

Varijabilitet kvalitativnih podataka – Zadaci

Deskriptivna statistika

Aleksandar Tomašević

April 2020.

1 Urađen primer

Zadatak 1. Stanovnici dve opštine anketirani su povodom predstojećih parlamentarnih izbora. Od njih je traženo da od 4 ponuđene političke partije izaberu onu za koju će najverovatnije glasati.

Tabela 1: Frekvencija odgovora ispitanika u opštinama A i B

Partija	Opština A	Opština B
P1	20	10
P2	18	30
P3	25	5
P4	17	5
Σ	80	50

- A) Izračunati Šenonov indeks entropije za opštinu A, a potom i njegovu maksimalnu i standardizovanu vrednosti
- B) Uraditi isto za opštinu B.
- C) Uporediti varijabilitet u ove dve opštine.

1.1 Formule

Osnovni princip merenja varijabiliteta kvalitativnih varijabli glasi:

1. Varijabilitet ne postoji onda kada su svi podaci pripadaju jednoj kategoriji, odnosno modalitetu.
2. Varijabilitet je najveći onda kada su podaci ravnomerno raspoređeni po kategorijama.

Formula 1. Formula za izračunavanje Šenonovog indeksa entropija ima sledeći oblik.

$$H = - \sum_{i=1}^k p \log p \quad (1)$$

Pri čemu su:

p — relativne frekvencije, odnosno proporcije svakog modaliteta $p = \frac{f}{\sum f}$

$\log p$ — logaritam vrednosti relativne frekvencije

k — broj modaliteta, odnosno kategorija koje poseduje kvalitativna varijabla

Varijabilitet je najveći kada su frekvencije **svih modaliteta jednake**. U tom slučaju, vrednost Šenonovog indeksa se približava maksimalnoj mogućoj vrednosti u odnosu na broj kategorija/modaliteta. U suprotnom, kada imamo nizak varijabilitet, imamo i nisku vrednost Šenonovog indeksa (teži ka 0).

Formula 2. *Maksimalna vrednost Šenonovog indeksa*

$$H_{max} = \log k \quad (2)$$

Da bismo lakše interpretirali vrednost Šenonovog indeksa, možemo izračunati i njegovu standardizovanu vrednost.

Formula 3. *Standardizovana vrednost Šenonovog indeksa*

$$H_S = \frac{H}{H_{max}} \quad (3)$$

Što je vrednost standardizovanog Šenonovog indeksa bliža 0, to je varijabilitet manji. Obrnuto, što je bliža 1 to je varijabilitet veći. Standardizovana vrednost će biti jednaka 1 samo onda kada su frekvencije svih modaliteta jednake.

1.2 A

Da bismo izračunali vrednosti Šenonovog indeksa za opštinu A prema formuli (1) potrebno je da izračunamo tri dodatne kolone. Prva od njih je kolona sa relativnim frekvencijama za Opštinu A. To ćemo uraditi tako što ćemo frekvenciju svakog modaliteta (P1, P2, P3 i P4) podeliti sa sumom frekvencija, odnosno 80.

Napomena. *Prilikom postupka izračunavanja Šenonovog indeksa, njegove maksimalne i standardizovane vrednosti, beležite rezultate na 3 decimale.*

Tabela 2: Dodata kolona p_A

Partija	Opština A	Opština B	p_A
P1	20	10	0.250

Partija	Opština A	Opština B	p_A
P2	18	30	0.225
P3	25	5	0.313
P4	17	5	0.213
Σ	80	50	1.000

Sledeći korak je izračunavanje logaritma relativne frekvencije.

Tabela 3: Dodata kolona $\log p_A$

Partija	Opština A	Opština B	p_A	$\log p_A$
P1	20	10	0.250	-0.602
P2	18	30	0.225	-0.648
P3	25	5	0.313	-0.505
P4	17	5	0.213	-0.673
Σ	80	50	1.000	

Na kraju, potrebno je da dodamo novu kolonu koja predstavlja proizvod prethodne dve. Negativna vrednost sume ove kolone zapravo predstavlja vrednost Šenonovog indeksa prema formuli (1).

Tabela 4: Dodata kolona $p_A \log p_A$

Partija	Opština A	Opština B	p_A	$\log p_A$	$p_A \log p_A$
P1	20	10	0.250	-0.602	-0.151
P2	18	30	0.225	-0.648	-0.146
P3	25	5	0.313	-0.505	-0.158
P4	17	5	0.213	-0.673	-0.143
Σ	80	50	1.000		-0.597

Na osnovu sume poslednje kolone možemo zaključiti da je vrednost Šenonovog indeksa za ovu opštinu $H_A = 0.597$

Maksimalnu vrednost Šenonovog indeksa izračunaćemo preko formule (2). Broj modaliteta je 4.

$$H_{A_{max}} = \log k = \log 4 = 0.602$$

Na kraju, standardizovanu vrednost Šenonovog indeksa izračunaćemo putem formule (3).

$$H_{A_S} = \frac{0.597}{0.602} = 0.991$$

1.3 B

Sličan postupak sa izračunavanjem relativnih frekvencija i dodatne dve kolone ponovićemo za opštinu B.

Tabela 5: Dodate izračunate vrednosti za opštinu B

Partija	Opština A	Opština B	p_A	$\log p_A$	$p_A \log p_A$	p_B	$\log p_B$	$p_B \log p_B$
P1	20	10	0.250	-0.602	-0.151	0.200	-0.699	-0.140
P2	18	30	0.225	-0.648	-0.146	0.600	-0.222	-0.133
P3	25	5	0.313	-0.505	-0.158	0.100	-1.000	-0.100
P4	17	5	0.213	-0.673	-0.143	0.100	-1.000	-0.100
Σ	80	50	1.000		-0.597	1.000		-0.473

Na osnovu sume poslednje kolone možemo zaključiti da je vrednost Šenonovog indeksa za ovu opštinu $H_B = 0.473$

Pošto je broj kategorija, odnosno modaliteta isti kao u i za opštinu A, maksimalna vrednost je identična: $H_{B_{max}} = 0.602$

Na kraju, standardizovanu vrednost Šenonovog indeksa izračunaćemo putem formule (3).

$$H_{B_s} = \frac{0.473}{0.602} = 0.786$$

1.4 C

Standardizovana vrednost Šenonovog indeksa je manja za opštinu B u odnosu na opštinu A, što nam govori da je varijabilitet podataka veći u slučaju opštine B.

Drugim rečima, kada je u pitanju podrška građana političkim partijama, ona više varira u opštini A nego u opštini B jer su frekvencije ujednačenije u prvoj opštini.

To je uočljivo u i u samim frekvencijama. U opštini B, partija 2 ima izraženo veću podršku u odnosu na ostale, dok su u slučaju opštine A partije relativno ujednačene po snazi podrške.

2 Primer za vežbanje

Zadatak 2. Stanovnici jednog grada anketirani su povodom predstojećih izbora za gradonačelnika. Od njih je traženo da od 5 kandidata izaberu onog za kojeg će najverovatnije glasati. Istrživanje je urađeno u januaru mesecu (na samom početku kampanje) i u aprilu (nekoliko dana pred izbore).

Tabela 6: Frekvencija odgovora ispitanika u mesecu januaru i aprilu

Kandidat	Januar	April
K1	61	50

Kandidat	Januar	April
K2	12	45
K3	33	38
K4	28	38
K5	2	30
Σ	136	201

- A) Izračunati Šenonov indeks entropije za januarske podatke, a potom i njegovu maksimalnu i standardizovanu vrednosti
- B) Uraditi isto za aprilske podatke.
- C) Uporediti varijabilitet ove dve grupe podataka.
- D) U kom trenutku (januar ili april) je delovalo da će izbori biti neizvesniji?