

ELMÉLETI KIEGÉSZÍTÉSEK A 3. FEJEZETHEZ

1. SST és σ^2 felbontása

A kiemelkedően fontos és a továbbiakban sokszor felhasználásra kerülő (3-15) összefüggés a következőképpen igazolható. (3-7) felhasználásával, majd elvégezve $(B_{ij} + K_j)$ négyzetre emelését azt kapjuk, hogy

$$\sum_{j=1}^M \sum_{i=1}^{N_j} d_{ij}^2 = \sum_{j=1}^M \sum_{i=1}^{N_j} (B_{ij} + K_j)^2 = \sum_{j=1}^M \sum_{i=1}^{N_j} B_{ij}^2 + 2 \sum_{j=1}^M \sum_{i=1}^{N_j} B_{ij} K_j + \sum_{j=1}^M N_j K_j^2.$$

Mivel azonban

$$\sum_{j=1}^M \sum_{i=1}^{N_j} B_{ij} K_j = \sum_{j=1}^M K_j \sum_{i=1}^{N_j} B_{ij},$$

s (2-14) miatt $\sum_{i=1}^{N_j} B_{ij} = 0$ minden részsokaságra nézve, a (3-15)-tel egyező

$$\sum_{j=1}^M \sum_{i=1}^{N_j} d_{ij}^2 = \sum_{j=1}^M \sum_{i=1}^{N_j} B_{ij}^2 + \sum_{j=1}^M N_j K_j^2$$

összefüggéshez jutunk. Ennek mindkét oldalát N -nel osztva a (3-14) összefüggéshez jutunk.

2. χ^2 maximális értéke

A (3-19) egyenlőtlenség alsó határa nyilvánvalóan teljesül. A felső határ χ^2 (3-20) formulájára támaszkodva igazolható. A (3-20) formulában szereplő

$$\frac{f_{ij}^2}{f_{i.} f_{.j}}$$

kifejezésre

$$\frac{f_{ij}^2}{f_{i.} f_{.j}} \leq \frac{f_{ij}}{f_{i.}}, \text{ illetve } \frac{f_{ij}^2}{f_{i.} f_{.j}} \leq \frac{f_{ij}}{f_{.j}} \quad (1)$$

teljesül $\forall i, j$ esetén, hiszen $\frac{f_{ij}^2}{f_{i.} f_{.j}} = \frac{f_{ij}}{f_{i.}} \cdot \frac{f_{ij}}{f_{.j}}$ és $\frac{f_{ij}}{f_{i.}} \leq 1$, illetve $\frac{f_{ij}}{f_{.j}} \leq 1$. Az (1)-beli mindkét egyenlőtlenség mindkét oldalát i és j szerint összegezve azt kapjuk, hogy

$$\sum_i \sum_j \frac{f_{ij}^2}{f_{i.} f_{.j}} \leq \sum_i 1 = r, \text{ illetve } \sum_i \sum_j \frac{f_{ij}^2}{f_{i.} f_{.j}} \leq \sum_j 1 = c \quad (2)$$

Ezért

$$\sum_i \sum_j \frac{f_{ij}^2}{f_{i.} f_{.j}} \leq \min(r, c),$$

és így (3-20) miatt valóban

$$\chi^2 \leq N[\min(r, c) - 1] = N \cdot \min\{(r - 1), (c - 1)\}$$

áll fenn.

3. Egy egyszerű PRE-mutató az asszociáció szorosságának mérésére

A 3.2.3. pontban leírt és használt *PRE* eljárás alkalmazásával többféle asszociációs együtttható is képezhető. Ezek közül itt csak a lehető legegyszerűbbeket, az ún. λ mutatókat említjük meg. A $\lambda_{Y|X}$ mutató azt méri, hogy az X szerinti hovatarozás ismerete hány százalékkal csökkenti azt a hibát, amit az Y szerinti hovatarozás becslésekor követünk el.

Ha nem ismerjük az egységek X szerinti hovatarozását, csak azoknak az Y szerinti feltétel nélküli megoszlását, akkor minden egység Y szerinti hovatarozását legkézenfekvőbb a legnagyobb – ún. *modális* – gyakoriságú Y -osztállyal becsülni. Mivel ennek az osztálynak a gyakorisága $\max_j f_{.j}$, ily módon eljárva összesen $N - \max_j f_{.j}$ számú egység Y szerinti besorolása során tévedünk, azaz követünk el hibát. Ezért

$$E_1 = N - \max_j f_{.j}.$$

Egy olyan egység Y szerinti hovatarozását, melyről tudjuk, hogy X szerint a C_i^X osztályba tartozik, azzal az Y osztállyal fogjuk becsülni, amelyre nézve f_{ij} a 3.7 táblázat i -edik során belül – tehát j szerint – maximális. Ezzel a C_i^X osztályba tartozó egységek Y szerinti besorolásakor nyilván $f_{i.} - \max_j f_{ij}$ számú esetben fogunk hibázni. Ezért az összes hiba ekkor

$$E_2 = \sum_i (f_{i.} - \max_j f_{ij}) = N - \sum_i \max_j f_{ij}.$$

Így végül az X és Y közötti asszociáció szorosságának *PRE* mérőszáma

$$PRE = \frac{E_1 - E_2}{E_1} = \frac{\sum_i \max_j f_{ij} - \max_j f_{.j}}{N - \max_j f_{.j}} = \lambda_{Y|X}. \quad (3)$$

E mutató hátrányos tulajdonsága, hogy a 0 értéket nemcsak a függetlenség esetén, hanem akkor is fel tudja venni, ha mind az Y szerinti feltételes eloszlások, mind a feltétel nélküli eloszlás *modális* osztálya megegyezik, de az eloszlások egyébként eltérőek.

Értelmezése ugyanakkor igen szemléletes, mert azt mutatja, hogy az egységek X szerinti hovatarozásának ismerete hány százalékkal csökkenti az azok Y szerinti hovatarozásával kapcsolatos bizonytalanságot, amit jelen esetben az Y szerint helytelenül osztályozott egységek számával mérünk. Ehhez teljesen hasonlóan konstruálható meg a $\lambda_{X|Y}$ mutató, ami azt mutatja, hogy az egységek Y szerinti hovatarozásának ismerete hogyan javítja azok X szerinti hovatarozásának becslését. Megjegyezzük, hogy általában $\lambda_{Y|X} \neq \lambda_{X|Y}$.

Megemlítjük végül, hogy a szakirodalomban számos olyan, a λ -hoz hasonló mutatószám is ismeretes, ami az egységek valamely ismértv szerinti hovatarozásának becslése során még azt is megköveteli, hogy az egységeknek az adott ismértv szerinti becsült abszolút megoszlása egyezzen meg az egységeknek az adott ismértv szerinti tényleges megoszlásával. Ezekkel a mutatószámokkal azonban a tananyagban nem foglalkozunk.

A 3.5. példa esetében a $\lambda_{Y|X}$ mutató számítása a következőképpen történik. Ha nem ismerjük a foglalkoztatottak nemét, akkor legmagasabb iskolai végzettségüként nyilván a középiskolát célszerű megjelölni, mert a 3.3. táblázat *Összesen* sora szerint a legtöbb foglalkoztatott ilyen legmagasabb végzettséggel rendelkezik. Ezzel összesen $E_1 = 3901,5 - 1299,3 = 2602,2$ ezer fő esetében tévedünk. Ha tudjuk egy foglalkoztatotról, hogy ő férfi, akkor a Szakmunkásképző, szakiskola végzettség a célszerű tipp a legmagasabb iskolai végzettségre, amivel $2116,1 - 862,5 = 1253,6$ ezer fő esetében hibázunk. A nők esetében a célszerű tipp a középiskola, az elkövetett hibák száma pedig $1785,4 - 713,5 = 1071,9$ ezer fő. Ezért

$$E_2 = 1253,6 + 1071,9 = 2325,5 \text{ ezer fő.}$$

Így végül (3) alapján

$$\lambda_{Y|X} = \frac{2602,2 - 2325,5}{2602,2} \approx 0,106.$$

Ez azt jelenti, hogy a jelen esetben a nem ismerete mintegy 11 %-kal csökkenti a legmagasabb iskolai végzettséget illető bizonytalanságot, azaz a két ismerv között meglehetősen laza a kapcsolat. Ugyanezt a következtetést lehetett levonni az asszociációs együtttható alapján is.