

## KOEFICIJENT KONTINGENCIJE

Mjerenje ovisnosti dviju pojava predloženih oblicima nominalnih obilježja nije moguće s pomoću opisanih koeficijenata korelacije. Na podatke koji predlažu oblike nominalnih obilježja ne mogu se primijeniti računski radnje zbrajanja (oduzimanja) i množenja (dijeljenja), a nema ni uređenja određenoga njihovim svojstvima. Stupanj povezanosti nominalnih varijabli izražava se *mjerama asocijacije*, kojih je više. Među mjere asocijacije ubraja se i *koeficijent kontingencije*. Temelj te mjere objasniti ćemo primjerom.

### PRIMJER 1.

Odjeljenje za istraživanje tržišta jednog poduzeća anketiranjem je prikupilo podatke o potrošnji bezalkoholnih napitaka „Višnja“, „Jabuka“, „Cola“ i „Tonic“. Prikupljeni podatci grupirani su prema vrsti napitaka i potrošačkoj skupini anketiranih. Svrha je ustanoviti postoji li povezanost izbora napitka i pripadnosti potrošačkoj skupini. Grupirani podatci prikazani su u tablici 1.

Tablica 1. Anketirani prema izboru napitka i potrošačkoj skupini

| Potrošačka skupina | Vrsta bezalkoholnog napitka |                 |                 |                 | Ukupno       |
|--------------------|-----------------------------|-----------------|-----------------|-----------------|--------------|
|                    | „Višnja“                    | „Jabuka“        | „Cola“          | „Tonic“         |              |
| kućanice           | 211<br>(165.30)             | 155<br>(169.50) | 167<br>(195.90) | 142<br>(144.30) | 675          |
| poslovni ljudi     | 83<br>(122.44)              | 98<br>(125.56)  | 195<br>(145.11) | 124<br>(106.89) | 500          |
| radnici            | 199<br>(178.77)             | 245<br>(183.31) | 157<br>(211.86) | 129<br>(156.06) | 730          |
| studenti           | 58<br>(84.49)               | 67<br>(86.63)   | 134<br>(100.13) | 86<br>(73.75)   | 350          |
| <b>Ukupno</b>      | <b>551</b>                  | <b>565</b>      | <b>653</b>      | <b>481</b>      | <b>2 250</b> |

Podatci u tablici tumače se na već objašnjeni način. Na primjer, frekvencija 221 govori da je od 2 250 anketiranih 211 u skupini kućanica koje su izabrale napitak „Višnja“. Frekvencija 86 pokazuje da je toliko broj anketiranih u skupini studenata koji su izabrali napitak „Tonic“.

Za ispitivanje povezanosti obilježja važno je poznavati značenje sadržaja rubnog stupca odnosno rubnog retka. U rubnom su stupcu podatci koji se odnose na obilježje „potrošačka skupina“. Tu je grupiranje neovisno o obilježju „vrsta bezalkoholnog napitka“. Frekvencija 675 predlaže broj anketiranih kućanica, neovisno o izboru pića. Od ukupnog broja anketiranih „kućanica“ je 30.00 %, „poslovnih ljudi“ 22.22 %, „radnika“ 32.45 %, a „studenata“ 15.33 %. Te su postotne relativne frekvencije dobivene diobom frekvencija rubnog stupca s njihovim zbrojem, tj. s 2 250, i množenjem omjera sa sto. Analogno se tumače frekvencije rubnog retka i računaju relativne frekvencije.

Dodatni sadržaj veže se uz sadržaj u udžbeniku  
Šošić, I., *Statistika*, Školska knjiga, Zagreb, 2014.  
na stranici 235.

Da bi se brojčano izrazila povezanost pripadnosti skupini potrošača i izbora pića, poći ćemo od pretpostavke da je izbor pića *neovisan* o pripadnosti potrošačkoj skupini. Ako se pođe od te pretpostavke, tada se očekuje da od 551 anketirane osobe koje su izabrale napitak „Višnja” 30 % pripada potrošačkoj skupini „kućanice”. Naime, 30 % od 511 iznosi 165.30 ili zaokruženo 165. Od 551 osobe s izborom napitka „Višnja” 22.22 % trebalo bi pripadati skupini „poslovni ljudi”. 22.22 % od 551 jest 122.44 ili zaokruženo 122. Jednako su izračunani postotni dijelovi i za druge potrošačke skupine. Postotni dijelovi za druge vrste napitaka izračunavaju se analogno. Oni su u tablici 1. napisani u zagradi i navedeni ispod polaznih frekvencija.

Izračunani postotni dijelovi nazivaju se *očekivanim frekvencijama*. Broj anketiranih kućanica s izborom napitka „Višnja” iznosi 211. Kada bi izbor napitka bio *neovisan* o pripadnosti potrošačkoj skupini, broj kućanica s izborom napitka „Višnja” bio bi 165. Broj studenata s izborom napitka „Tonic” je 86, a očekivani 73.75, odnosno 74, *uz uvjet neovisnosti* pripadnosti potrošačkoj skupini i izbora napitka. Očekivane frekvencije mogu se jednostavnije izravno izračunavati primjenom formule:

$$e_{ij} = \frac{f_i \cdot f_j}{N}, i = 1, 2, \dots, 3; j = 1, 2, \dots, c..$$

U formuli su  $f_i$  frekvencije u rubnom stupcu, a  $f_j$  frekvencije u rubnom retku. Može se pokazati da je zbroj očekivanih frekvencija jednak zbroju frekvencija  $N$ .

Razlike između stvarnih i očekivanih frekvencija upućuju na odstupanje od neovisnosti, tj. postojanje ovisnosti (asocijacije) dvaju obilježja.

**KOEFICIJENT KONTINGENCIJE** kojim se mjeri stupanj asocijacije nominalnih varijabli dan je izrazom:

$$(1) C = \sqrt{\frac{\chi^2}{N + \chi^2}}, \quad \chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(f_{ij} - e_{ij})^2}{e_{ij}}.$$

U navedenom<sup>1</sup> izrazu  $N$  je zbroj frekvencija (u primjeru 2 250),  $r$  je broj redaka u tablici (bez redaka zaglavlja i rubnog retka),  $c$  je broj stupaca u tablici (bez predstupca i rubnog stupca),  $f_{ij}$  su izvorne frekvencije u polju tablice. Oznaka  $i$  u simbolu označuje redni broj retka, a  $j$  redni broj stupca tablice u kojoj je prikazana frekvencija. Simbolom  $e_{ij}$  označena je *očekivana frekvencija*  $i$ -tog retka i  $j$ -tog stupca.

Najmanja vrijednost koeficijenta  $C$  jednaka je nuli, a najveća ovisi o dimenziji tablice. Za tablicu s istim brojem redaka i stupaca ( $r = c$ ) koeficijent je broj

iz intervala:  $0 \leq C \leq \sqrt{\frac{c-1}{c}}$ . Koeficijentu se ne pridružuje predznak jer je porijeklo oblika nominalnih obilježja teorijski proizvoljan.

<sup>1</sup> U izrazu (1)  $\chi^2$  je pomoćna veličina. Čita se hi-kvadrat. Pri računanju vrijednosti hi-kvadrata nije dopuštena uporaba relativnih frekvencija.

Koeficijent kontingencije izračunat ćemo za podatke u tablici 1. Iz tablice je vidljivo da stvarne frekvencije nisu jednake očekivanima pa se ne može reći da je izbor napitaka potpuno neovisan o pripadnosti potrošačkoj skupini. Da bismo izračunali mjeru ovisnosti, najprije ćemo utvrditi vrijednost pomoćne veličine  $\chi^2$ . Ona je:

$$\begin{aligned}\chi^2 &= \frac{(f_{11} - e_{11})^2}{e_{11}} + \frac{(f_{12} - e_{12})^2}{e_{12}} + \dots + \frac{(f_{44} - e_{44})^2}{e_{44}} = \\ &= \frac{(211 - 165.30)^2}{165.30} + \frac{(155 - 169.50)^2}{169.50} + \dots + \frac{(86 - 73.75)^2}{73.75}, \quad \chi^2 = 125.014,\end{aligned}$$

a koeficijent kontingencije:

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}} = \sqrt{\frac{125.014}{2250 + 125.014}}, \quad C = 0.22942.$$

Polje tablice sadržava četiri retka i četiri stupca pa je maksimalna vrijednost koeficijenta:

$$C_{\max} = \sqrt{\frac{c-1}{c}} = \sqrt{\frac{4-1}{4}}, \quad C_{\max} = 0.86603.$$

Koeficijent kontingencije u primjeru pokazuje relativno slab stupanj povezanosti izbora napitka i pripadnosti određenoj potrošačkoj skupini.