

ZNAMO LI ŠTA ISTRAŽIVAČKE AGENCIJE PRODAJU: PRIMENA BENFORDOVOG ZAKONA U CILJU PROVERE POUZDANOSTI PODATAKA?

DO WE KNOW WHAT RESEARCH AGENCIES ARE SELLING: APPLICATION OF BENFORD'S LAW FOR DATA RELIABILITY VERIFICATION?

Sanja Mitić*

Jelena Stanojević**

Apstrakt: U toku primenjenog ekonomskog istraživanja korišćenje i analiza podataka dobrog kvaliteta je osnovna pretpostavka koju istraživač treba da obezbedi. Kako se mnogo veća pažnja pridaje samoj analizi postavljenog problema, merenju posmatranih veličina i donošenju zaključaka sprovedenog istraživanja, ovaj rad ima za cilj da skrene pažnju na važnost procesa samog prikupljanja podataka, proveri njihove pouzdanosti i kvaliteta. Naime, kako je u toku tog procesa moguća pojava odgovarajućih grešaka, koje se mogu javiti slučajno, na primer usled pogrešnog zapisivanja ili zaokruživanja podatka, ali i usled namernog podmetanja odgovarajućih fingiranih vrednosti, važno je imati metodu ili postupak koji će smanjiti rizik javljanja takvih nenamernih grešaka i otkriti moguće prisustvo onih namernih. U radu je analizirana mogućnost primene Benfordovog zakona, kao jedne od metoda za otkrivanje iregularnosti, nepravilnosti i moguće manipulativnosti posmatranog skupa podataka. Sam zakon govori o frekventnosti pojavljivanja prvih cifara u numeričkom skupu podataka, koje bi trebalo da slede odgovarajuću logaritamsku zakonitost. Na primeru podataka o vrednosti prodaje različitih brendova i konkretnih jedinica proizvoda, koje je za proizvođača pića obezbedila globalna istraživačka agencija, analiziran je postavljeni problem. Rezultati su predstavljeni za dve evropske zemlje (Češka Republika i Srbija). Cilj rada je prepoznavanje mogućnosti primene Benfordovog zakona na podacima koje istraživačke agencije prodaju korporativnim klijentima. Analizirani su

isti setovi podataka koje su različite filijale jedne globalne istraživačke agencije prodale klijentu u dve različite zemlje. Rezultati će biti diskutovani u kontekstu određivanja pouzdanosti podataka i potencijalnih odstupanja u različitim zemljama.

KLJUČNE REČI: POUZDANOST PODATAKA, KVALITET PODATAKA, GLOBALNE ISTRAŽIVAČKE AGENCIJE, BENFORDOV ZAKON

JEL KLASIFIKACIJA: C10, C80, M30

Abstract: In applied economic research, the use and analysis of good quality data is a basic assumption that the researcher should ensure. As much more attention is paid to the analysis of the problem, the measurement of observed variables and the conclusion of the conducted research, this paper aims to draw attention to the importance of the process of data collection itself, to check their reliability and quality. During that process, some errors can occur, which may occur accidentally, for example, due to incorrect recording or rounding of data, but also due to the intentional planting of fabricated values. Therefore, it is important to have a method or procedure that will reduce the risk of such unintentional occurrences errors and detect the possible presence of intentional ones. The paper analyzes the possibility of applying Benford's law, as one of the methods for detecting irregularity and potential manipulation of the observed data set. The law itself speaks about the frequency of occurrence of the first digits in a numerical data set, which should follow a corresponding

* Ekonomski fakultet, Univerzitet u Beogradu. E-mail: sanja@ekof.bg.ac.rs.

** Ekonomski fakultet, Univerzitet u Beogradu. E-mail: jelena.stanojevic@ekof.bg.ac.rs.

logarithmic law. Using the example of data on the value of sales of different brands and specific product units, which was provided for a beverage manufacturer by a global research agency, the problem was analyzed. The results are presented for two European countries (Czech Republic and Serbia). The goal of the paper is to recognize the possibility of applying Benford's law to the data that research agencies sell to corporate clients. The same data sets sold by different branches

of a global research agency to a client in two different countries were analyzed. The results will be discussed in relation to data reliability and potential deviations in different countries.

KEYWORDS: DATA RELIABILITY, DATA QUALITY, GLOBAL RESEARCH AGENCIES, BANFORD'S LAW

JEL CLASSIFICATION: C10, C80, M30

1. UVOD

U istraživanjima koja se odnose na primenjenu ekonomiju za istraživače je ključno da obezbede bazu kvalitetnih podataka koja će se koristiti u toku istraživanja. Iako je fokus na analizi problema, merenju i zaključku, važno je posvetiti pažnju samom procesu prikupljanja podataka i proveriti njihove pouzdanosti. Tokom ovog procesa mogu se pojaviti greške, bilo slučajne, poput pogrešnog unosa ili zaokruživanja, ili namerne, kroz svesno unošenje lažnih vrednosti. Stoga je neophodno imati metodu koja smanjuje rizik od nenamernih grešaka i otkriva moguće manipulacije. Ovaj rad istražuje primenu Benfordovog zakona kao metode za detekciju nepravilnosti i potencijalnih manipulacija u podacima. To je matematička metoda koja se koristi za otkrivanje nepravilnosti u numeričkim podacima. Njegova primena pomaže u identifikaciji slučajnih grešaka, manipulacija i potencijalnih prevara, što je ključno za sprečavanje negativnih posledica po poslovanje. Ovaj model digitalne analize omogućava brzo i efikasno prepoznavanje sumnjivih unosa i anomalija, posebno u finansijskim podacima. Zakon predviđa učestalost pojavljivanja prvih cifara u numeričkim podacima, gde se cifra 1 pojavljuje najčešće (oko 30%), a cifra 9 najređe (manje od 5%), dok ostale cifre slede opadajući logaritamski obrazac. Naime, američki naučnik Simon Newcomb¹ je 1881. godine je izveo zaključak da se cifra 1 pojavljuje češće kao prva cifra u brojevima, a ovo zapažanje postalo je poznato kao Benfordov zakon, nazvan po fizičaru Franku Benfordu² koji je 1938. godine potvrdio i proširio ovo otkriće. Nigrini je 1996. godine proširio primenu Benfordovog zakona u računovodstvu, reviziji i oporezivanju radi otkrivanja prevara, nakon čega su se mnogi istraživači bavili primenom Benfordovog zakona u ovoj oblasti.³ Benfordov zakon se smatra pouzdanim sredstvom za identifikaciju potencijalnih prevara, pronevera, utaje poreza, kao i grešaka nastalih od strane računovođa ili računarskih sistema, što ga čini popularnim u forenzičkom računovodstvu.

Primena Benfordovog zakona u oblasti marketinga je još uvek u povelju. Kvalitetni poslovni podaci, bilo da je reč o finansijskim ili marketinškim podacima su pretpostavka donošenja i kvalitetnih poslovnih odluka. Način prikupljanja i obrade marketinških podataka je takođe podložan namernim ili nenamernim greškama, zbog čega je neophodno razviti određene kontrolne mehanizme. Takođe, marketinški podaci mogu biti generisani interno ili nabavljeni eksterno. U drugom slučaju postoji još niži stepen kontrole i uvida u

1 Newcomb (1881), str. 39-40

2 Benford (1938), str. 551-572

3 Nigrini (1996), str. 72; Durtschi et al. (2004), str. 17-34; Krakar et al. (2009), str. 39-51; Nigrini et al. (1997); Papić et al. (2017), str. 153-172; Tota et al. (2016), str. 73-77

kvalitet podataka. Cilj ovog rada je utvrđivanje mogućnosti primene Benfordovog zakona u marketingu, na primeru marketinških podataka koje kompanije kupuju na tržištu od istraživačkih agencija, kada ne mogu da kontrolišu proces prikupljanja i obrade dobijenih podataka, a na osnovu kojih donose poslovne odluke. U nedostatku sličnih istraživanja, ova studija će predstaviti prve rezultate provere potencijalnih nepravilnosti i manipulacija podataka koje na tržište, kao svoj komercijalni „proizvod“ plasiraju istraživačke agencije.

Nakon uvodnog dela u radu je prikazana teorijska osnova Benfordovog zakona, a zatim je sistematizovan pregled literature kada je reč o primeni ovog zakona u marketingu. Četvrti deo rada je posvećen opisu metodologije istraživanja, nakon čega su prikazani i diskutovani dobijeni rezultati. Sprovedena je analiza na podacima o vrednostima ostvarene prodaje različitih brendova i proizvoda, koje je globalna istraživačka agencija obezbedila za dve filijale proizvođača pića u Češkoj i Srbiji. U radu je testirana saglasnost podataka koje istraživačke agencije pružaju korporativnim klijentima sa Benfordovom raspodelom. Rezultati će biti razmotreni u kontekstu pouzdanosti podataka i mogućih razlika u dobijenim rezultatima između različitih zemalja.

2. BENFORDOV ZAKON – TEORIJSKA OSNOVA

Benfordov zakon je postao popularan zbog svoje sposobnosti da otkrije greške nastale tokom prikupljanja i beleženja podataka ili usled njihove manipulacije. Zbog toga se smatra jednom od osnovnih i najprimenljivijih metoda za proveru ispravnosti podataka. Važno je napomenuti da neslaganje sa Benfordovim zakonom ne ukazuje nužno na falsifikovane podatke, kao što ni podudaranje sa zakonom ne garantuje odsustvo manipulacije. Stoga se Benfordov zakon koristi kao početni korak, nakon čega sledi detaljnija analiza podataka ili cifara koje odstupaju od očekivane raspodele prema zakonu.

Benfordov zakon se najefikasnije primenjuje na velike skupove podataka koji potiču iz različitih raspodela, gde je aritmetička sredina veća od medijane, a koeficijent spljoštenosti pozitivan⁴. Ovakvi skupovi podataka obično sadrže više manjih nego većih vrednosti. Važno je istaći da se Benfordov zakon ne može primeniti na podatke koji su kodirani (npr. brojevi ličnih karata, bankovnih računa ili brojevi socijalnog osiguranja), na zaokružene brojeve sa specifičnom namenom (poput zaokruživanja cena), niti na brojeve sa prirodnim minimumom ili maksimumom, osim u slučajevima gde je prirodni minimum nula, kao što su rezultati izbora ili broj stanovnika grada. Više o ograničenjima ovog zakona pogledati u radovima Nigrini i Durtschi et al.⁵

U radu Fang & Chen⁶ definisan je pojam značajnosti broja, što je osnova za određivanje slučajne promenljive koja prati raspodelu prema Benfordovom zakonu.

Definicija 1: (Značajnost) Za svaki pozitivan broj $x > 0$, posmatrano u dekadnom sistemu, x se može predstaviti u sledećoj formi:

$$x = S(x) \cdot 10^{k(x)}.$$

4 Wallace (2002), str. 16

5 Nigrini, M. J. (2012); Durtschi et al. (2004), str. 17-34

6 Fang et al. (2020), str. 123-129

gde se $S(x) \in [1,10)$ naziva značajnost broja x , a celobrojno $k(x)$ (obavezno jedinstveno) predstavlja eksponent. Za negativan broj x važi $S(x) = S(-x)$ i po definiciji važi $S(0) = 0$.

Definicija 2: (Benfordov zakon) Realna slučajna veličina X ima Benfordov zakon raspodele ako $\forall s \in [1,10)$ važi:

$$P\{S(X) < s\} = \log s.$$

Posebno, za slučajnu veličinu koja ima Benfordov zakon raspodele važi sledeće:

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

gde je $D_1(X)$ slučajna veličina koja predstavlja cifru na prvoj poziciji sa leve strane slučajne veličine X .

Verovatnoća pojavljivanja cifara na ostalim pozicijama je:

$$P\{D_k(X) = d_k\} = \sum_{d_1=1}^9 \sum_{d_2=0}^9 \dots \sum_{d_{k-1}=0}^9 \log\left(1 + \frac{1}{\sum_{i=1}^k 10^{k-i} d_i}\right),$$

gde je $D_k(X)$ slučajna veličina koja predstavlja cifru na k -toj poziciji slučajne veličine X i d_k je cifra na toj poziciji, $d_k \in \{0,1,2, \dots, p\}$. Verovatnoća pojavljivanja cifara na višim pozicijama je uniformno raspodeljena i iznosi 0.1. U sledećoj tabeli je prikazana verovatnoća pojavljivanja odgovarajuće cifre na različitim pozicijama posmatranog broja.

Tabela 1: Verovatnoća pojavljivanja odgovarajuće cifre na različitim pozicijama posmatranog broja (u procentima)

Cifra	1. poz.	2. poz.	3. poz.	4. poz.	5. poz.
0	-	11.968	10.178	10.018	10.000
1	30.103	11.389	10.138	10.014	10.000
2	17.609	10.882	10.097	10.010	10.000
3	12.494	10.433	10.057	10.006	10.000
4	9.691	10.031	10.018	10.002	10.000
5	7.918	9.668	9.979	9.998	10.000
6	6.695	9.337	9.940	9.994	10.000
7	5.799	9.035	9.902	9.990	10.000
8	5.115	8.757	9.864	9.986	10.000
9	4.576	8.500	9.827	9.982	10.000
	100.000	100.000	100.000	100.000	100.000

Izvor: Račun autora

3. PRIMENA BENFORDOVOG ZAKONA U MARKETINGU

Benfordov zakon u poslovnoj ekonomiji ima najintenzivniju primenu u oblasti računovodstvene forenzike.⁷ U oblasti marketinga primena Benfordovog zakona, kao alata u funkciji detekcije potencijalnih nepravdnosti i zloupotreba je još uvek u pvoju. Kao najznačajnije oblasti marketinga u kojima je istaknuta primena Benfordovog zakona ističu se psihološke cene, onlajn marketing i digitalna marketinška forenzika i tržišna istraživanja.

Primena Benfordovog zakona u oblasti digitalnog marketinga je vrlo izražena u poslednjih deset godina. Koristi se u cilju utvrđivanja potencijalnih manipulacija kada je reč o podacima o web saobraćaju, onlajn metrici, generisanju sadržaja, kao što su komentari i recenzije i slično. Studije Golbeck i Striga & Podobnik, pokazale su da se podaci o broju pratilaca i objava na društvenim mrežama ponašaju u skladu sa Benfordovim zakonom.⁸ Golbeck je pratila bazične podatke sa društvenih mreža, o broju prijatelja ili pratilaca i broju objava korisnika, na pet društvenih mreža (Facebook, Twitter, Google Plus, Pinterest, and LiveJournal).⁹ Navedene društvene i bihevioralne karakteristike korisnika ovih društvenih mreža ne pokazuju odstupanje od Benfordovog zakona. Odstupanja nisu identifikovana u većini slučajeva, ni na nivou posmatranja pojedinačne društvene mreže, niti na nivou veza koje pojedinac razvija na ovim platformama.

Značaj primene Benfordovog zakona na ovakvim podacima se ogleda u mogućnosti identifikacije neobičnog ponašanja korisnika društvenih mreža, ponašanja koje može odstupati od takozvanog „prirodnog“ ponašanja. U ovom kontekstu Benfordov zakon je prema Golbeck alat koji treba da pomogne analizi i razumevanju ponašanja korisnika društvenih mreža, lakše identifikacije situacija i razloga zbog kojih dolazi do devijacija, razumevanja takozvanih „suptilnih obrazaca prirodnog ponašanja koje ljudi mogu teško da repliciraju u neprirodnim okolnostima.“¹⁰

Interesantna primena ovog zakona je u oblasti analize sadržaja koji korisnici generišu. Tako su se autori Zhao & Wang bavili analizom recenzija i komentara korisnika na veb-sajtovima velikih platformi za onlajn kupovinu, naručivanje i platformi za ostavljanje recenzija.¹¹ Recenzije koje se pojavljuju na sajtovima su često pod velom sumnje da su predmet manipulacije, s jedne strane, dok u isto vreme predstavljaju važan informativni sadržaj za korisnike usluga onlajn platformi. Razvijani su različiti metodi za detekciju autentičnosti komentara i recenzija korisnika, ali oni nisu u stanju da prepoznaju sistematsku manipulaciju do koje može doći.¹² Zato autori predlažu primenu Benfordovog zakona i na uzorku recenzija sa četiri velike onlajn platforme rezultati ukazuju na manipulaciju onlajn komentarima i recenzijama, na svim analiziranim platformama, kao i da je na osnovu poređenja rezultata uočeno da platforme koje dolaze iz zrelih tržišta pokazuju manji nivo odstupanja, jer su razvijeni mehanizmi sprečavanja sistematske manipulacije recenzijama. Takođe, pokazano je da u slučaju pozitivnih i ekstremno pozitivnih onlajn komentara

7 Kinanti et al. (2023), str. 462–485

8 Golbeck (2015), str. 1-10 & Striga et al. (2018), str. 14629–14642

9 Golbeck (2015), str. 1-10

10 Golbeck (2015), str. 1-10

11 Zhao et al. (2023), str. 365–406

12 Zhao et al. (2023), str. 365–406

postoji značajno odstupaju od Benfordove raspodele. Kada je reč o generisanju sadržaja, korišćen je Benfordov zakon u funkciji detekcije onlajn automatski generisanih sadržaja, botova na društvenim medijima koji se koriste za kreiranja malicioznog onlajn sadržaja.¹³

Primena Benfordovog zakona u oblasti istraživanja tržišta je vrlo obimna. Podaci koji su prikupljeni nekom od istraživačkih tehnika, najčešće predstavljaju vrednu podlogu za donošenje odluka, bilo da je reč o poslovnim odlukama ili odlukama povodom javnih politika. U tom smislu, kvalitet podataka predstavlja važan preduslov i za donošenje kvalitetnih odluka. Podaci koji su prikupljeni terenskim istraživanjem mogu pokazati odstupanje u odnosu na ovaj zakon bilo zbog metodoloških grešaka, bilo zbog namerne zloupotrebe i manipulacija. Iako Benfordov zakon ukazuje na jednu dimenziju kvaliteta podataka, na dimenziju pouzdanosti, treba imati na umu da je kvalitet podataka širi pojam. Kvalitet podataka se odnosi na njihovu sposobnost da odgovore zahtevima korisnika, u specifičnom kontekstu, bilo da je reč o poslovnoj ili nekoj drugoj upotrebi.¹⁴ Dimenzije kvaliteta podataka koje se ističu u literaturi su kompletnost, preciznosti, relevantnost, konzistentnosti, aktuelnost, pouzdanost,¹⁵ s tim da i one zavise od upotrebe i konteksta. Do ugrožavanja kvaliteta podataka može doći u bilo kojoj fazi takozvanog životnog ciklusa podatka, od pripreme za prikupljanje podataka, tokom samog procesa prikupljanja, zatim unosa i analize podataka, čuvanja i deljenja podataka.¹⁶

Brojni autori ukazuju da Benfordov zakon može biti vredan alat u kontroli pouzdanosti, kao jedne dimenzije kvaliteta podataka koji su prikupljeni anketnim istraživanjem. Judge & Schechter su pokazali da veća odstupanja od Benfordovog zakona postoje u slučaju podataka prikupljenih anketnim istraživanjem u manje razvijenim zemljama u odnosu na slične setove podataka u SAD-u, kao primeru razvijene zemlje u njihovom uzorku.¹⁷ Takođe, ukazano je na pojavu da podaci koji su prikupljeni pod supervizijom akademskih istraživača pokazuju manju distorziju u odnosu na podatke koje su prikupljale državne institucije ili istraživačke agencije.

Kaiser je pokazao da podaci prikupljeni anketnim istraživanjem pokazuju nizak stepen pouzdanosti, sa izuzetkom podataka o dohotku.¹⁸ Podaci o dohotku pokazuju usaglašenost sa Benfordovom raspodelom, kako na agregatnom nivou (GDP), tako i kada je reč o podacima iz anketnih upitnika. U slučaju podataka prikupljenih anketom, autori su utvrdili veću pouzdanost podataka o dohotku na nivou domaćinstva u odnosu na isti podatak na individualnom nivou, zbog čega preporučuju da se kada god je to moguće postavljaju pitanja i prikupljaju podaci na nivou domaćinstva. Kundt je koristio Benfordov zakon u cilju identifikacije najboljeg metoda prikupljanja osetljivih podataka putem ankete, kao što su na primer podaci o utaju poreza.¹⁹ Uspešno korišćenje Benfordovog zakona u funkciji identifikacije prevara i grešaka pokazano je na primeru socio-ekonomskog panela u Ne-

13 Madahali et al. (2020), str. 1-8; Mbona et al. (2022), str. 369-381

14 Ehrlinger et al. (2022), str. 1-30

15 Mahanti (2018); Batini et al. (2006), str. 19-49; Wang et al. (2024), str. 1159–1178

16 Mahanti (2018)

17 Judge et al. (2009), str. 1–24

18 Kaiser (2019), str. 1602–1618

19 Kundt (2014), str. 1-15

mačkoj, tako što je u jedan uzorak ubačeno nekoliko fabrikovanih intervjua, što je dovelo do odstupanja u odnosu na ovaj zakon.²⁰ Takođe, koristeći navedeni metod, uspešno je identifikovano 10 falsifikovanih intervjua od strane jednog anketara, što u prethodnoj terenskoj proveru nije bilo prepoznato. Zapravo, ovom metodologijom ukazano je na sumnju u određeni set podataka, što je naknadnom analizom i pozivima navodnim ispitanicima provereno i tako potvrđeno da je reč o falsifikovanim intervjuima.

Kao poseban segment podataka koji su važni u marketinškim istraživanjima su podaci o prodaji proizvoda i usluga. Na primeru podataka o prodaji vojne opreme na inostranim tržištima, na osnovu faktura, pokazano je da navedeni podaci, tj. obrasci kupovine ne odstupaju od Benfordovog zakona, i kako autori naglašavaju zakon predstavlja sredstvo za razumevanje kupovnog ponašanja.²¹ Veliki broj kompanija eksterno nabavlja podatke o ostvarenoj prodaji različitih proizvodnih jedinica, po različitim prodajnim kanalima i maloprodajnim objektima. Provajderi tih podataka, istraživačke agencije, često ne nude garanciju da je reč o kvalitetnim podacima, pre svega u pogledu preciznosti i pouzdanosti. S druge strane, reč je o vrednim podacima, na osnovu kojih menadžeri donose poslovne odluke i za čiju nabavku su opredeljenje značajne stavke u korporativnim budžetima. U ovom radu, u cilju provere pouzdanosti ovakvih podataka koristićemo Benfordov zakon i metodologiju koja će biti opisana u nastavku.

4. METODOLOGIJA

Podaci koji su predmet analize su dobijeni od multinacionalne kampanje koja posluje u prehrambenoj industriji, i ima filijale u većem broju zemalja u SAD-u i Evropi. Predmet analize će biti dve serije podataka iz Srbije i Češke, o vrednostima prodaje različitih brendova i konkretnih jedinica proizvoda, koja je ostvarena u različitim maloprodajnim kanalima. Reč je o podacima koje je za potrebe kompanije obezbedila globalna istraživačka agencija, kao eksterni provajder, koja takođe ima svoje filijale u pomenutim zemljama. Svaka filijala je nezavisno jedna od druge prikupila i pripremila podatke za naručioca. Dobijene baze podataka sadrže 480065 unosa za tržište Češke i 12359 unosa za tržište Srbije. Značajne razlike u veličini baza za dve zemlje su rezultat različitog proizvodnog asortimana po zemljama i broja raspoloživih distributivnih kanala.

Prilikom testiranja saglasnosti sa Benfordovim zakonom, nulta hipoteza je tvrđenje koje govori o tome da su podaci u saglasnosti sa Benfordovom raspodelom, dok alternativna hipoteza tvrdi da nisu. Sve odluke su u daljem tekstu donete na nivou značajnosti od 5%. U radu se ispituje saglasnost sa Benfordovim zakonom pomoću dva testa, koji su već predloženi u literaturi (Nigrini i Costa et al.²²): Hi-kvadrat test i test zasnovan na srednjoj apsolutnoj devijaciji (skraćeno MAD test (mean absolute deviation test)). Treba istaći sličnost između ovih testova, oba testa daju odgovor za sve cifre odjednom (simultano), dok je razlika između njih to što MAD test ne zavisi od broja podataka koji se ispituje, dok Hi-kvadrat test zavisi, tako da se on može primeniti i daje pouzdan odgovor i kada je posmatrani skup podataka veoma veliki, kao što je slučaj u ovom radu.

20 Schräpler (2011), str. 685-718

21 Tichenor et al. (2008), str. 77-86

22 Nigrini (2012), str. 153-160; Costa (2013), str.16-30

Prvi test koji smo koristili u ovom radu, u cilju testiranja saglasnosti podataka sa Benfordovom raspodelom, je Hi-kvadrat test. Ovim testom se ispituje da li ukupna raspodela ostvarenih frekvencija prve cifre odstupa od raspodele očekivanih frekvencija koja je u skladu sa Benfordovim zakonom, tako da test testira sve cifre odjednom. Odgovarajuća statistika testa se računa pomoću sledeće formule:

$$\chi^2 = \sum_{i=1}^9 \frac{(O_i - E_i)^2}{E_i} = n \sum_{i=1}^9 \frac{(p_{oi} - p_i)^2}{p_i},$$

gde je O_i ostvarena frekvencija cifre i , E_i je očekivana frekvencija cifre prema Benfordovom zakonu ($E_i = np_i$), p_{oi} je ostvarena procentualna frekvencija cifre i i p_i je očekivana procentualna frekvencija cifre i u skladu sa Benfordovim zakonom, n je broj opservacija posmatrane promenljive. Kritična vrednost iznosi 15.51 (za 8 stepeni slobode) i ukoliko je vrednost statistike Hi-kvadrat testa veća od ove kritične vrednosti, nulta hipoteza se odbacuje, tj. što je statistika testa veća to je veće odstupanje od Benfordove raspodele.

Drugi test koji smo u ovom radu primenili za testiranje saglasnosti raspodele posmatranih podataka sa Benfordovim zakonom jeste test srednjeg apsolutnog odstupanja (MAD test). Formula po kojoj se računa statistika testa je sledeća:

$$MAD = \frac{1}{n} \frac{\sum_{i=1}^K |O_i - E_i|}{K} = \frac{\sum_{i=1}^K |p_{oi} - p_i|}{K},$$

gde su O_i , E_i , p_{oi} , p_i , i i n već uvedene promenljive, dok je K broj cifara koji se posmatra. Nigrini²³ je sugerisao kritične vrednosti za slaganje, prihvatljivo slaganje, marginalno prihvatljivo slaganje i neslaganje, na osnovu svog ličnog iskustava, pa ta činjenica da ne postoji jasno definisana kritična vrednost ovog testa, smatra se osnovnim njegovim nedostatkom. U tabeli 2 su dati rangovi odgovarajućih kritičnih vrednosti.

Tabela 2: Rangovi MAD kritičnih vrednosti

Rang slaganja	Prva cifra
Blisko slaganje	0.000-0.006
Prihvatljivo slaganje	0.006-0.012
Marginalno prihvatljivo slaganje	0.012-0.015
Neslaganje	Iznad 0.015

Izvor: Nigrini, 2012

5. REZULTATI

Hi-kvadrat test pokazuje neslaganje sa Benfordovim zakonom za obe zemlje (tabela 3), jer su odgovarajuće test statistike veće od kritične vrednosti, koja iznosi 15.51. I to, za

²³ Nigrini (2012), str. 158-160

Srbiju je ta statistika znatno veća, što ukazuje na to da podaci značajno više odstupaju od Benfordovog zakona u odnosu na podatke za Češku. Dobijeni rezultati MAD testa ukazuju na mogućnost postojanja anomalija u podacima samo za Srbiju, jer je test statistika MAD testa za te podatke u rangu neslaganja, dok je test statistika MAD testa za podatke dobijene iz filijale u Češkoj u rangu bliskog slaganja, prema tabeli 2.

Mogući razlozi za odstupanja od Benfordove raspodele su sledeći: prvi, ako su podaci ručno uneti, mogu biti modifikovani ili zaokruženi na određene vrednosti, čime se narušava priroda raspodele i slaganje sa zakonom; drugi, Benfordova raspodela najbolje funkcioniše kada podaci pokrivaju više redova veličine, tj. ako su brojevi ograničeni na uski interval (npr. svi su između 100 i 999) raspodela prve cifre neće odgovarati Benfordovom zakonu; treći, administrativni faktori, kao na primer propisi koji mogu dovesti do toga da su određeni brojevi češći, npr. cene proizvoda, mogu dovesti do nesrazmere između očekivane i stvarne raspodele; i četvrti, ovde najočigledniji razlog za ovakav rezultat jeste, ako je veličina uzorka manja, slučajna varijacija može uzrokovati da raspodela ne prati Benfordov zakon, što je slučaj za Srbiju u odnosu na Češku.

Tabela 3: Statistike MAD testa i Hi-kvadrat testa za dve zemlje

Test	Češka	Srbija
MAD test	0.0007	0.023
Hi-kvadrat test	29.61	907.66

Izvor: Račun autora

Rezultati ukazuju na mogućnost manipulacije podacima koju treba detaljnije istražiti u cilju izvođenja jasnog zaključka. Kako je reč o podacima koji se očitavaju na kasama u maloprodajnim objektima, koje od maloprodavaca otkupljuje istraživačka agencije, zatim ih sistematizuje, po potrebi vrši ekstrapolaciju za nedostajuće pozicije i dalje distribuiraju, do anomalije je moglo doći u bilo kojoj fazi životnog ciklusa podatka. U ovom slučaju pouzdanost podatka može da zavisi i od kvaliteta maloprodajne infrastrukture, koja može značajno odstupati po zemljama, pre svega u delu načina na koji se očitavaju i čuvaju podaci sa POS terminala. Drugi uzrok može biti način na koji je urađeno objedinjavanje i sistematizacija dobijenih podataka od strane istraživačke agencije, kao i načina i stepena u kojem je bilo potrebno upotpuniti pozicije koje nedostaju. Ovakav rezultata odgovara jednom od zaključaka studije Judge-a i Schechter-a, koji su identifikovali veća odstupanja od Benfordovog zakona u slučaju manje razvijenih u odnosu na visoko razvijene zemlje.²⁴ Konkretni uzrok identifikovanih razlika u pouzdanosti podataka iz Češke i Srbije je potrebno utvrditi naknadnim analizama. Menadžeri mogu koristiti navedenu metodologiju, kao jedan kontrolni mehanizam pouzdanosti podataka na osnovu kojih donose poslovne odluke. Ovakva kontrola je od posebnog značaja u slučaju postojanja eksternih provajdera podataka, ili jednog raspoloživog provajdera podatak, kada je moguća i zloupotreba dominantnog položaja na tržištu.

²⁴ Judge et al. (2009), str. 1–24

6. ZAKLJUČAK

U radu je primenjen Benfordov zakon na podatke o prodaji koje je globalna istraživačka agencija obezbedila multinacionalnoj kompaniji, kao svom klijentu. Pouzdanost tih podataka je testirana i testovi ukazuju na mogućnosti manipulacije podataka, s tim što je veća pouzdanost podataka u slučaju Češke u odnosu na Srbiju. Ova studija ima značaj za marketing menadžere, koji mogu koristiti navedenu metodologiju kako bi povećali stepen kontrole kvaliteta, kako interno generisanih podataka, tako i podataka koje kupuju od eksternih provajdera. Od posebnog je značaja u situaciji kada provajder podataka ne obezbeđuje garanciju kvaliteta podataka koje prodaje, kao što je to slučaj u primeru koji je analiziran. Takođe, na ovaj način marketing menadžeri mogu da redefinišu način na koji ugovaraju nabavku podataka i svoju pregovaračku poziciju u odnosu na globalne provajdere podataka. Korišćenje Benfordovog zakona predstavlja i mogućnost za istraživačke agencije, u cilju identifikacije metodoloških ili drugih manjkavosti tokom takozvanog životnog ciklusa podatka.

U ovoj studiji su prvi put testirani ovakvi setovi podataka, koji su predmet komercijalne transakcije između globalne istraživačke agencije i korporativnog klijenta, i ona predstavlja vredan doprinos akademskoj marketinškoj i statističkoj literaturi. U tom smislu postoji i značajan prostor za dalja istraživanja, pre svega za primenu dodatnih statističkih testova, kao i razvoj novih i uključivanje i testiranje podataka iz filijala iz većeg broja zemalja, kako bismo mogli izvesti jasniji zaključak o potencijalnim razlikama između visoko i manje razvijenih zemalja.

Ograničenja ove studije se odnose na neujednačenu veličinu baza podataka za dve zemlje, kako broj podataka može i ne mora uticati na rezultate primenjenih testova. Naime, broj podataka koji se ispituje ne utiče na rezultate MAD testa, dok ima uticaja na rezultate Hi-kvadrat testa. Treba istaći da se Hi-kvadrat test može primeniti i u situaciji sa velikim skupom podataka i da daje pouzdan odgovor i u tom slučaju. U ovom radu, dostupna baza znatno je veća za Češku nego za Srbiju. Iako Hi-kvadrat statistike za obe zemlje pokazuju neslaganje sa Benfordovim zakonom, dobijena statistika za Češku znatno je manja, i sa promenom nivoa značajnosti (koji je u radu unapred određen na 5%) odluka testa bi mogla da bude i drugačija.

LITERATURA

1. Batini, C. and Scannapieca, M. (2006), *Data Quality: Concepts, Methodologies and Techniques*, Springer, Berlin.
2. Benford, F. (1938), "The law of anomalous numbers", *Proceedings of the American philosophical society*, Pp. 551-572.
3. Cerqueti, R., et al. (2024), "Statistical methods for decision support systems in finance: how Benford's law predicts financial risk", *Annual Operation Research*, Vol. 342, Pp. 1445–1469, doi: 10.1007/s10479-022-04742-z.
4. Costa J., et al. (2013), "Application of Newcomb-Benford law in accounting audit: a bibliometric analysis in the period from 1988 to 2011", *10th International Conference on Information Systems and Technology Management-CONTECSI*, June, 12 to 14, 2013, São Paulo, Brazil.

5. Durtschi, C., et al. (2004), "The effective use of Benford's law to assist in detecting fraud in accounting data", *Journal of forensic accounting*, Vol. 5, Pp. 17-34.
6. Ehrlinger, L. and Wöß, W. (2022), "A Survey of Data Quality Measurement and Monitoring Tools", *Frontiers in Big Data*, Vol. 5, Pp. 1-30, doi: 10.3389/fdata.2022.850611.
7. Fang, G. and Chen, Q. (2020), "Several common probability distributions obey Benford's law", *Physica A: Statistical Mechanics and its Applications*, Vol. 540, 123129.
8. Golbeck, J. (2015), "Benford's Law Applies to Online Social Networks" *PLoS ONE*, Vol. 10, Pp. 1-10, doi: 10.1371/journal.pone.0135169.
9. Judge, G. and Schechter, L. (2009), "Detecting Problems in Survey Data Using Benford's Law", *Journal of Human Resources*, Vol. 44, Pp. 1–24. doi: 10.3368/jhr.44.1.1.
10. Kaiser, M. (2019), "Benford's law as an indicator of survey reliability—can we trust our data?" *Journal of Economic Surveys*, Vol. 33, Pp. 1602–1618. doi: 10.1111/joes.12338.
11. Kinanti, K. P. et al. (2023), "Forensic Accounting, Preventing And Detecting Fraud: A Systematic Literature", *Journal of Accounting and Finance Management*, Vol. 4, Pp. 462–485, doi: 10.38035/jafm.v4i6.284.
12. Krakar, Z. and Žgela, M. (2009), "Application of Benford's law in payment systems auditing", *Journal of Information and Organizational Sciences*, Vol. 33, Pp. 39-51.
13. Kundt, T. (2014), "Applying „Benford's law" to the Crosswise Model: Findings from an online survey on tax evasion", Helmut Schmidt Universität Hamburg, Department of Economics, Working Paper No. 148, July, 2014.
14. Madahali, L. and Hall, M. (2020), "Application of the Benford's law to Social bots and Information Operations activities", *2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)*, IEEE, 15-19 June 2020, Dublin, Ireland.
15. Mahanti, R. (2018), *Data Quality: Dimensions, Measurement, Strategy, Management, and Governance*, ASQ Press, Milwaukee.
16. Mbona, I. and Eloff, J. (2022), "Feature selection using Benford's law to support detection of malicious social media bots", *Information Sciences*, Vol. 582, Pp. 369-381, doi: 10.1016/j.ins.2021.09.038.
17. Newcomb, S. (1881), "Note on the frequency of use of the different digits in natural numbers", *American Journal of mathematics*, Vol. 4, Pp. 39-40.
18. Nigrini, M. J. (1996), "A taxpayer compliance application of Benford's law", *The Journal of the American Taxation Association*, Vol. 18, Pp. 72.
19. Nigrini, M. J. (2012), *Benford's Law: Applications for forensic accounting, auditing, and fraud detection*, John Wiley & Sons, New Jersey.
20. Nigrini, M. J. and Mittermaier, L. J. (1997), "The use of Benford's Law as an Aid in Analytical Procedures", *Auditing: A Journal of Practice & Theory*, Vol. 16, Pp. 52.
21. Papić, M., et al. (2017), "Benfordov zakon i njegova primjena u forenzičkom računovodstvu", *Zbornik sveučilišta Libertas*, Vol. 1, Pp. 153-172.
22. Schräpler, J. (2011), "Benford's Law as an Instrument for Fraud Detection in Surveys Using the Data of the Socio-Economic Panel (SOEP)", in Diekmann (ed.), *Methodological Artefacts, Data Manipulation and Fraud in Economics and Social Science: Themenheft 5+6/Bd. 231(2011) Jahrbücher für Nationalökonomie und Statistik*, De Gruyter Oldenbourg, Berlin, Boston, Pp. 685-718. doi: 10.1515/9783110508420-008.

23. Striga, D. and Podobnik, V. (2018), “Benford’s Law and Dunbar’s Number: Does Facebook Have a Power to Change Natural and Anthropological Laws?” *IEEE Access*, Vol. 6, Pp. 14629–14642. doi:10.1109/access.2018.280571.
24. Tichenor, C. and Davis, B. (2008), “The Applicability of Benford’s Law to the Buying Behavior of Foreign Military Sales Customers”, *Global Journal of Business Research*, Vol. 2, Pp. 77-86.
25. Tota, I., et al. (2016), “The use of Benford’s law as a tool for detecting fraud in accounting data”, *Interdisciplinary Journal of Research and Development*, Vol. 3, Pp. 73-77.
26. Wallace, W. A. (2002), “Assessing the quality of data used for benchmarking and decision-making”, *The Journal of Government Financial Management*, Vol. 51, Pp. 16-22.
27. Wang, J., et al. (2024), “Overview of Data Quality: Examining the Dimensions, Antecedents, and Impacts of Data Quality”, *Journal of Knowledge Economy*, Vol. 15, Pp. 1159–1178. doi: 10.1007/s13132-022-01096-6.
28. Zhao, C. and Wang, C.A. (2023), “A cross-site comparison of online review manipulation using Benford’s law”, *Electronic Commerce Research*, Vol. 23, Pp. 365–406, doi: 10.1007/s10660-020-09455-8.