. (2019). Challenges of Benford's Law goodness-of-fit testing in discovering the distribution of first digits: comparison of two industries. *Proceedings of the 15th International Symposium on Operational Research SOR'19* (pp. 290-295). Ljubljana: Slovenian Society Informatika, Section for Operational Research.
18. Durtschi, C., Hillison, W. A., & Pacini, C. (2004). The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data. *Journal of Forensic Accounting*, 5 (1), 17-34.
19. Frechtling, D. C., & Hara, T. (2016). State of the world's tourism statistics and what to do about it. *Tourism Economics*, 22 (5), 995-1013. <https://doi.org/10.5367/2Fte.2015.0474>
20. Frechtling, D. C. (2006). An Assessment of Visitor Expenditure Methods and Models. *Journal of Travel Research*, 45 (1), 26-35. <https://doi.org/10.1177/2F0047287506288877>
21. Geyer, C. L., & Pepple Williamson, P. (2004). Detecting Fraud in Data Sets Using Benford's Law. *Communications in Statistics - Simulation and Computation*, 33 (1), 229-246. <https://doi.org/10.1081/SAC-120028442>
22. Hickman, M. J., & Rice, S. K. (2010). Digital Analysis of Crime Statistics: Does Crime Conform to Benford's Law? *Journal of Quantitative Criminology*, 26 (3), 333-349. <https://doi.org/10.1007/s10940-010-9094-6>
23. Hill, T. (1995). A Statistical Derivation of the Significant-Digit Law. *Statistical Science*, 10 (4), 354-363.
24. Holz, C. A. (2014). The quality of China's GDP statistics. *China Economic Review*, 30, 309-338.
25. Ivandić, N., & Marušić, Z. (2009). Je li sustav statistike turizma u Hrvatskoj spreman za izradu satelitskog računa turizma? *Turizam: znanstveno-stručni časopis*, 57 (2), 185-205.
26. Javor A. & Kalčić, I. (2003). Private Accommodation in Tourism Statistics in the Republic of Croatia. In J. J. Lennon (Ed.), *Tourism Statistics: International Perspectives and Current Issues* (pp. 92-104). London, New York: Continuum.
27. Jawabreh, O., Bader, A., Saleh, M., & Alrabei, A. M. A. (2018). Fraud Inference Investigation within Tourism Sector in Jordan. *Journal of Economic & Management Perspectives*, 12 (1), 5-20.
28. Jayasree, M., Pavana Jyothi, C. S., & Ramya, P. (2018). Benford's Law and Stock Market - The Implications for Investors: The Evidence from India Nifty Fifty. *Jindal Journal of Business Research*, 7(2), 103-121.
29. Jerman, M. (2018). The Importance of Intangible Assets in the Hotel Industry: The Case of Croatia and Slovenia. *Scientific Annals of Economics and Business*, 65(3), 333-346. DOI: <http://dx.doi.org/10.2478/saeb-2018-0017>
30. Kelebuh, I., & Javor, A. (1998). Postojeća i nova statistička istraživanja iz područja turizma u Republici Hrvatskoj. *Tourism and hospitality management*, 4 (1), 43-60.
31. Kopal, R., Nemeth, T., & Leinweber, G. (2019). Primjena Benfordovog zakona u otkrivanju računovodstvenih prijevара. *FIP - Financije i pravo*, 7 (2), 7-22.
32. Krajnović, A. Jeličić, N., & Šćiran-Rizner, M. (2020). Digitalni informacijski sustavi i poslovna inteligencija turističke destinacije – primjer eVisitor. *CroDiM: International Journal of Marketing Science*, 3 (1), 113-133.
33. Milićević, M., & Galić, M. (2018). Istraživanje razine neformalnoga gospodarstva u sektoru turizma u federaciji Bosne i Hercegovine. *Mostariensia: časopis za društvene i humanističke znanosti*, 22 (2), 163-190.
34. Newcomb, S. (1881). Note on the Frequency of Use of the Different Digits in Natural Numbers. *American Journal of Mathematics*, 4 (1), 39-40.

35. Nigrini, M. J. (2012). *Benford's Law: Applications for forensic accounting, auditing, and fraud detection*. Hoboken: John Wiley & Sons.
36. Nigrini, M. J. (2011). *Forensic Analytics: Methods and Techniques for Forensic Accounting Investigations*. Hoboken: John Wiley & Sons Inc.
37. Nigrini, M. J., & Miller, S. J. (2009). Data Diagnostics Using Second-Order Test of Benford's Law. *Auditing: A Journal of Practice & Theory*, 28 (2), 305-324.
38. Omerzu, N., & Kolar, I. (2019). Do the Financial Statements of Listed Companies on the Ljubljana Stock Exchange Pass the Benford's Law Test? *International Business Research*, 12 (1), 54-64. <https://doi.org/10.5539/ibr.v12n1p54>
39. Papić, M., Vudrić, N., & Jerin, K. (2017). Benfordov zakon i njegova primjena u forenzičkom računovodstvu. *Zbornik sveučilišta Libertas*, 1-2 (1-2), 153-172.
40. Parroco, A. M., & Vaccina, F. (2004). Estimates of hidden tourism to plan local services: the Sicilian case". *Proceedings of the SCORUS 2004 Conference* (pp. 86-93). Minneapolis: Scorus.
41. Rauch, B., Götttsche, M., Brähler, G., & Kronfeld, T. (2014). Deficit versus social statistics: empirical evidence for the effectiveness of Benford's law. *Applied Economics Letters*, 21 (3), 147-151. <https://doi.org/10.1080/13504851.2013.844319>
42. Shi, J., Ausloos, M., & Zhu, T. (2018). Benford's law first significant digit and distribution distances for testing the reliability of financial reports in developing countries. *Physica A: Statistical Mechanics and its Applications*, 492, C, 878-888. <https://doi.org/10.1016/j.physa.2017.11.017>
43. Shrestha, I. (2016). Validity of Financial Statements: Benford's Law. Retrieved from <https://www.overleaf.com/articles/validity-of-financial-statements-benfordslaw/mrmypkcsbsy>.
44. Slijepčević, S., & Blašković, B. (2014). Statistical detection of fraud in the reporting of Croatian public companies. *Financial Theory and Practice*, 38 (1), 81-96.
45. Stoerk, T. (2015). *Statistical corruption in Beijing's air quality data has likely ended in 2012*. Working Paper No. 194, Grantham Research Institute on Climate Change and the Environment.
46. Stipetić, V. (1998). Plaidoyer za jedinstveni sustav turističkih računa (dileme, teškoće i otvorena pitanja). *Tourism and hospitality management*, 4 (2), 283-293.
47. Stučka, T. (2002). Neslužbeno gospodarstvo u turizmu. *Financijska teorija i praksa*, 26 (1), 173-195.
48. Škrinjarić, B., Recher, V., & Budak, J. (2017). Consumption in the Dark: Estimating Unrecorded Expenditures of Households in Croatia. *Croatian Economic Survey*, 19 (2), 135-167.
49. Vlahović, D. (1998). Inoviranje sustava statistike turizma. *Tourism and hospitality management*, 4 (1), 193-203.
50. Volo, S. (2020). Tourism statistics, indicators and big data: a perspective article. *Tourism Review*, 75 (1), 304-309.
51. Volo, S. (2004). A journey through tourism statistics: accuracy and comparability issues across local, regional and national levels. *Proceedings of the SCORUS 2004 Conference* (pp. 210-216). Minneapolis: Scorus.
52. Volo, S., & Giambalvo, O. (2008). Tourism Statistics: Methodological Imperatives and Difficulties: The Case of Residential Tourism in Island Communities. *Current Issues in Tourism*, 11 (4), 369-380.
53. Vranar, K. (2015). Paušalno oporezivanje iznajmljivanja stanova, soba i postelja putnicima i turistima. *Računovodstvo i porezi u praksi*, (3), 63-71.
54. World Travel and Tourism Council (2020). Croatia 2020 Annual Research: Key Highlights. Retrieved from <https://wttc.org/Research/Economic-Impact/moduleId/704/itemId/96/controller/DownloadRequest/action/QuickDownload>.
55. Wöber, K. W. (2000). Standardizing city tourism statistics. *Annals of Tourism Research*, 27 (1), 51-68.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution - Non Commercial - No Derivatives 4.0 International License.