МАТЕМАТИЧКИ ОМНИБУС, 3 (2018), 117 – 143
http://im-pmf.weebly.com/matematicki-omnibus-kniga-3.html

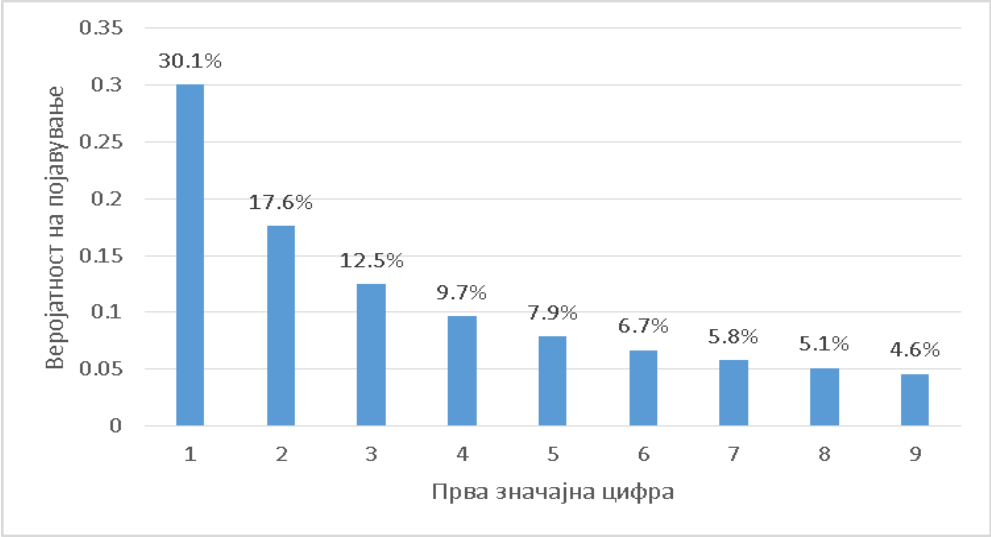
ЗАКОНОТ НА БЕНФОРД ЗА ПРВАТА ЗНАЧАЈНА ЦИФРА

Ирена Стојковска [1]

Една од убавините на математиката е нејзината универзалност, односно способноста навидум различни појави од различни области да ги опишува на еден ист начин. Познат пример за универзалноста на математиката е *Централната гранична теорема* во теоријата на веројатност според која величините кои се збир од голем број величини со незначителни поединечни влијанија, се приближно нормално распределени, како на пример, висината на луѓето, електронскиот шум, коефициентот на интелегенција (IQ), поените на тестовите и слично. Друг пример за универзалноста на математиката е *законот на Бенфорд за првата значајна цифра*, кој последните години сосема заслужено добива сè поголемо внимание. Накратко, овој закон тврди дека кај броевите кои „потекнуваат од природата“, многу поверојатно е првата значајна цифра да е помал број, отколку поголем, наспроти очекуваната нејзина рамномерна распределба. *Првата значајна (децимална) цифра* на еден реален број x се дефинира како (единствен) цел број $j \in \{1,2,...,9\}$ за кој важи $10^k\, j \le |x| < 10^k\,(j+1)$ за некој (единствен) $k \in \mathbb{Z}$, [3]. На пример, првата значајна цифра на бројот $\sqrt{2} = 1.41421356...$ е цифрата 1, додека првата значајна цифра на бројот $\frac{1}{\sqrt{2}} = 0.70710678...$ е цифрата 7, а првата значајна цифра на бројот $\frac{2}{300} = 0.00666666...$ е цифрата 6. За да го искажеме попрецизно законот на Бенфорд, со D_1 ја означуваме првата значајна цифра на бројот, при што D_1 е случајна променлива, па *законот на Бенфорд* вели дека за броевите кои „потекнуваат од природата“, веројатноста првата значајна цифра да е d е еднаква на:

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

Ова значи дека во 30,1% од случаите, првата значајна цифра е 1, во 17,6% од случаите е 2, па сè така до 4,6% од случаите таа е 9 (Слика 1).



Слика 1. Распределба на првата значајна цифра кај колекциите од броеви кои „потекнуваат од природата“, според законот на Бенфорд.

Многу научници низ историјата се обидувале емпириски да ја покажат универзалноста на овој закон, т.е дека важи за која било колекција од нумерички податоци, правејќи трансформации и калкулации со оние податоци за кои директната примена на законот се покажало дека не е можна. Така биле откриени „природни“ колекции од нумерички податоци за кои важи овој закон, но биле откриени и такви за кои не важи. Многумина, пак, се обидувале математички да ја докажат точноста на законот и со тоа да ја потврдат неговата универзалност. Сите овие обиди се успешни помалку или повеќе и секој од нив има дадено свој придонес кон разбирањето на овој феномен. Од друга страна, примената на законот на Бенфорд во најразлични дисциплини како сметководството, компјутерските науки, динамичките системи, економијата, инженерството, медицината, теоријата на броеви, психологијата, веројатноста и статистиката, зборува за вистинското значење на универзалноста на овој закон.

Во продолжение, прво ќе зборуваме за откривањето на законот на Бенфорд, а потоа ќе презентираме дел од емпириските резултати со цел да ја воведеме потребата од математичката теорија која го објаснува законот и условите при кои тој важи. На крајот, ќе изложиме некои примени на овој закон.

1. ОТКРИВАЊЕ

Иако законот на Бенфорд е именуван според американскиот физичар Франк Бенфорд (1883-1948), првиот печатен труд во врска со оваа законитост датира од 1881 година, кога американскиот астроном Сајмон Њукомб (1835-1909) го објавил својот труд *Note on the frequency of use of the different digits in natural numbers* (прев. *Забелешка за честотата на употреба на различни цифри во броевите од природата*) во списанието American Journal of Mathematics на само две страници, [12]. Њукомб во тоа време бил признат астроном, кој работел на планетарни теории и изведување на астрономски константи. Тој не бил математичар во строга смисла на зборот, но и самата интуитивна природа на законот на Бенфорд, не е за чудење што повеќе ги привлекувала физичарите и научниците од сродни области, отколку математичарите. Во својот труд Њукомб дури изразува и чудење, како ваква појава не била и порано откриена. Но, како дошол тој до тоа откритие?

Набљудувајќи ја книгата со логаритамски таблици, многупати позајмувана од библиотеката, Њукомб забележал дека почетните страници биле многу повеќе користени, отколку средните, а крајните најмалку. Ако знаеме дека логаритамските таблици ги содржат мантисите на логаритмите на броевите подредени во растечки редослед, тогаш јасен е заклучокот на Њукомб кој вели дека кај броевите од природата „првата значајна цифра е почесто 1, отколку која било друга цифра и честотата опаѓа одејќи кон 9“. *Мантиса* (во декаден систем) на еден реален број x се дефинира како единствениот реален број $r \in [\frac{1}{10},1)$ за кој $x = r10^n$ за некој $n \in \mathbb{Z}$, [5]. Така, мантиса на бројот $\sqrt{2}$ е 0.141421356..., мантиса на бројот $\frac{1}{\sqrt{2}} = 0.70710678...$ е 0.70710678..., додека мантиса на бројот $\frac{2}{300} = 0.00666666...$ е 0.666666.... Њукомб исто така забележал дека „законот на веројатноста на појавување на броевите е таков што мантисите на нивните логаритми се еднаквоверојатни“. Оваа своја забелешка, Њукомб не ја запишал преку формула, но ги пресметал веројатностите за појавување на првата и втората значајна цифра кај броевите „од природата“ (види Слика 2). Може да се покаже дека оваа забелешка на Њукомб ја повлекува распределбата на првата значајна цифра, дадена со (1), [7]. По пресметките на веројатностите на појавување на првата и

втората значајна цифра, Њукомб забележува дека „за третата значајна цифра распределбата на веројатностите на појавување е скоро еднаква за сите цифри, а од четвртата цифра натаму, разликата е занемарлива.“

Dig.	First Digit.	Second Digit.
0 . . .	. . .	0.1197
1 . . .	0.3010	0.1139
2 . . .	0.1761	0.1088
3 . . .	0.1249	0.1043
4 . . .	0.0969	0.1003
5 . . .	0.0792	0.0967
6 . . .	0.0669	0.0934
7 . . .	0.0580	0.0904
8 . . .	0.0512	0.0876
9 . . .	0.0458	0.0850

Слика 2. Веројатности на појавување на првата и втората значајна цифра, пресметани од Њукомб (1881), [12].

Второто важно согледување кое го забележал Њукомб е дека нумеричките вредности на физичките величини во голема мера зависат од изборот на единицата мерка, па затоа Њукомб препорачува броевите „да се земаат како односи на количини“. На тој начин би биле неименувани броеви, односно броеви независни од мерните единици, а појавата во врска со распределбата на првата значајна цифра ќе биде независна од изборот на единицата мерка (својство на скаларна инваријантност). На крајот Њукомб дава забелешка за примената на ова сознание, имено тој вели дека „овој закон би ни овозможил да одлучиме дали некоја голема колекција од независни нумерички резултати е составена од броеви од природата или логаритми“.

Од досега кажаното, може да се заклучи дека Њукомб изложил важни согледувања за овој феномен, од емпириски, интуитивен и применлив аспект. Тоа што недостасувало била математичката теорија на која би се изградило посолидно објаснување за оваа појава, како и оправдување, зошто е важно изучувањето на оваа појава. Но затоа, Бенфорд во 1938 година, иако не презентирал солидна математичка теорија за оваа појава, прв дал понапредно објаснување зошто првите значајни цифри ја имаат таа распределба и презентирал оправдување зошто оваа појава вреди да се истражува, [2].

Педесетина години подоцна и Бенфорд, како Њукомб, најнапред приметил дека почетните страници на логаритамските таблици се повалкани (од користење), отколку крајните. Потоа, расудувајќи каде сè се користат овие таблици, тој се одлучил да ја провери распределбата на првата значајна цифра кај 20 колекции од нумерички податоци, различни по потекло, вкупно 20229 податоци, како на пример должини на реки, големини на области, број на население, физички константи, низи од броеви, но и некои чудни колекции од податоци, како сите броеви кои се среќаваат на страниците во еден број од списанието *Reader Digest*, броевите од првите 342 адреси на улици од списанието *American Men of Science* итн. Резултатите од неговата проверка се дадени на Слика 3.

Title	1	2	3	4	5	6	7	8	9	Count
Rivers, Area	31.0	16.4	10.7	11.3	7.2	8.6	5.5	4.2	5.1	335
Population	33.9	20.4	14.2	8.1	7.2	6.2	4.1	3.7	2.2	3259
Constants	41.3	14.4	4.8	8.6	10.6	5.8	1.0	2.9	10.6	104
Newspapers	30.0	18.0	12.0	10.0	8.0	6.0	6.0	5.0	5.0	100
Spec. Heat	24.0	18.4	16.2	14.6	10.6	4.1	3.2	4.8	4.1	1389
Pressure	29.6	18.3	12.8	9.8	8.3	6.4	5.7	4.4	4.7	703
H.P. Lost	30.0	18.4	11.9	10.8	8.1	7.0	5.1	5.1	3.6	690
Mol. Wgt.	26.7	25.2	15.4	10.8	6.7	5.1	4.1	2.8	3.2	1800
Drainage	27.1	23.9	13.8	12.6	8.2	5.0	5.0	2.5	1.9	159
Atomic Wgt.	47.2	18.7	5.5	4.4	6.6	4.4	3.3	4.4	5.5	91
$n^{-1}, \sqrt{n}$	25.7	20.3	9.7	6.8	6.6	6.8	7.2	8.0	8.9	5000
Design	26.8	14.8	14.3	7.5	8.3	8.4	7.0	7.3	5.6	560
Digest	33.4	18.5	12.4	7.5	7.1	6.5	5.5	4.9	4.2	308
Cost Data	32.4	18.8	10.1	10.1	9.8	5.5	4.7	5.5	3.1	741
X-Ray Volts	27.9	17.5	14.4	9.0	8.1	7.4	5.1	5.8	4.8	707
Am. League	32.7	17.6	12.6	9.8	7.4	6.4	4.9	5.6	3.0	1458
Black Body	31.0	17.3	14.1	8.7	6.6	7.0	5.2	4.7	5.4	1165
Addresses	28.9	19.2	12.6	8.8	8.5	6.4	5.6	5.0	5.0	342
$n, n^2, \ldots, n!$	25.3	16.0	12.0	10.0	8.5	8.8	6.8	7.1	5.5	900
Death Rate	27.0	18.6	15.7	9.4	6.7	6.5	7.2	4.8	4.1	418
Average	30.6	18.5	12.4	9.4	8.0	6.4	5.1	4.9	4.7	1011
Benford's Law	30.1	17.6	12.5	9.7	7.9	6.7	5.8	5.1	4.6	

Слика 3. Распределбата на првата значајна цифра кај 20 различни колекции од нумерички податоци, тестирања кои ги направил Бенфорд (1938), [2].

Бенфорд забележал дека иако некои колекции од податоци не го задоволуваат законот, сепак кога сите тие различни колекции од податоци ќе се измешаат се добива нова низа кај која распределбата на првата значајна цифра има блиско однесување на Бенфордовата распре-

делба, односно велиме дека го поседува *Бенфордовото својство*. Ова посебно се забележува кај низите $n, n^2, \ldots$ која секоја одделно не го поседува Бенфордовото својство, но сите заедно се приближуваат кон тоа својство (претпоследната колекција од податоци на Слика 3). Исто така, распределбата на првата значајна цифра кај средната вредност на сите колекции е многу поблиска до Бенфордовата распределба, отколку истата распределба кај секоја колекција поединечно.

Меѓу другото, на Бенфорд му се припишува и неуспешниот обид да покаже дека веројатноста еден природен број да има прва значајна цифра 1 е еднаква на $\log_{10} 2$. Имено, тој сакал да ја пресмета веројатност при случаен избор на природен број да се избере број од множеството $\{1,10,11,12,\ldots,100,101,\ldots\}$, што е проблем за кој не постои границата на релативните честоти на појавување на тој настан, [7]. Неуспешноста на овој обид лежи и во неговата бесмисленост, имено обидот да се докаже вакво тврдење за природните броеви е како да сакаме да покажеме дека законот на Бенфорд важи за севкупноста од броеви, за која било природна случајна колекција од нумерички податоци, што не е точно. Сепак, Бенфорд многу подлабоко навлегол во анализирањето на оваа појава. Како што подоцна ќе биде покажано, неговите согледувања за среќавањето на овој феномен кај мешаните колекции од податоци се посуштински, а дал и геометриско објаснување на оваа појава, според кое процесите со константна стапка на раст го поседуваат Бенфордовото својство (види го делот 3.1. од овој труд за геометриското објаснување на Бенфорд).

2. ЕМПИРИСКИ РЕЗУЛТАТИ И СОГЛЕДУВАЊА

Пред да се зборува за тоа кои колекции од нумерички податоци го поседуваат Бенфордовото својство, треба да се споменат оние кои не го поседуваат и зошто е тоа така. Ако имаме податоци за дневни температури за некое место во текот на летото, за кое време температурите се движат од 12 до 31 степени, јасно е дека не се застапени сите цифри како прва значајна цифра, па не може ни да стане збор за проверка на законот на Бенфорд. Слично е и со висината кај луѓето, резултатите од тестовите и слични величини кои се нормално распределни, што значи дека нивните вредности се симетрично распределени околу очекуваната

вредност, па во случај на мала дисперзија, кај овие податоци не може да се очекува првата значајна цифра да ја прати Бенфордовата распределба.

Овие согледувања водат кон едно од објаснувањата на законот на Бенфорд – *хипотезата за распространетост,* која вели дека ако вредностите на податоците се распростанети низ неколку степени на интензитет, тогаш може да се очекува да првата значајна цифра има Бенфордова распределба. Меѓутоа, оваа хипотеза не треба да се зема како правило за поседување на Бенфордовото својство, туку само како показател зошто некоја колекција податоци не го поседува својството, во случај на недоволна распространетост на податоците. Имено, множеството податоци $\{1,10,100,1000,...,10^{2000}\}$ опфаќа 2000 степени на интензитет, а јасно е дека не го поседува Бенфордовото својство (секој број има прва значајна цифра 1).

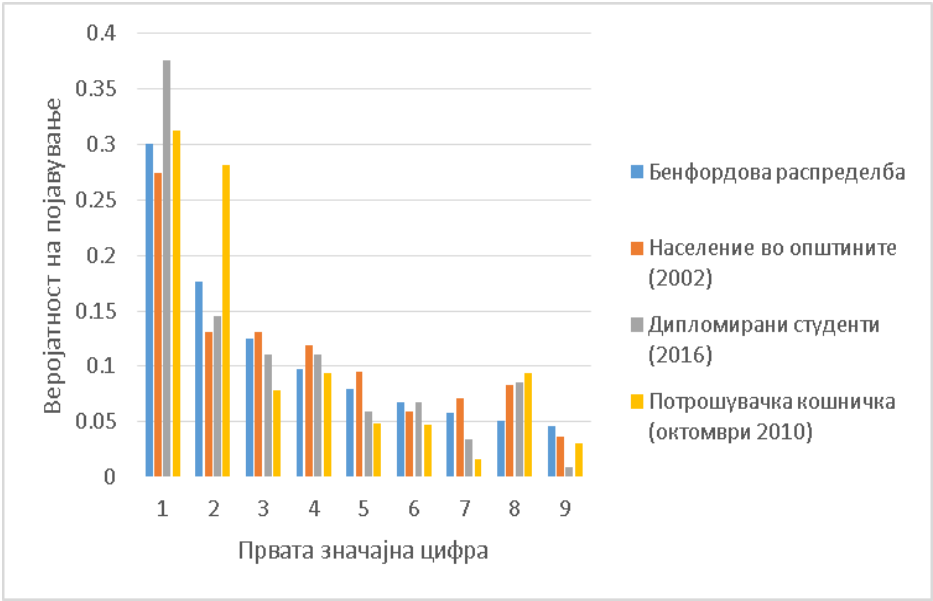
При проверката на законот на Бенфорд, пронајдени се различни колекции од податоци кои го поседуваат Бенфордовото својство. Некои од тие колекции ги има откриено и Бенфорд во неговото обемно истражување, како на пример: должините на реките, големина на областите, број на население, броеви од еден број на списание, воздушен притисок, радиоактивно зрачење, и други, [2]. Други примери за колекции кои за кои барем приближно важи законот на Бенфорд се: научни пресметки во случај на долги обемни примени на пресметковни операции; сметководствени и финансиски податоци во вид на платежни сметки, односно дневен приход од акции; некои физички мерења и други, [7].

2.1. НЕКОИ НОВИ ЕМПИРИСКИ РЕЗУЛТАТИ И СОГЛЕДУВАЊА

Законот на Бенфорд е толку фасцинантен и предизвикувачки, што ретко кој, додека го истражува овој феномен, може да оддолее, а да не го тестира на „свои“ податоци, и покрај мноштвото постоечки емпириски резултати. Тоа е и една од главните причини за тестирањата чии резултати ги презентираме во овој дел. Прво, беа тестирани три колекции од податоци – бројот на жители во општините во Република Македонија според пописот од 2002 година, бројот на дипломирани студенти на факултетите во Р. Македонија во 2016 година и просечната

потрошувачка кошничка на просечен граѓанин на Р. Македонија за месец октомври 2010 година. Податоците на кои е тестиран законот на Бенфорд се преземени од Statoids, [18] и Државниот завод за статистика на Р. Македонија, [17].

Првата колекција податоци, бројот на жители во општините во Р. Македонија според пописот од 2002 година, се состои од податоци за 84 општини со опсег од 1322 жители во Вранештица до 105485 жители во Куманово (жителите на градот Скопје се разгледуваат по општини). Втората колекција податоци, бројот на дипломирани студенти на факултетите во Р. Македонија во 2016 година, се состои од податоци за број на дипломирани студенти на 117 факултети со опсег од 1 дипломиран студент (на некои од факултетите) до 696 дипломирани студенти на Економскиот факултет при УКИМ во Скопје. И третата колекција податоци, просечната потрошувачка кошничка на просечен граѓанин на Р. Македонија за месец октомври 2010 година, се состои од потрошувачки месечни износи на 64 артикли со опсег од 14.83 ден. за квасец до 1172.60 ден. за бел леб. Распределбите на првата значајна цифра кај овие три колекции податоци, во споредба со Бенфордовата распределба, се дадени на Слика 4.



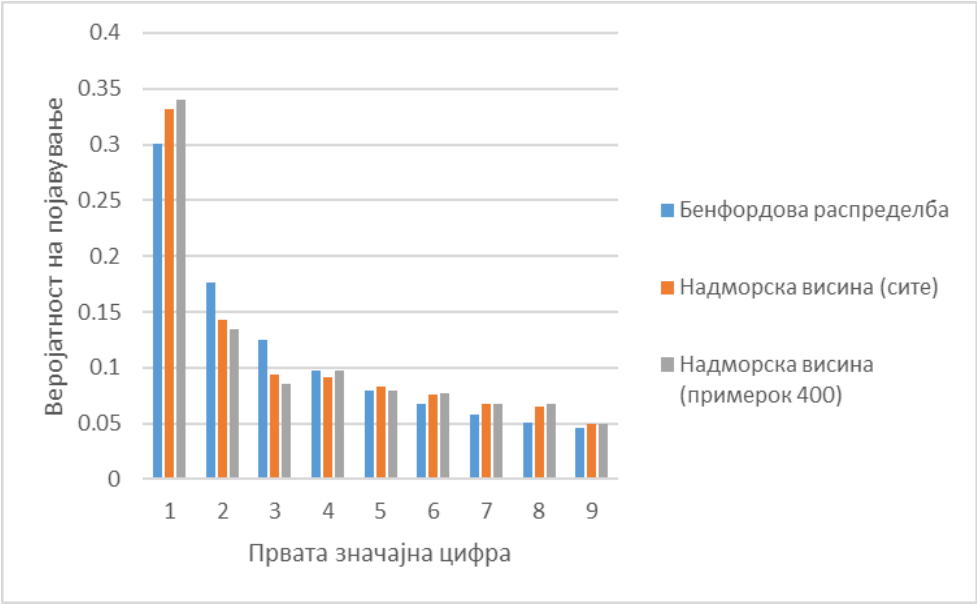
Слика 4. Распределбите на првата значајна цифра кај трите колекции податоци, во споредба со Бенфордовата распределба.

Според Слика 4, „визуелно“ најблиска распределба на првата значајна цифра до Бенфордовата распределба има бројот на жители т.е. население во општините во Република Македонија во 2002 година, што е и за очекување, заради самата природа на овие податоци. Имено, и во резултатите на Бенфорд на Слика 3, распределбата на првата значајна цифра кај бројот на население т.е. популацијата, незначително отскокнува од Бенфордовата распределба, најверојатно заради хипотезата на распространетост, бидејќи имаме вредности кои преминуваат преку неколку прагови на степените на бројот 10, но и заради тоа што бројот на население се должи на повеќе случајни фактори со различни распределби. Исто така, и при проверка на согласноста на распределбата на првата значајна цифра кај населението во општините со Бенфордовата распределба, за Пирсоновата χ^2 статистика се добива $\chi^2 = 4.118 < 15.507$, што значи дека статистички гледано не ја отфрламе претпоставката за Бенфордова распределба (статистичкиот Пирсонов χ^2 тест е со 8 степени на слобода).

Кај останатите две распределби на првите значајни цифри, на бројот на дипломирани студенти на факултетите во 2016 година и на месечните портошувачки износи на артиклите од потрошувачката кошничка за октомври 2010 година, се забележува визуелно можно отстапување од Бенфордовата распределба, едните имаат поголем број 1-ци како први значајни цифри, додека другите имаат поголем број на 2-ки за први значајни цифри. Од друга страна, за очекување е заради природата на овие податоци да важи Бенфордовото својство, имено бројот на дипломирани студенти е величина која зависи од повеќе случајни фактори со различни распределби, додека потрошувачките месечни износи се добиваат како резултат од повеќе пресметковни операции. Статистичкиот Пирсонов χ^2 тест за согласност го потврдува ова наслутување, имено Пирсоновата χ^2 статистика за бројот на дипломирани студенти е $\chi^2 = 11.175 < 15.507$, а за потрошувачките месечни износи изнесува $\chi^2 = 10.951 < 15.507$, што значи и во овие два случаја не ја отфрламе претпоставката за Бенфордова распределба на првата значајна цифра.

Овие три колекции податоци се незначително големи, 84, 117 и 64 податоци соодветно, што сосема одговара за примена на Пирсоновиот χ^2 тест за согласност кој се препорачува за помали колекции од податоци, [8]. Затоа, дополнителен предизвик претставува проверка на Бенфордовото својство кај голема колекција од нумерички податоци. За таа цел беа тестирани нумерички податоци за надморската височина на земјиштето во југоисточниот регион на Република Македонија. Податоците беа добиени со помош на дигиталниот висински модел ASTER GDEM V2 на NASA и METI, [16], при просторна резолуција од 38 m. Добиената матрица со нумерички вредности за надморските височини е со димензии 1254 на 1690, односно вкупно 2119260 податоци во опсег од 14.39 до 2171.95 метри. Примената на Пирсоновиот χ^2 тест за согласност на податоци од оваа димензија резултираше со неприфаќање на хипотезата за Бенфордова распределба на првата значајна цифра кај надморската височина на земјиштето, со многу голема вредност на Пирсоновата χ^2 статистика од $\chi^2 = 51889 > 15.507$. Од една страна, добиениот резултат не е за очекување заради природноста на податоците и нивната визуелна споредба со Бенфордовата распределба (види Слика 5), но од друга страна големината на колекцијата и несоодветната примена на Пирсоновиот χ^2 тест за согласност е причина за отфрлање на хипотезата. Имено, кај многу големи колекции од податоци, збирното отстапување од претпоставената распределба доведува до статистичко значајно несовпаѓање со претпоставената распределба, [8]. Во овој случај статистички оправдана постапка е тестирањето на законот на Бенфорд да се изврши на дел од податоците со разумна големина. Големината на делот од податоците беше одредена според симплифицираната формула за пропорции на Yamane, [6]. На случаен начин, од севкупноста на податоците, беше извлечен примерок со големина 400. Опсегот на оваа значајно помала колекција од податоци е од 18.07 до 2078.84 метри, а на овој начин беше надминато и „вештачкото“ поставување на просторната резолуција од 38 m, со тоа што 400-те вредности беа одбрани на случаен начин. На Слика 5 прикажана е распределбата на првата значајна цифра кај оваа потколекција од 400 податоци во споредба со Бенфордовата распределба. Овој пат Пирсоновиот χ^2 тест за согласност не ја отфрли претпос-

тавката за Бенфордова распределба на првата значајна цифра, со вредност на Пирсоновата χ^2 статистика од $\chi^2 = 14.501 < 15.507$. Овој резултат, односно идентификацијата на Бенфордовото својство кај податоците за надморската височина на земјиштето добиени од дигиталниот висински модел ASTER GDEM V2, би можел да послужи како потврда за веродостојноста на овој модел, односно неговата точност при проценка на надморската височина на земјиштето (види го делот 4 од овој труд за некои примени на законот на Бенфорд).



Слика 5. Распределбата на првата значајна цифра кај сите податоци за надморската височина на земјиштето во југоисточниот регион на Република Македонија, односно на примерок со големина 400 во споредба со Бенфордовата распределба.

3. МАТЕМАТИЧКИ ОБЈАСНУВАЊА

Откако видовме некои примери за податоци за кои важи и такви за кои не важи законот на Бенфорд, природно се наметнува потребата за карактеризирање на условите при кои важи законот. Во овој дел ќе изложиме две математички објаснувања на законот на Бенфорд. Првото објаснување е геометриското објаснување кое го дал Бенфорд за процесите со константна стапка на раст, [2], а второто е статистичката теорија на Теодор Хил, [5].

3.1. ГЕОМЕТРИСКОТО ОБЈАСНУВАЊЕ НА БЕНФОРД

Според ова објаснување, кај процесите со константна стапка на раст, првата значајна цифра има Бенфордова распределба. Примери за вакви процеси се радиоактивното зрачење, растот на популација од бактерии, Фибоначиевите броеви, степените на бројот 2, степените на кој било друг број и други. Но, да видиме како се доаѓа до тој заклучок.

Да претпоставиме дека имаме акции чија вредност се зголемува за 4% на годишно ниво, [11]. Ако n_d е бројот на години потребни за вредноста на акциите да порасне од d во $d+1$ денари, тогаш $d \cdot (1.04)^{n_d} = d+1$, од каде што

$$n_d = \frac{\log\left(\frac{d+1}{d}\right)}{\log 1.04}. \qquad (2)$$

Во Табела 1 е дадено времето за кое вредноста на акциите има прва значајна цифра еднаква на d, при фиксна стапка на раст од 4%. Од табелата може да се види дека се потребни повеќе од 17 години, акциите да пораснат од 1 до 2 денари, но потребни се помалку од 3 години тие да пораснат од 9 до 10 денари.

Првата значајна цифра	Години	Процент на време	Закон на Бенфорд
1	17.6730	0.30103	0.30103
2	10.3380	0.17609	0.17609
3	7.3350	0.12494	0.12494
4	5.6894	0.09691	0.09691
5	4.6486	0.07918	0.07918
6	3.9303	0.06695	0.06695
7	3.4046	0.05799	0.05799
8	3.0031	0.05115	0.05115
9	2.6863	0.04576	0.04576

Табела 1. Време за кое вредноста на акциите има прва значајна цифра еднаква на d, при фиксна стапка на раст од 4%, [11].

Ако се пресмета процентот на времето за кое вредноста на акциите има прва значајна цифра еднаква на d може да се воочи дека тоа се поклопува со Бенфордовиот закон (види Табела 1). Имено, ако n

е времето потребно вредноста на акциите да се зголеми од 1 на 10 денари, тогаш $1\cdot(1.04)^n = 10$ или $n = \frac{\log 10}{\log 1.04}$, па со користење на (2) добиваме дека процентот на времето за кое вредноста на акциите има прва значајна цифра еднаква на d е еднакво на

$$n_d \,/\, n = \frac{\log\left(\frac{d+1}{d}\right)}{\log 1.04} \Big/ \frac{\log 10}{\log 1.04} = \frac{\log\left(\frac{d+1}{d}\right)}{\log 10} = \log_{10}\left(\frac{d+1}{d}\right),$$

што навистина се поклопува со законот на Бенфорд даден со (1). Стапката на раст од 4% е случајно избрана за илустрирање, до ист заклучок би дошле и со која било друга вредност на фиксна стапка на раст.

Многу математички и природни појави имаат геометриски раст, како на пример радиоактивното зрачење, растот на популација од бактерии, па сè до Фибоначиевите броеви. Една од причините за геометрискиот раст е тоа што овие појави претставуваат решенија на диференцни равенки кои се линеарни комбинации на геометриски прогресии, [11]. На пример, низата Фибоначиеви броеви ја задоволува линеарната рекурентна равенка од втор ред

$$F_{n+2} = F_{n+1} + F_n,$$

за дадени почетни вредности на првите два Фибоначиеви броја $F_0 = F_1 = 1$. За n-тиот Фибоначиев број постои и експлицитна формула, позната како Бинеова формула т.е.

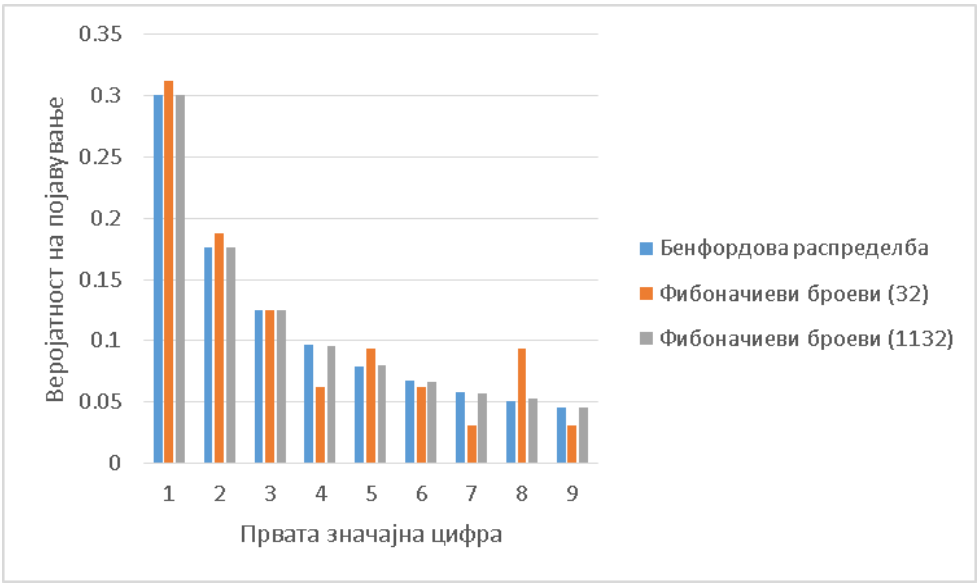
$$F_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n,$$

чие обопштување е дел од решение на систем од линеарни рекурентни равенки. Бидејќи, $\left|\frac{1+\sqrt{5}}{2}\right| > 1$ и $\left|\frac{1-\sqrt{5}}{2}\right| < 1$, па за големи вредности на n, се добива приближното претставување на n-тиот Фибоначиев број како

$$F_n \approx \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n,$$

односно $F_{n+1} \approx \frac{1+\sqrt{5}}{2} F_n$ или $F_{n+1} \approx 1.61803 F_n$. Ова значи дека низата Фибоначиеви броеви има константна стапка на раст од 61.803%, па

според претходната анализа, за очекување е да го поседува Бенфордовото својство, за големи вредности на n. Навистина, на Слика 6 се прикажани распределбите на првите значајни цифри кај првите 32, односно 1132 Фибоначиеви броеви во споредба со Бенфордовата распределба, од каде се согледува дека поголема вредност на n дава поблиска распределба до Бенфордовата.



Слика 6. Распределбата на првата значајна цифра кај првите 32, односно 1132 Фибоначиеви броеви во споредба со Бенфордовата распределба.

3.2. СТАТИСТИЧКАТА ТЕОРИЈА НА ТЕОДОР ХИЛ

Сите до сега спомнати обиди за објаснување на законот на Бенфорд имаат главно детерминистичка природа. За разлика од нив, теоријата на Теодор Хил од 1995 година, дава статистичко објаснување на законот на Бенфорд, [5]. Суштината на теоријата на Хил е најнапред во конструкција на соодветен простор на веројатност. Имено, вообичаениот простор на веројатност во кој настани се Бореловите множества во $\mathbb{R}$, не одговара за разгледување на проблемот на првата значајна цифра и нејзината распределба, бидејќи нема да постои единствен интервал кој ќе ги содржи сите реални броеви чии мантиси припаѓаат на одредено Борелово множество, па следствено нема начин како едноз-

начно да се доделат веројатностите на појавување на настаните. На пример, секој реален број од интервалите $[1,2)$, $[10,20)$, $[100,200)$, … има мантиса во интервалот $[1,2)$. Овде, терминот *мантиса* ќе го користиме за *значајниот дел* на реален број x, имено станува збор за бројот $y \in [1,2)$ од единствениот запис на $x = y10^n$, за некој $n \in \mathbb{Z}$. Затоа, Хил конструира нов простор на веројатност во кој настани се множествата од облик $\bigcup_{n=-\infty}^{\infty} B \cdot 10^n$, кои ги содржат сите позитивни реални броеви чии мантиси припаѓаат на Бореловото множество $B \subseteq [1,10)$. Тоа значи дека се разгледува мерливиот простор $(\mathbb{R}^+, M)$, каде M е σ-алгебра од настани дефинирана со:

$$M = \left\{ \bigcup_{n=-\infty}^{\infty} B \cdot 10^n \mid B \subseteq [1,10) \text{ е Борелово множество} \right\}. \qquad (3)$$

Основни карактеристики на σ-алгебрата M се тоа дека секое непразно множество од M е бесконечно со точки на натрупување 0 и +∞, што значи дека во секое множество од M може да се најдат произволно големи и прозволно мали ненулти броеви, потоа M е затворена во однос на множење со скалар, M е затворена во однос на коренување со коренов показател $m \in \mathbb{N}$ и M е себеслична т.е. ако $S \in M$ тогаш и $10^m S \in M$ за произволен $m \in \mathbb{N}$, [5]. Овие интересни особини на σ-алгебрата M дозволуваат воведување на претпоставките за скаларна инваријантност и базна инваријантност, кои ќе ги изложиме во продолжение.

Скаларната инваријантност е една од наједноставните претпоставки кога станува збор за „универзалноста“ на законот на Бенфорд. Да се потсетиме дека и Њукомб укажал на ова својство, препорачувајќи броевите „да се земаат како односи на количини“ за да не зависат од единицата мерка, [12]. Впрочем, на Универзумот не му е грижа која единица мерка ја користиме во нашите експерименти, па природно е еден универзален закон, како законот на Бенфорд да го поседува својството на скаларна инваријантност. На пример, ако законот на Бенфорд важи за вредностите на месечните сметки изразени во долари, треба да важи и за истите сметки конвертирани во денари.

Со следното тврдење се покажува дека скаларната инваријантност е карактеристика на законот на Бенфорд.

Теорема 1. [5] *Веројатносната мера P дефинирана над мерливиот простор* $(\mathbb{R}^+, M)$ *е скаларно инваријантна т.е.*

$$(\forall s \in \mathbb{R}^+)(\forall S \in M)\ \ P(S) = P(sS) \qquad (4)$$

ако и само ако P го задоволува законот на Бенфорд т.е.

$$P\left(\bigcup_{n=-\infty}^{\infty} [1,t) \cdot 10^n \right) = \log_{10} t \text{ , за сите } t \in [1,10). \qquad (5)$$

Да забележиме дека обликот (5) е општ облик на законот на Бенфорд, со кој е опфатена распределбата не само на првата значајна цифра, туку заедничката распределба на значајните цифри. Ако *k*-тата значајна цифра ја означиме со D_k, тогаш *општиот облик на законот на Бенфорд*, еквивалентен на обликот (5) е следниот:

$$P(D_1 = d_1, D_2 = d_2, ..., D_k = d_k) = \log_{10}\left[1 + \left(\sum_{i=1}^{k} d_i \cdot 10^{k-i} \right)^{-1} \right], \qquad (6)$$

за $d_1 \in \{1,2,3,...,9\}$, $d_j \in \{0,1,2,3,...,9\}$, $j = 2,...,k$.

Претпоставката за базна инваријантност е многу посуптилна од претпоставката за скаларна инваријантност, но исто така потекнува од универзалноста на законот. Имено, ако некој универзален закон важи за броевите во декаден броен систем, систем со основа 10, би требало да важи за броевите во систем со произволна основа *b*. За оваа намена ја означуваме со M_b, σ-алгебрата од мантиси во систем со основа *b*, па за $b = 10$ ја добиваме σ-алгебрата M дефинирана со (3). Тогаш, важат сите досегашни тврдења, при што во распределбата на веројатностите, $\log_{10}$ се заменува со $\log_b$. Следното тврдење ја дава врската меѓу базната инваријантност и законот на Бенфорд.

Теорема 2. ([5]) *Веројатносната мера P дефинирана над мерливиот простор* $(\mathbb{R}^+, M_b)$ *е базно инваријантна т.е.*

$$(\forall m \in \mathbb{N})(\forall S \in M_b)\ \ P(S) = P(S^{1/m}) \qquad (7)$$

ако и само ако $P = qP_b + (1-q)\delta_1$, *за некој* $q \in [0,1]$. (8)

Притоа, P_b е веројатносната распределба дефинирана со (5), каде што 10 се заменува со b, а δ_1 е Дираковата мера на множеството $S_1 = \bigcup_{n=-\infty}^{\infty} \{1\} \cdot b^n$ дефинирана со $\delta_1(S) = 1$, ако $S \supseteq S_1$ и $\delta_1(S) = 0$, во спротивен случај.

Од Теорема 1 и Теорема 2 може да се види дека скаларната инваријантност ја повлекува базната инваријантност, но обратното не важи, имено δ_1 е базно, но не и скаларно инваријантна мера.

Третиот придонес на статистичката теорија на Хил е статистичкото објаснување на експериментот на Бенфорд кој се состои од собрани случајни податоци од случајни распределби. Хил дефинира случајна веројатносна мера μ, мера на очекувана распределба $E\mu$ на случајната веројатносна мера μ и случајна низа од μ-случајни k-примероци, со чија помош ја моделира постапката на Бенфорд во неговиот експеримент и покажува зошто една таква постапка води кон распределба дефинирана со законот на Бенфорд.

Дефиниција 3. ([5]) *Случајна веројатносна мера* μ е случајна променлива, над разгледуваниот простор на веројатност (Ω, F, P), чии вредности се Борелови веројатносни мери на $\mathbb{R}$ и која е регуларна т.е. за секое Борелово множество $B \subset \mathbb{R}$, $\mu(B)$ е случајна променлива.

Дефиниција 4. ([5]) *Мерата на очекуваната распределба* на случајна веројатносна мера μ е веројатносната мера $E\mu$ (на Борелови подмножества од $\mathbb{R}$) дефинирана со:

$$(E\mu)(B) = \mathbf{E}(\mu(B)) \text{ за секое Борелово множество } B \subset \mathbb{R},$$

каде што $\mathbf{E}(\cdot)$ е математичкото очекување во однос на соодветната веројатносна мера $\mathbf{P}$.

Дефиниција 5. ([5]) За случајна веројатносна мера μ и природен број k, *случајна низа од* μ*-случајни* k*-примероци* е низа од случајни променливи $X_1, X_2, \ldots$ над (Ω, F, P) така што постои низа од независни и еднакво распределени случајни веројатносни мери $\mu_1, \mu_2, \ldots$ со иста распределба како μ и за секој $j = 1, 2, \ldots$

(i) случајните променливи $X_{(j-1)k+1},...,X_{jk}$ се независни и еднакво распределни со распределба $P=\mu_j$, и

(ii) $X_{(j-1)k+1},...,X_{jk}$ се независни од $\mu_i, X_{(i-1)k+1},...,X_{ik}$ за $i\neq j$.

Колку и да изгледа апстрактна последната Дефиниција 5 за случајна низа од μ-случајни k-примероци, впрочем станува збор за колекција од (независни) примероци со големина k, при што секој примерок одговара на распределба, случајно и независно избрана со помош на случајната веројатносна мера μ. Може да се покаже дека, дури и кога $X_1, X_2,...$ не се независни и еднакво распределни случајни променливи, разгледуваната честота на настан во низата од μ-случајни k-примероци сепак конвергира кон веројатност која е мера на настан според очекуваната распределба на μ, [5]. Односно, за да ја знаеме честотата на настанот, треба да ја познаваме само очекуваната распределба.

Откако ги поставивме основите на статистичката теорија на Хил, може да го изложиме главниот резултат во скратен облик, [7]. Проширената верзија на овој резултат може да се најде во [5].

Теорема 3. ([5]) (**Логаритамско граничен закон за значајните цифри**) *Нека μ е случајна веројатносна мера на $(\mathbb{R}^+, M)$. Следните тврдења се еквивалентни:*

(i) $E\mu$ е скаларно инваријантна,

(ii) $E\mu$ е базно инваријантна и безатомска на $(\mathbb{R}^+, M)$,

(iii) $E\mu\left(\bigcup_{n=-\infty}^{\infty}[1,t)\cdot 10^n\right)=\log_{10} t$ *за сите* $t\in[1,10)$,

(iv) за секоја низа од μ-случајни k-примероци $X_1, X_2,...$,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}1_{\mu(X_i)\in[1,t)}=\log_{10} t \text{ за сите } t\in[1,10). \qquad (9)$$

Притоа, *безатомска веројатносна мера* е онаа која не содржи *атоми* т.е. мерливи множества со позитивна мера кои не содржат помали множества со позитивна мера. Да забележиме дека во (iv) од Теорема 3, ознаката $1_{[\cdot]}$ е индикатор функција, па во нашиот случај $1_{\mu(X_i)\in[1,t)}$ прима вредност 1, ако $\mu(X_i)\in[1,t)$, а во спротивно, прима вредност 0.

Последната Теорема 3 дава објаснување зошто во експериментот на Бенфорд, иако некои од колекциите податоци ни приближно не го задоволувале законот на Бенфорд (на пример, одредени низи од броеви), сепак средната вредност од честотите на појавување на првите значајни цифри во севкупноста од сите колекции, сепак го задоволува законот на Бенфорд. Имено, во низата од μ-случајни k-примероци, само очекуваната распределба влијае на пресметките со честотите, додека сите останати поединечни распределби се на некој начин неважни. Или како што констатира Хил, „постојат многу (природни) начини за избор на примерок кои водат кон логаритамска распределба“, [5].

На крајот, да забележиме само дека претпоставките за скаларна и базна инваријантност се скоро исто оправдани како и претпоставката за независни (и еднакво распределени) случајни променливи кај строгиот закон на големите броеви или централната гранична теорема - ниедни од овие претпоставки со сигурност не можат да се докажат дека важат, а сепак во многу реални примени разумно е да се претпостават, [5].

4. ПРИМЕНИ

Статистичката теорија на Теодор Хил и главниот резултат даден со Теорема 3, оправдува многу од примените на законот на Бенфорд. Сепак, постојат три поголеми области на примена: дизајнирање на ефикасни сметачи, математичко моделирање и откривање на измами, [5], [7].

При дизајнирањето на ефикасни сметачи, важно е одредувањето на меморијата за складирање и брзината на процесирање при правењето на пресметките. Така, ако распределбата на влезните податоци е позната, тогаш таа информација може да се искористи за дизајнирање на оптимален сметач во однос на таа распределба. Може да се покаже дека под претпоставка на логаритамски влез (влезни податоци чии значајни цифри пратат логаритамска распределба), системот со основа $b = 2^3$ е оптимален во однос на минимален простор за складирање, [14]. Од друга страна, податоците добиени од научни пресметки се однесуваат како случајно избрани примероци од случајни распределби, па според Теорема 3, овие податоци би одговарале на споменатиот логаритамски влез.

Втората поголема примена на законот на Бенфорд се основа на идејата дека ако одредено множество податоци го следи законот на Бенфорд, тогаш и математичките модели за предвидување основани на овие податоци исто така го следат законот на Бенфорд, [15]. На овој начин, познавајќи дека за одредени податоци важи законот на Бенфорд, може да се тестираат математичките модели. Имено, ако за предвидувањата не важи законот на Бенфорд, разгледуваниот модел треба да се отфрли и да се бара нов.

Согледувањето дека манипулираните податоци не го следат законот на Бенфорд, туку само природно добиените, навестува дека законот на Бенфорд може да се користи за откривање на измами базирани на манупилирање со нумерички податоци. Па така, законот на Бенфорд може да се искористи и за откривање на дуплирани податоци во базите, како на пример повторни наплатувања. Можеби и најголемата слава законот на Бенфорд ја доживува со изработката на тестот за согласност со Бенфордовата распределба од страна на Нигрини, со чија помош се откриени финансиски измами во седум компании во Њујорк во 1995 година. Имено, ова расудување се заснова на откритието на Нигрини дека даноците, сметководствените податоци и трансакциите на берзата приближно го следат Бенфордовиот закон, [13].

Во насока на откривањето на измами, забележана е примената на законот на Бенфорд и при откривањето на измама за време на претседателските избори во Иран во 2009 година, кога при пребројувањето на бројот на гласови, за еден од кандидатите биле пријавени број на гласови од избирачките места добиени со замена на првата цифра со поголема. Имено, статистичките анализи на бројот на гласови, основани на примена на Бенфордовите тестови за првата и втората значајна цифра, како и големиот број на аутлеери, покажале дека „средно силно“ може да се верува дека за време на изборите била направена измама [9]. Потоа, законот на Бенфорд во комбинација со моделот на Бениш (Messod D. Beneish) и Алтмановиот Z-тест помогнал и при откривање на финансиската измама во фабриката Тошиба во периодот 2008 – 2014 година, [10]. Законот на Бенфорд е применет и при анализа на фотографии и проверка на нивната природност, при што добиено е дека обработуваните фотографии не го следат Бенфордовиот закон, [1]. Исто така, регресионите коефициенти во научните трудови покажуваат

согласност со законот на Бенфорд, па во овој случај законот на Бенфорд може да биде искористен и за откривање на научни измами, поточно лажни податоци во научните трудови, [4].

5. ЗАКЛУЧОК

Можеби историјата за законот на Бенфорд, од неговото откривање, преку емпириските потврдувања, математичката теорија и примените, изгледа како една завршена приказна, сепак вниманието кое го добива во последните години, зборува нешто сосема поинаку. Имено, во секое од спомнатите полиња има простор за нови истражувања, нови идеи, нови примени, па и нови подобри математички теории.

Благодарност. Авторот му благодари на д-р Свемир Горин, вонреден професор на Институтот за географија при Природно-математичкиот факултет во Скопје, за стручната помош околу добивањето и обработката на податоците за надморската височина на земјиштето во југоисточниот регион во Република Македонија.

ЛИТЕРАТУРА

[1] E. Acebo, M. Sbert, *Benford's Law for Natural and Synthetic Images*, in Computational Aesthetics in Graphics, Visualization and Imaging, L. Neumann, M. Sbert, B. Gooch and W. Purgathofer Eds., 2005, 169-176.

[2] F. Benford, *The Law of Anomalous Numbers*, Proceedings of the American Philosophical Society, Vol. 78, No. 4 (1938), 551–572.

[3] A. Berger, T. P. Hill, *A basic theory of Benford's Law*, Probability Surveys, Vol. 8 (2011) 1–126.

[4] A. Diekmann, *Not the First Digit! Using Benford's Law to detect fraudulent scientific data*, J Appl Stat, 34 (3) (2007), 321–329.

[5] T. P. Hill, *A Statistical Derivation of the Significant-Digit Law*, Statist. Sci., Volume 10, Number 4 (1995), 354–363.

[6] G. D. Israel, *Determining Sample Size*, PEOD6, University of Florida, November 1992.

[7] A. Jamain, *Benford's Low*, Imperial College of London, Department of Mathematics, April-September 2001.

[8] R. Johnson, D. Wichern, *Applied Multivariate Statistical Analysis*, Prentice Hall, Englewood Cliffs, New Jersey, 1992.

[9] W. R. Mebane, Jr., *Note on the presidential election in Iran, June 2009*, (2009), http://www-personal.umich.edu/~wmebane/note18jun2009.pdf

[10] A. Mehta, G. Bhavani, *Application of Forensic Tools to Detect Fraud: The Case of Toshiba*, Journal of Forensic and Investigative Accounting, 9(1) (2017), 692–710.

[11] S. J. Miller, ed., *Benford's Law: Theory and Applications*, Princeton University Press, 2015.

[12] S. Newcomb, *Note on the frequency of use of the different digits in natural numbers*, Amer. J. Math. 4 (1881), 39–40.

[13] M. Nigrini, *A taxpayer compliance application of Benford's Law*, The Journal of the American Taxation Association, Vol. 18, Iss. 1 (1996), 72–91.

[14] P. Schatte, *On mantissa distributions in computing and Benford's Law*, Journal of Information Processing and Cybernetics, Vol.24 (1988), 443–455.

[15] H. Varian, *Benford's Law*, American Statistician 23 (1972), 65–66.

[16] ASTER Global Digital Elevation Map, https://asterweb.jpl.nasa.gov/gdem.asp

[17] Државен завод за статистика на Република Македонија, http://www.stat.gov.mk/

[18] Macedonia Municipalities, Statoids, http://www.statoids.com/umk.html

[1] Универзитет „Св. Кирил и Методиј“, Скопје,
Природно-математички факултет,
Архимедова 3, 1000 Скопје, Р. Македонија
e-mail: irenatra@pmf.ukim.mk

Примен: 25. 04. 2018
Поправен: 07. 06. 2018
Одобрен: 23. 06. 2018

Објавен на интернет: 28.08.2018