

26. - 28. oktober 2020 | Kongresni center Bernardin, Portorož

ODKRIVANJE PREVAR Z ORODJI POSLOVNE INTELIGENCE

mag. Tomaž Dular, Špela Povrženič, dr. Julija Lapuh Bele, Rok Pirnat
Skupina B2

26.10.2020

Odkrivanje prevar z orodji poslovne intelligence

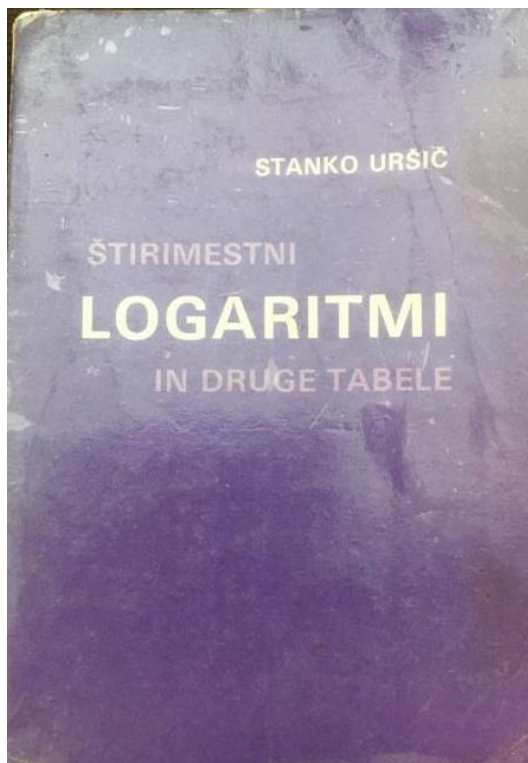
- Benfordov zakon
 - Definicija
 - Načelo invariantnosti
 - Testiranje podatkovne zbirke
- Testni primer
- Demo

Benfordov zakon

- Empirični zakon
- Nizke številke (1, 2,..) se pojavijo pogostejše kot visoke številke (7, 8 in 9)
- Primeri: čas.intervali med potresi, molske mase kemičnih spojin, količina vode v rečnih tokovih, populacija v občini ali državi, rezultati volitev po mestih, dohodki gospodinjstev, velikost datotek v megabajtih na poljubnem računalniku,...

Zgodovina

- Simon Newcomb (1881) in Frank Benford (1938)



	Sinus	Tangens	Secans
14			
31	2506616	2589280	10329781
32	2509432	2592384	10330559
33	2512248	2595488	10331339
34	2515063	2598593	10332119
35	2517879	2601699	10332901
36	2520694	2604805	10333683
37	2523508	2607911	10334467
38	2526323	2611018	10335251
39	2529137	2614126	10336037
40	2531952	2617234	10336823
41	2534766	2620342	10337611
42	2537579	2623451	10338399
43	2540393	2626560	10339188
44	2543206	2629670	10339979
45	2546019	2632780	10340770
46	2548832	2635891	10341563
47	2551645	2639002	10342356
48	2554458	2642114	10343151
49	2557270	2645226	10343945
50	2560082	2648339	10344743
51	2562894	2651452	10345540
52	2565705	2654566	10346338
53	2568517	2657680	10347138
54	2571328	2660794	10347938
55	2574139	2663909	10348740
56	2576950	2667025	10349542
57	2579760	2670141	10350346
58	2582570	2673257	10351150
59	2585381	2676374	10351955
60	2588190	2679492	10352762

Matematični zapis

$$P(1. \text{ številka} = d) = \log_B(1 + 1/d)$$

Vodilna številka	Vodilna številka - math	Porazdelitev	Pogoji
Zakon prve vodilne številke	1. številka = d	$\log(1 + 1/d)$,	$d \in \mathbb{Z}, d \in [1, 9]$.
Zakon druge vodilne številke	2. številka = d	$\sum_{k=1:9} \log(1 + 1/(10k + d))$	$d \in [0, 9]$.
Zakon tretje vodilne številke	3. številka = d	$\sum_{m=1:9} \sum_{n=0:9} \log(1 + 1/(100m + 10n + d))$	$d \in [0, 9]$.
Zakon n-te vodilne številke	n. številka = d	$10 \sum_{n-1-k=10^{n-2}} \log(1 + 1/(10k + d))$	$n \in \mathbb{N}, d \in [0, 9]$.
Kombinacija prvih dveh števk	1. številka = p, 2. številka = q	$\log(1 + 1/(10p + q))$	$p \geq 0$ in $q \geq 0$.
Kombinacija prvih treh števk	1. številka = p, 2. številka = q, 3. številka = r	$\log(1 + 1/(100p + 10q + r))$	$p \geq 1, q \geq 0, r \geq 0$.
Kombinacija zadnjih dveh števk		1/100 oz. 1 % - enakomerna porazdelitev	

Prva vodilna številka je prva neničelna številka na levi strani števila. Predznaka ne upoštevamo.

...

- Pogoji veljavnosti:

- Red velikosti = $\log(X_{\text{Max}}) - \log(X_{\text{Min}}) > 3$

- Načelo invariantnosti:

- ne glede na merski sistem enot bo slučajna spremenljivka enako porazdeljena

- če je množica dovolj velika in ima dovolj velik red velikosti

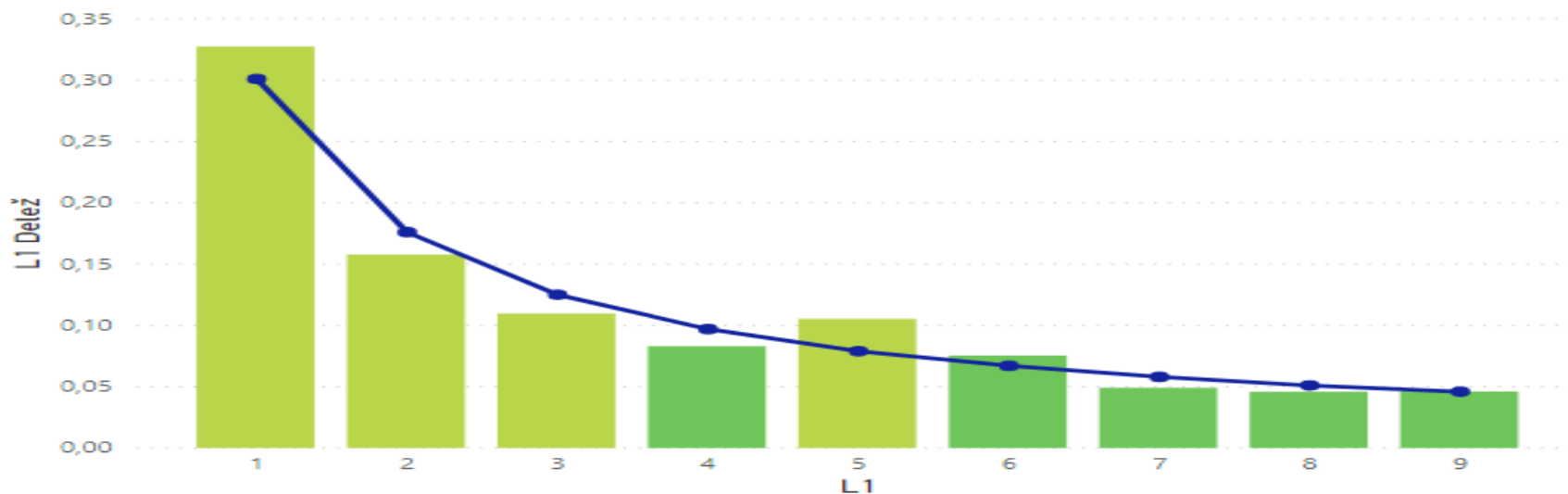
Testiranje podatkov

- Test Z
- Test hi-kvadrat
- Odstopanje vsote kvadratov SSD

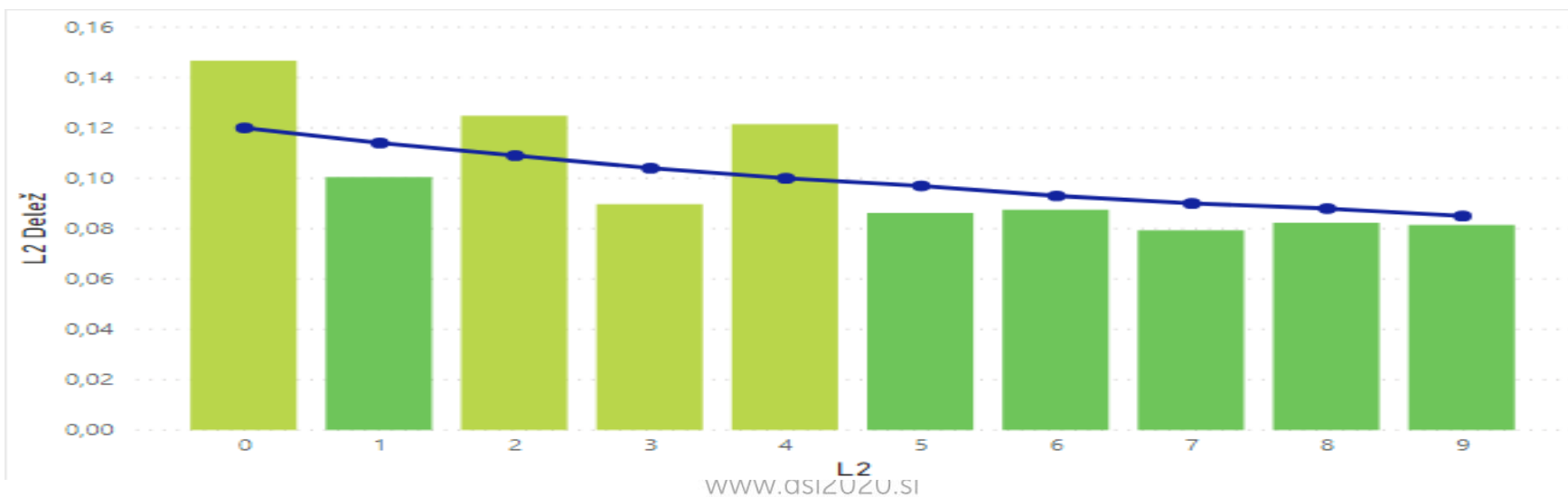
$$SSD = \sum_{d=1}^{RD} (O_d^{\%} - E_d^{\%})^2,$$

SSD (%)	Popolni Benford	Sprejemljiv Benford	Mejni Benford	Ni Benford
Zakon prve vodilne številke	<2	2-25	25-100	>100
Zakon druge vodilne številke	<2	2-10	10-50	>50
Komb. prvih dveh števk	<2	2-10	10-50	>50
Komb. zadnjih dveh števk	<4	4-40	40-100	>100

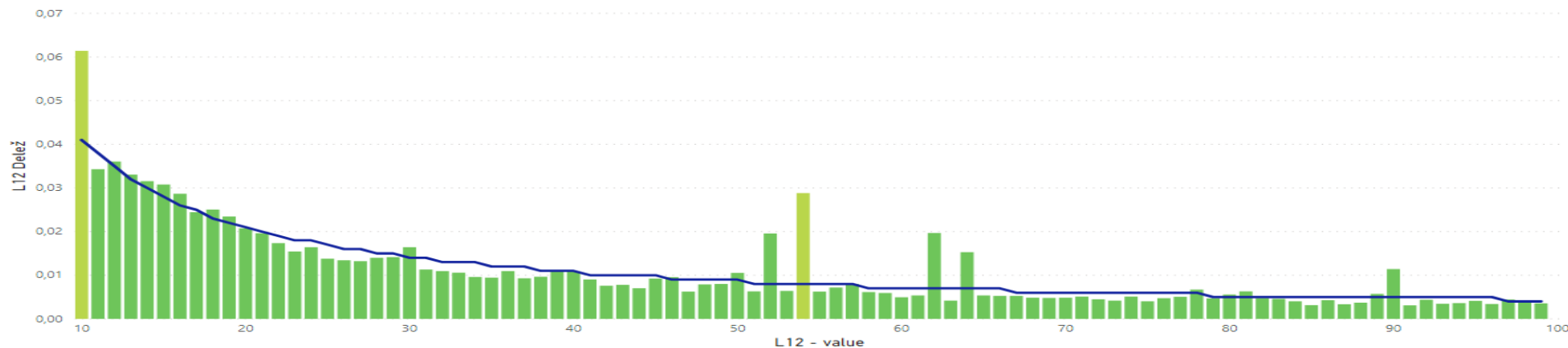
Slika 1: Delež prve vodilne številke pri neto zneskih računov



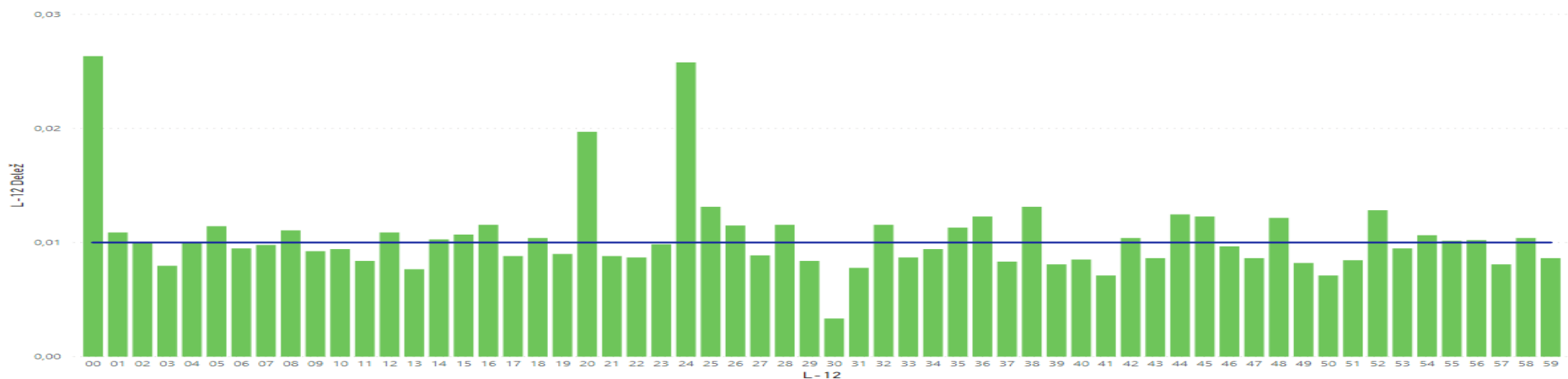
Slika 2: Delež druge vodilne številke pri neto zneskih računov



Slika 3: Delež kombinacije prve in druge vodilne številke pri neto zneskih računov



Slika 4: Delež kombinacije zadnjih dveh števk pri neto zneskih računov - izsek



Hvala za pozornost!