## Mérőszám folyóiratok tartalmi szóródásának jellemzésére

A folyóirat által lefedett tárgykör (vagy a felhasználók érdeklődésének tárgykör szerinti) szóródásának mértékét úgy határozzák meg, hogy a tárgykört részekre osztják, majd az egyes részeket előfordulási gyakoriságuk szerint rangsorolják.

Mivel az eljárás önkényes lépést is tartalmaz, az adatoknak csak olyan egyszerű feldolgozása lehetséges, amelynek eredménye még az eredeti részkategóriák felhasználásával interpretálható.

Ennek megfelelően az adott szakterületet nem túl nagyszámú részkategóriára osztották fel. Ezek számát $\underline{n}$-nel jelölték. Ezután a vizsgált folyóiratban közölt cikkeket besorolták a kategóriákba, majd az egyes kategóriákat gyakoriságuk sorrendjében megszámozták. Az összes cikkek számát $\underline{N}$-nel, a rangszámot $\underline{r}$-rel, a gyakoriságot $\underline{f(r)}$-rel jelölték. Egy példa számításban $\underline{N} = 20$ cikk, $\underline{n} = 5$ kategóriába a következő gyakorisággal volt sorolható:

| $\underline{r}$ | 1 | 2 | 3 | 4 | 5 | összes cikk |
|---|---|---|---|---|---|---|
| $\underline{f(r)}$ | 10 | 5 | 3 | 1 | 1 | 20 |

Az alsó sorban leolvasható gyakorisági eloszlás súlyozott átlaga:

$$\underline{m} = \frac{1 \cdot 10 + 2 \cdot 5 + 3 \cdot 3 + 4 \cdot 1 + 5 \cdot 1}{20} = 1{,}90$$

A folyóiratokban közölt cikkek szóródásának mérésére azonban olyan mutatószámot kell használni, amelynek értéke 0 és 1 közé esik, és 0, ha a folyóirat csak egyetlen kategóriából közöl cikkeket, illetve 1, ha a cikkek eloszlása teljesen egyenletes az egyes kategóriák között. Mivel a fenti módon számolva az $\underline{m}$ értéke az első esetben 1, míg az utóbbiban $0{,}5(n+1)$, ezért célszerű a $\underline{D}$ szóródási együtthatót a következő módon definiálni:

$$\underline{D} = 2\frac{m-1}{n-1},$$

s ez a mutatószám már teljesíti a fenti követelményeket.

Látható, hogy rögzített $\underline{n}$ esetén $\underline{D}$ az $\underline{m}$ növekedésével monoton nő. Látható az is, hogy ha egy folyóiratban $\underline{n}$ közül csupán $\underline{r}$ kategóriára oszlanak meg egyenletesen a cikkek, akkor $\underline{D} = 2(r-1):(n-1)$; és a $\underline{D}$ értéke csak ennél kisebb lehet, ha az eloszlás nem egyenletes. Mindebből következik, hogy ha a számított $\underline{D}$ nagyobb mint $(r-1):(n-1)$, akkor a folyóirat legalább $\underline{r}$ kategóriát ölel fel. Például $\underline{n} = 11$ kategória esetén ha azt találjuk, hogy $\underline{D} = 0{,}61$, akkor biztosan legalább hét kategóriából közöl a folyóirat cikkeket.

A következő kis példában négy folyóiratot (P, Q, R, S) hasonlítunk össze, $\underline{n} = 5$ és N = 15 esetére:

| r | 1 | 2 | 3 | 4 | 5 | Összesen | m | D |
|---|---|---|---|---|---|---|---|---|
| P | 3 | 3 | 3 | 3 | 3 | 15 | 3 | 1 |
| Q | 5 | 4 | 3 | 2 | 1 | 15 | 2,35 | 0,66 |
| R | 11 | 1 | 1 | 1 | 1 | 15 | 1,66 | 0,33 |
| S | 15 | – | – | – | – | 15 | 1 | 0 |

Megfigyelhető, hogy $\underline{D}$ e négy tipikus eloszlás esetén egyenlő lépésekben csökken, ami még meggyőzőbben alátámasztja, hogy elfogadhatjuk szóródási mutatónak.

Felmerül a kérdés, vajon $\underline{n}$ növelésével $\underline{D}$ értéke pontosabbá tehető-e. A matematikai statisztika szerint nagyobb $\underline{n}$ esetén lényegesen nagyobb mintára ($\underline{N}$) van szükség, ha azonos megbízhatósági szintet kívánunk elérni. Nagyon nagy minták vizsgálatára azonban – a dolog természetéből következően – nincs lehetőség.

A fenti szóródási mutatószám hibája, hogy nem veszi figyelembe az egyes kategóriák eltérő méreteit, tehát azt a tényt, hogy bizonyos kategóriák eleve ritkábban szerepelnek az adott témakör irodalmában. Ezt kiküszöböli a $\Delta$ *relatív szóródási mutatószám.*

Tegyük fel, hogy ismerték az adott időszakban az egyes kategóriák százalékos gyakoriságát az egész adott területre vonatkozó irodalomban ($p_1$, $p_2 \ldots p_n$). Az ismertetett módon kiszámítják az f(r) gyakoriságokat, rendre képezik és rangsorolják az $f(r):p_r$ hányadosokat, majd ezekből kiszámítják az $\underline{m}$ átlagot és a már ismert módon a $\Delta$ relatív szóródási mutatót. Ennek értéke szintén 0 és 1 közé esik, és csak akkor lesz 1, ha az adott folyóirat a szakterület átlagos gyakorisági rangsora szerint közli a cikkeket.

Minél közelebb esik $\Delta$ a nullához, annál jobban eltér az adott folyóirat szóródása az átlagostól. Az ismertetett mutatószámok nemcsak folyóiratok értékelésében, hanem más területeken is felhasználhatók, ahol valamilyen nem számszerű eloszlást akarunk kvantifikálni.

*/BROOKES, B. C.: A measure of categorical dispersion = The Information Scientist, 11. köt. 1. sz. 1977. p. 11–17./*

(Valkó Péter)